

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Crop type mapping by using transfer learning

Artur Nowakowski^{a,*}, John Mrziglod^{a,b}, Dario Spiller^c, Rogerio Bonifacio^b, Irene Ferrari^b, Pierre Philippe Mathieu^a, Manuel Garcia-Herranz^d, Do-Hyung Kim^d

^a Φ -Lab, EOP, European Space Agency, ESRIN, Frascati, Rome, Italy

^b Climate and Earth Observation Team – Research, Assessment and Monitoring Division, World Food Program HQ, Rome, Italy

^c Italian Space Agency, Rome, Italy

^d Office of Innovation, UNICEF, NY, USA

ARTICLE INFO

Keywords:

Crop detection
Transfer learning
Convolutional neural networks
Drone images

ABSTRACT

Crop type mapping currently represents an important problem in remote sensing. Accurate information on the extent and types of crops derived from remote sensing can help managing and improving agriculture especially for developing countries where such information is scarce. In this paper, high-resolution RGB drone images are the input data for the classification performed using a transfer learning (TL) approach. VGG16 and GoogLeNet, which are pre-trained convolutional neural networks (CNNs) used for classification tasks coming from computer vision, are considered for the mapping of the crop types. Thanks to the transferred knowledge, the proposed models can successfully classify the studied crop types with high overall accuracy for two considered cases, achieving up to almost 83% for the Malawi dataset and up to 90% for the Mozambique dataset. Notably, these results are comparable to the ones achieved by the same deep CNN architectures in many computer vision tasks. With regard to drone data analysis, application of deep CNN is very limited so far due to high requirements on the number of samples needed to train such complicated architectures. Our results demonstrate that the transfer learning is an efficient way to overcome this problem and take full advantage of the benefits of deep CNN architectures for drone-based crop type mapping. Moreover, based on experiments with different TL approaches we show that the number of frozen layers is an important parameter of TL and a fine-tuning of all the CNN weights results in significantly better performance than the approaches that apply fine-tuning only on some numbers of last layers.

1. Introduction

Crop type maps are vital for many applications in land monitoring and policy. It includes, but it is not limited to, improvement of crop yield models (Kogan et al., 2013), producing accurate agricultural statistics (Gallego et al., 2010), and developing hydrological models (Yin and Williams, 1997). It can help modelling of flood damage estimation and water quality (Foerster et al., 2012) as well as managing food supplies and children's welfare in developing countries (Jensen et al., 2006). In this paper, we address the problem of improving the quality of crop types maps by using supervised transfer learning (TL) and remote sensing (RS) data from drones.

Supervised machine learning (SML) is the ability of machines to learn from data by designing predictive models. For classification

purposes, descriptive information known as label are associated to each feature in the input data and the general task consists of learning an unknown predictive function mapping from data to labels. Every predictive model is characterized by specific parameters (or weights) that must be optimized on the input data and labels during the training phase. After the training, the model can be used to predict the corresponding label of new input instances. The usage of SML in agriculture has been considered in the review studies reported in (Liakos et al., 2018) and (Kamilaris and Prenafeta-Boldú, 2018). The combination of aerial imagery and deep learning for scene and environment monitoring has been already proposed and is analysed in (Maktab Dar Oghaz et al., 2019). Based on the application, five major research groups have been identified: vegetation identification, classification and segmentation (Fan et al., 2018; Ji et al., 2018; Jin et al., 2018; Rebetz et al., 2016),

* Corresponding author.

E-mail addresses: artur.nowakowski@esa.int (A. Nowakowski), john.mrziglod@esa.int (J. Mrziglod), dario.spiller@esa.int (D. Spiller), rogerio.bonifacio@wfp.org (R. Bonifacio), irene.ferrari@wfp.org (I. Ferrari), pierre.philippe.mathieu@esa.int (P.P. Mathieu), mgarciherranz@unicef.org (M. Garcia-Herranz), dokim@unicef.org (D.-H. Kim).

<https://doi.org/10.1016/j.jag.2021.102313>

Received 15 October 2020; Received in revised form 9 January 2021; Accepted 30 January 2021

Available online 18 February 2021

0303-2434/© 2021 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

crop counting and yield predictions (Dijkstra et al., 2019; Rahneemofar and Sheppard, 2017; Tri et al., 2017), crop mapping (Nijhawan et al., 2018), weed detection (Dyrmann et al., 2016), and crop disease and nutrient deficiency detection (di Gennaro et al., 2016; Mardanisamani et al., 2019).

In the past few years, the concept of TL has been introduced as it greatly improves the learning performance on reduced datasets thus avoiding time-expensive data-labelling efforts. A precise definition of TL is reported in (Pan and Yang, 2010), and can be briefly summarized as the ability to use a pre-existent model trained on a given different (usually huge) source dataset to improve the prediction on the target problem. This operation is often performed by training the given model on the target dataset starting from the weights already optimized on the pre-existent source problem. TL is emerging as a new learning framework to address SML problems due to the inherent simplification in modelling and training.

In image analysis the most successful TL solutions use deep neural networks (DNN) trained on huge computer vision datasets, e.g. ImageNet (J. Deng et al., 2010), MNIST (L. Deng, 2012), and Open Images Dataset (Kuznetsova et al., 2018). The success is based on the observation that general features are learned at the first layers of DNN regardless of the type of image classification task and dataset (Yosinski et al., 2014). Consequently, learned weights from all or selected layers can be copied to the same architecture used for another problem. The most used models for transfer learning applied for agriculture scopes are Faster R-CNN (Bargoti and Underwood, 2017; Jin et al., 2018; Sa et al., 2016), AlexNet (Lee et al., 2015; Mohanty et al., 2016; Nijhawan et al., 2018; Reyes et al., 2015; Valente et al., 2019; Yalcin, 2017), GoogLeNet (Mohanty et al., 2016; Ramcharan et al., 2017; Tri et al., 2017), SegNet (Bosilj et al., 2019), and VGG (Xie et al., 2016).

With advances in satellite, airborne and ground based RS, reflectance data are increasingly being used in agriculture (Wójtowicz et al., 2016). Various RS methods designed to optimize profitability of agricultural crop production and protect the environment have been developed, e.g. ground-based RS, airborne RS, and satellite imagery. Moreover, different sensors and technologies can be used, such as digital cameras, multispectral cameras, hyperspectral cameras, thermal infrared imagers and LiDAR. A comparison of RS platforms, along with their pros and cons, is reported in (Pádua et al., 2017), describing how unmanned aerial vehicles (UAVs) can complement the established manned aircraft and satellite platforms. Precision agriculture based on UAVs represents a robust, timely, cost effective way to obtain viable data on the farm to improve yields and overall profitability in sustainable farming systems (Abdullahi et al., 2015). With this respect, McCabe et al. (2016) reported that the high resolution sensing from UAVs is currently more advantageous than RS from Earth-observing satellites when considering latency in satellite data processing and delivery as well as the overall management costs. Some examples of applications of UAVs for agriculture applications are classification of maize in complex farming systems (Hall et al., 2018; Wahab et al., 2018), land use and vegetation mapping (Berni et al., 2009; C. Swain and Uz Zaman, 2012; Heng et al., 2018; Senthilnath et al., 2017), crop and weed classification for smart farming (Lottes et al., 2017; Peña et al., 2013), phenotyping (Yang et al., 2017), automatic identification of plant disease (Boulent et al., 2019).

To the best of authors' knowledge, the combination of transfer learning (TL) and UAV data for crop type classification has been introduced only very recently, as it is shown in (Chew et al., 2020). The authors proposed a TL technique based on VGG16 on a specific region in Rwanda but neither a comparison with other models nor a parameter analysis has been carried out. The proposed approach is applied on areas of Malawi and Mozambique, where a small number of similar studies has been carried out so far, e.g. (De Bie et al., 2008; Gumma et al., 2019). With respect to the current state of the art, the contributions of this paper are:

1. TL is applied to crop type mapping using UAV data, providing a confirmation that TL can be successfully applied reducing the user design effort and the computational effort.
2. The robustness of the TL approach is shown using two different pre-trained models, i.e. GoogLeNet and VGG16.
3. A performance analysis on the results varying the number of frozen layers in the models is carried out.

The rest of the paper is organized as follows. Section 2 deals with the description of the study areas in Malawi and Mozambique, while Section 3 focuses on the analysis of the input datasets. In Section 4 the transfer learning approach is described and the results from GoogLeNet and VGG16 are reported in Section 5. A critical discussion of the output of this study is reported in Section 6 and conclusions are given in Section 7.

2. Study areas

In this research, two different datasets are used. Both of them consist of drone images and reference data for each crop type collected through extensive field visits. One dataset was acquired in Kasungu District Malawi, the other in Gaza Province, Mozambique. In both areas, the growing season unfolds from October to May with planting viable from November and harvests possible from late March onwards. Most of Gaza is drier than Kasungu, however its near coastal areas receive plentiful rains. In both areas, maize is the heavily dominant crop.

2.1. Malawi dataset

The first dataset was collected by drone and field campaigns in the Kasungu district in Malawi during May 2018. The dataset includes 40 orthomosaic maps (each created from about 100–200 single images) with RGB spectral bands acquired using a CMOS camera mounted on a drone flying on altitude less than 400 m. Each of the mosaics depicts an area of about 1 km² with the ground resolution about 4 cm. Localization of the data along with some sample mosaics are presented in Fig. 1. As the reader can see, it is noteworthy that each mosaic contains more than one cultivated field. Indeed, along with the drone campaign, different polygons (shown in red in Fig. 1) are identified and labelled with the support of enumerators sent in the areas of drone data acquisitions. On average, a little more than 41 polygons were identified in each mosaic. Among the different crop types identified with the drone campaign, only the five most representative types have been considered in this study, i.e. Cassava, Groundnut, Maize, Sweet potatoes, and Tobacco. The fallows and minority crops class collects fallows and all the remaining crop types, which are: Beans, Cow peas, Pigeon peas and Soya beans. Collected information about agricultural fields includes perimeters of plots, crop types and general development stages. The overall area and the number of samples for each crop type within the input dataset are shown in Table 1. One can note that the number of fields and the related covered area is not uniform among the five classes.

2.2. Mozambique dataset

The second dataset was collected in the south part of Gaza province in Mozambique during spring 2019. The dataset includes 30 orthomosaic maps with RGB spectral bands acquired using a CMOS camera (Sequoia 4.9 with resolution of 4608 × 3456 pixels) mounted on a fixed-wing drone flying on altitude of about 200 m. Each of the mosaics depicts an area varying from about 0.1 to 0.3 km² with ground resolution of about 1.3 cm. Note that in this drone campaign the ground resolution is different with respect to the Malawi campaign, as can be also noted by looking at Fig. 1. Crop type reference data were collected and processed similarly to the Malawi case described above. In this case, the crop types chosen for the classification are Cassava, Maize, and Rice, whereas the remaining crop types have been included in the minority crops class. Statistics about the reference crop types in the Mozambique datasets are

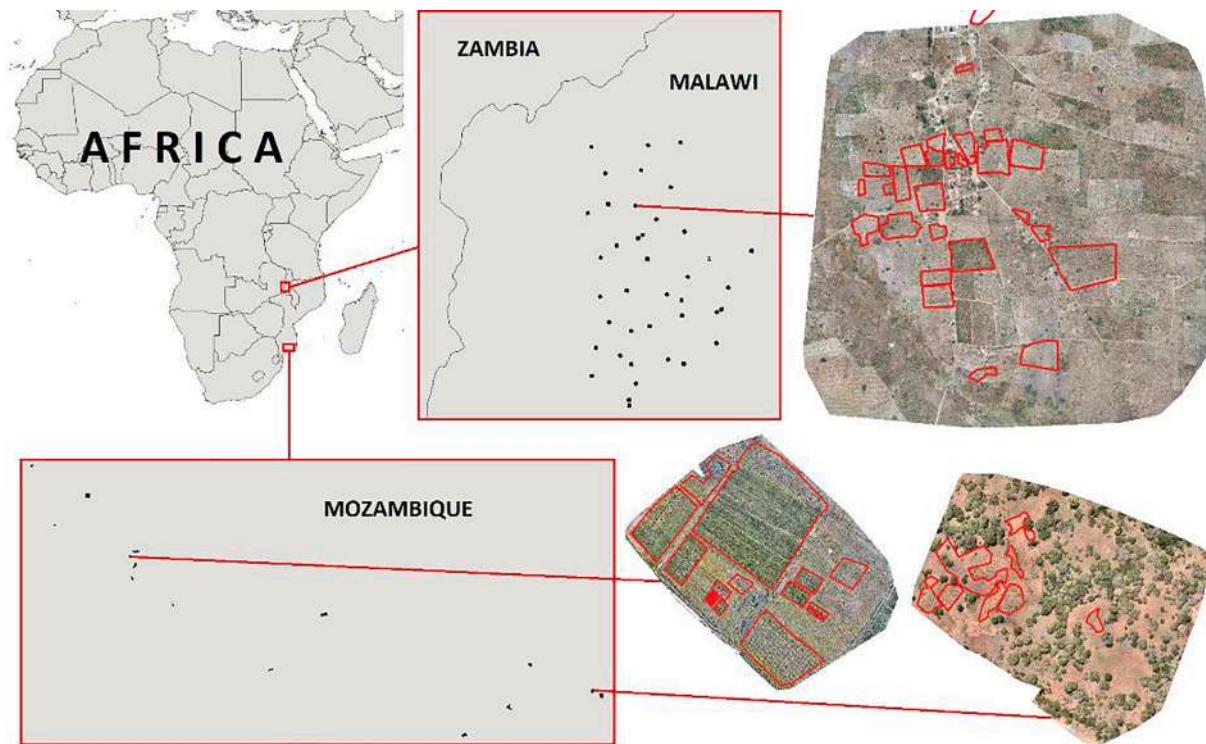


Fig. 1. Localization of drone mosaics within the input datasets. Three examples of mosaics with perimeters of reference agricultural plots are shown.

Table 1
Number of reference samples in Malawi dataset.

Crop type	Developing stage					
	Cultivated field		Post-harvested field		Sparse planting	
	Area (km ²)	N. of items	Area (km ²)	N. of items	Area (km ²)	N. of items
Cassava	0.123	32	0	0	0	0
Groundnut	0.127	78	0.443	123	0.01	6
Maize	1.068	216	1.328	287	0.0232	5
Sweet potatoes	0.025	37	0.013	8	0.001	1
Tobacco	0.142	61	0.390	90	0.099	9
Fallows and minority crops	5.215 km ² , 706 items					
Total	9.00 km ² , 1659 items					

shown in Table 2. As in the previous Malawi dataset, the number of fields and the related covered area is not uniform among the three classes.

Table 2
Number of reference samples in Mozambique dataset.

Crop type	Developing stage					
	Cultivated field		Post-harvested field		Sparse planting	
	Area (km ²)	N. of items	Area (km ²)	N. of items	Area (km ²)	N. of items
Cassava	0.010	16	0	0	0.032	20
Maize	0.296	110	0.028	17	0.036	14
Rice	0.317	57	0.001	2	0.002	2
Fallows and minority crops	0.215 km ² , 95 items					
Total	0.938 km ² , 333 items					

3. Analysis of the input datasets

3.1. Preliminary analysis and corrections

As it is commonly known, the characteristics of the input training dataset determine the success of any classification algorithm. Correct features with reliable labels are key factors for the success of the training procedure. With this regard, a preliminary inspection of the given input datasets underlined some accidental labelling errors and co-registration shifts. A manual correction of these errors has been carried out in order to use a consistent input dataset. Polygons indicating field boundaries have been checked and adjusted to fit the real geometry of the cultivated field (for instance, the red polygons in Fig. 1). Other manual interventions included corrections of harvested crops marked as standing (and vice versa), partially harvested fields marked uniformly as standing crops (and vice versa), different crop types within one polygon.

Another important modification of the original dataset was the introduction of a further attribute describing the cultivation stage in order to discriminate among cultivated field, post-harvested field, and sparse planting. Indeed, only cultivated fields are associated to the crop class during the testing and validation process, whereas post-harvested and sparse planting fields are merged with the fallows and minority crop class to form the “other” class.

3.2. Challenges of drone-based crop type detection

The drone images shown in Fig. 1 have been used to generate input images for the networks. To be consistent with the input shape default parameters of the convolutional networks described in Section 4.2, images with dimension 224 × 224 × 3 are sampled from the mosaics, where 224 is the number of horizontal and vertical pixel and 3 is the number of channels (RGB) (there was no need to change the input shape and require additional effort in the fine-tuning process). Some reference images are shown in Fig. 2 for the five crop types within the Malawi dataset. The reader can notice the differences among the crop type looking at the colours and distribution of the plants. It is noteworthy that

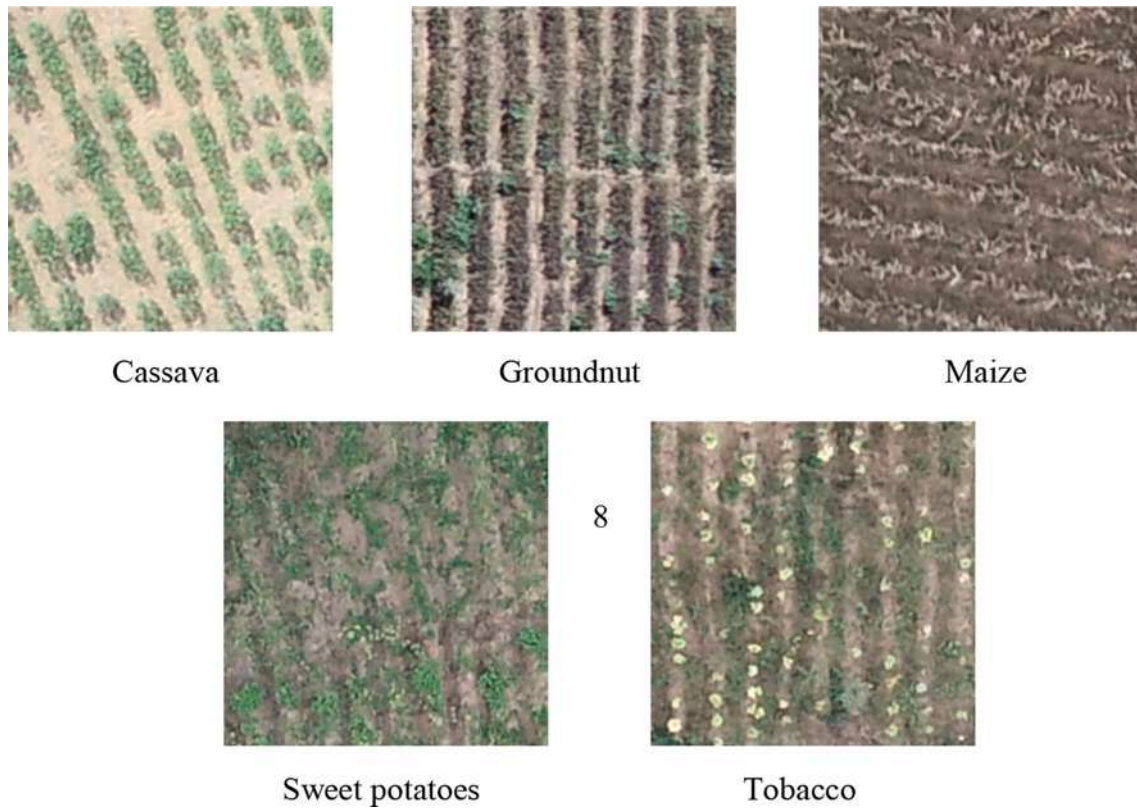


Fig. 2. Example of images for the five classes of crops in Malawi.

the images have been collected when a large number of maize fields was already harvested or dry. Moreover, the dataset contains images collected in different hours during the day, with different lightning

conditions passing from the morning to the evening.

Even inside one crop class, the collected images can appear very different from each other. Without loss of generality, in Fig. 3 six

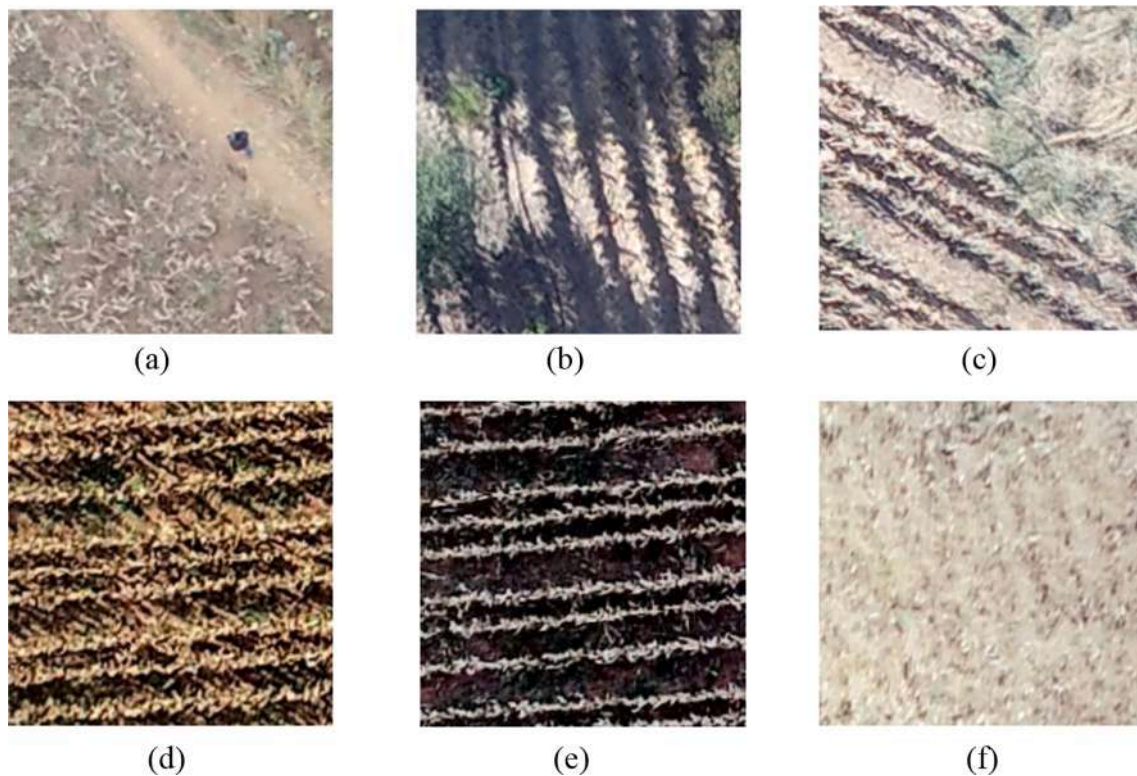


Fig. 3. Examples of different images labelled as cultivated maize fields in the Malawi dataset.

different images within the cultivated maize class are shown where the reader can appreciate some of the most relevant issues. For instance, in Fig. 3(a) a road passes through the cultivated field, whereas in Fig. 3(b) shadows from clouds or close trees are observed. In Fig. 3(c) the field is partially affected by the presence of weeds. Finally, in Fig. 3(d-e-f) crop fields with very different lightning conditions and/or soil states are reported. These issues are denoted for all the crop types used in this study. To cope with the problems of distinguish the crop types even in the presence of all these disturbances, convolutional neural networks have been chosen as they have been designed to identify class objects in different pose and light conditions. Indeed, such models are not pixel based since they take into account the spatial distribution of the information by means of the convolutional operations. In this way, the network does not only consider the spectral information contained in the colour bands but it takes into account the features distributions such as the presence of patterns associated to each different crop (Zhang, 2018).

4. Transfer learning approach

Supervised machine learning (SML) is the ability of machines to learn from data, x , when provided with descriptive information, y , known as label. Given a domain $D = \{X, P(X)\}$ consisting of a feature space X and a marginal probability distribution $P(X)$, where $X = \{x_1, \dots, x_n\} \subset X$, the task $T = \{Y, f(\cdot)\}$ consists of learning the unknown predictive function $f(\cdot)$ using the pairs $\{x_i, y_i\}$, where $x_i \in X$ and $y_i \in Y$. The function $f(\cdot)$ can be used to predict the corresponding label, $f(x)$, of a new instance x .

In these last years, the concept of TL has been introduced as it greatly improves the learning performance on reduced datasets thus avoiding time-expensive data-labelling efforts. As stated in (Pan and Yang, 2010), given a source (S) domain D_S and learning task T_S , a target (T) domain D_T and learning task T_T , TL aims to help improve the learning of the

target predictive function $f_T(\cdot)$ in D_T using the knowledge in D_S and T_S , where $D_S \neq D_T$, or $T_S \neq T_T$. In this paper, TL is applied by pre-training deep learning models on T_S and fine-tuning the same models on T_T , i. e. using the weights of the pre-trained model as initial guess for the weights of the target model.

The application of the transfer learning approach for DNN from computer vision to crop types mapping must take into account the following steps: 1) choosing a computer vision dataset, 2) choosing an architecture to be pre-trained on this dataset, 3) defining the method for transferring learnable weights from the pre-trained network to the target one aiming at crop types classification, and 4) applying data augmentation for reduced target datasets. These points are explained in the following sections. A flowchart of the proposed transfer learning approach is shown in Fig. 4 with details about the required operations along with input/output data and involved AI models

4.1. Computer vision dataset

Among the different image datasets existing in computer vision domain, ImageNet (J. Deng et al., 2010) has now become the standard for pre-training image features to be transferred for solving many image analysis problems (Huh et al., 2016). This dataset consists of 1.2 million of labelled images divided into a thousand classes. The classes are organized hierarchically and includes images of natural and artificial objects, including animals, plants, devices, goods, vehicles, houses, geological structures, etc. Very often the most detailed sub-classes represented in the dataset are very similar to each other and allow for fine-grained recognition between objects (e.g. different plants or different breeds of dogs and cats). Using ImageNet, good classification results can only be achieved when the DNN can learn very specific image features to discriminate images basing on very accurate and detailed analysis. The ImageNet database contains some classes that are directly helpful for

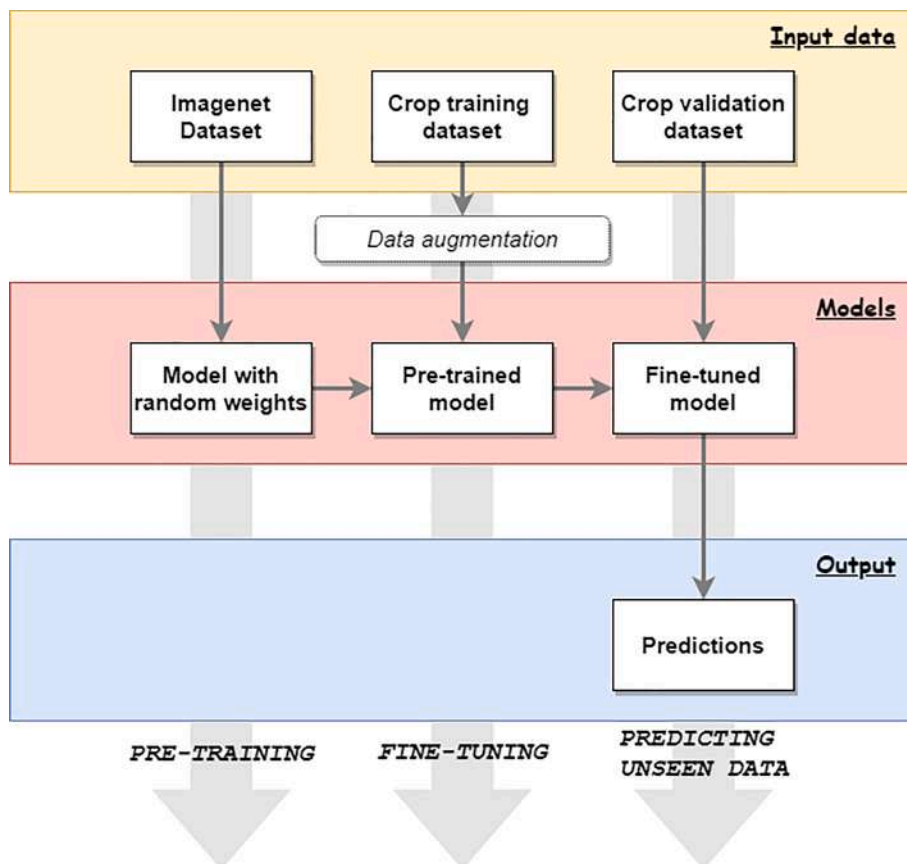


Fig. 4. Flowchart of the proposed transfer learning approach.

crop type mapping (e.g. geological formations, plants).

4.2. Pre-trained architecture

During the last two decades a lot of network architectures trained on ImageNet have been published. In this paper, GoogLeNet and VGGNet are adopted. Both of them have already been successfully applied in transfer learning problems (Huh et al., 2016; Mehdipour Ghazi et al., 2017; Xie et al., 2016; Du et al., 2020). Both architectures are examples of convolutional neural networks (CNN) which have been inspired by the working characteristics of human brain. CNN can successfully explore spatial dependencies in images through the ability to learn two dimensional features using convolutional filters.

The GoogLeNet architecture (Szegedy et al., 2015) is a deep and wide architecture with 22 layers (only counting convolutional or fully connected layers, see Fig. 5). The number of parameters (about 6.8 million) is considerably low compared to other common networks (e.g., AlexNet has about 60 million parameters). The core features of GoogLeNet are the Inception module and the convolutional layers for dimension reduction (i.e. 1×1 convolutions followed by Relu also known as network in network aiming at enhancing model discriminability for local patches within the receptive field (Lin et al., 2013)). Usually, at each layer of a traditional convolutional network either a max pooling operation or a convolutional operation is found. With the Inception module, these two operations are performed in parallel. Indeed, the inception module uses parallel 1×1 , 3×3 and 5×5 convolutions along with a max-pooling layer in parallel, hence enabling it to capture a variety of features in parallel. To lower the computation needs, 1×1 convolutions before the above mentioned 3×3 , 5×5 convolutions (and after the max-pooling layer) are added for dimensionality reduction. Finally, a filter concatenation layers concatenates the outputs of all these parallel layers. As depicted in Fig. 5, GoogLeNet is composed by nine inception modules, three convolutional layers (the second one after the input is for dimension reduction), two normalization layers, four max-pooling layers, one average pooling, one fully connected layer, and a linear layer with softmax activation in the output. During training, GoogLeNet connects two auxiliary classifiers to the intermediate layers of the network to improve the backpropagation performances thus

making the earlier stages more discriminative. VGG16 is a convolutional neural network model proposed by K. Simonyan and A. Zisserman from the University of Oxford (it is also known as VGGNet with 16-layer, see configuration D in the original paper, (Simonyan and Zisserman, 2015)). The network is represented in Fig. 5. It uses a homogeneous architecture to investigate the effects of increased convolutional network depth on performance. There are 16 layers with learnable weights, divided into 13 convolutional layers with 3×3 convolutional filters and 3 fully connected layers. The number of learnable parameters is 138 million (quite bigger than other networks as AlexNet and GoogLeNet). Five max pooling layers are introduced to reduce the spatial size of the input volume. It is noteworthy that the number of filters doubles after each maxpool layer, reinforcing the idea of shrinking spatial dimensions and growing in depth. Given the very high number of learnable parameters, VGG16 should be trained using GPUs and parallel computation.

4.3. Transfer learning method

The DNNs described in the previous section were adapted to the crop types mapping problem by applying the TL methodology. However, the TL approach can be used in different ways by varying the number of ‘frozen’ layers. In the frame of this paper, a layer is called frozen when the learnable weights are directly taken from the computer vision training, i.e. weights are optimized on a different problem and not optimised to the target problem. With this regard, the simplest approach deals with replacing the last (classification) layer in order to produce the desired number of crop types classes instead of the original one thousand categories. Accordingly, both in the GoogLeNet and VGG16 architecture the fully connected layer with one thousand outputs was removed and a new fully connected layer with 6 outputs was inserted (the following softmax layer was also changed). All the remaining weights for the rest of the layers have been preserved from the computer vision training.

However, the open question that needs to be addressed is whether it will be more profitable to fine tune all the weights or only part of them. According to (Yosinski et al., 2014), when the target dataset is small and the number of parameters is high it is better to freeze weights in the first layers instead of fine-tuning them in order to avoid overfitting. This is especially applicable in the present case where only a small dataset is available. Indeed typical DNNs requires huge amount of input data (dozens or hundreds of thousands samples) to avoid overfitting. However, the optimal number of first layers to be frozen remains unknown and selecting the correct setup is difficult for many reasons. First of all, transfer learning is applied between different objectives, starting from a much larger classification dataset. Another reason comes from co-adapted features that can be found through many consecutive layers typically present in pre-trained DNN (Yosinski et al., 2014). Accordingly, a parametric analysis with different numbers of frozen layers is performed to empirically select the best condition and investigate if there exist any tendency.

During learning process we used stochastic gradient descent with momentum (SGDM) optimizer, batch size was equal to 100. The initial learning rate was set to $3e-4$ and was constant during all training epochs. The networks were trained for 20 epochs in order to assure that the plateau in learning curve was reached in every case (an example of training and validation loss and accuracy is presented in Fig. 6). Training data were shuffled every training epoch.

4.4. Definition of training and validation datasets

An accurate definition of training and validation datasets is required for the obtainment of good classification scores as well as for a valid interpretation of the results. To tackle the problem of spatial autocorrelation arising in closely distributed features, training and validation images have been selected from separate mosaics. For the Malawi campaign, training images have been selected from 36 mosaics whereas the other 4 mosaics have been used for validation. For the Mozambique

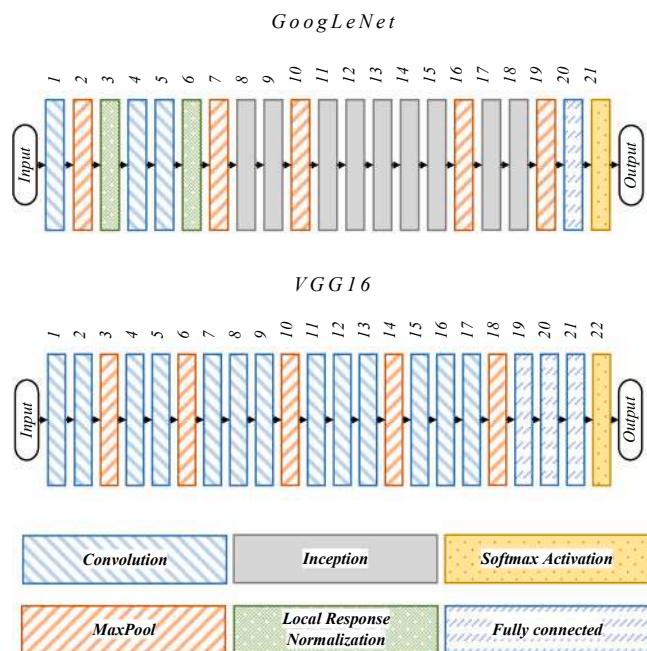


Fig. 5. Schematic view of the architecture of GoogLeNet (top) and VGG16 (bottom). The legend shows the type of layer.

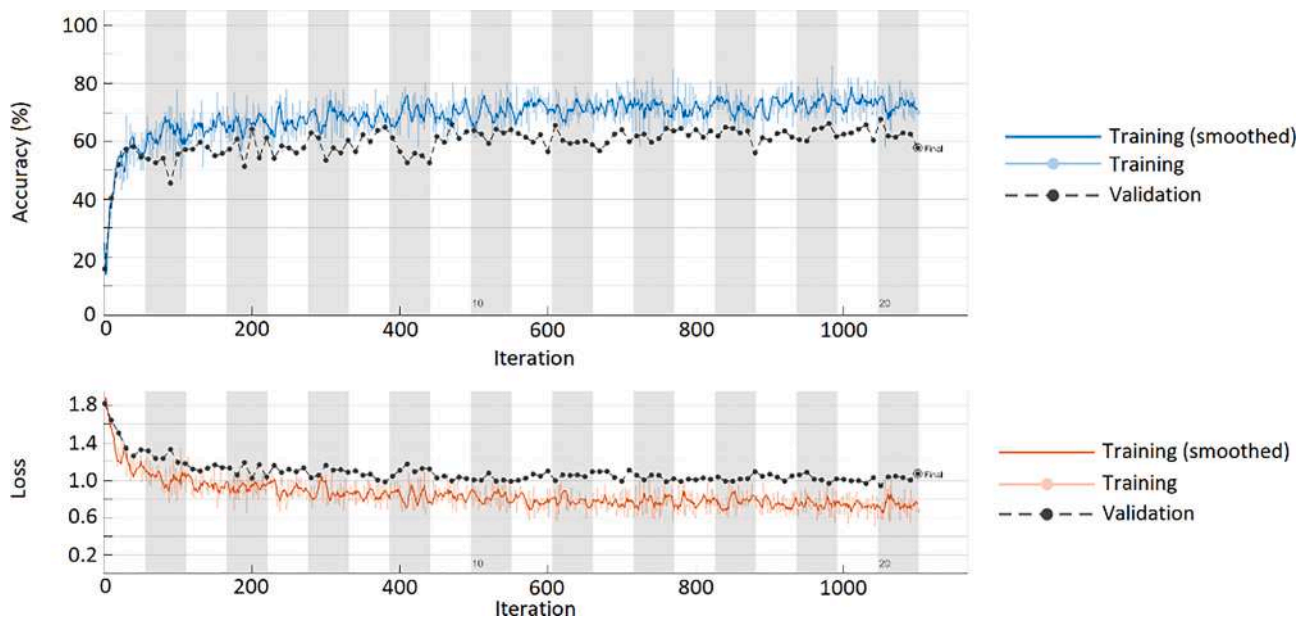


Fig. 6. An example of training and validation accuracy and loss for the Malawi dataset for 20 epochs.

campaign, training images have been selected from 26 mosaics whereas the other 4 mosaics have been used for validation. The selection of the drone mosaics used for validation has been automatically performed using a greedy approach assuring that the classes occurrences in the validation dataset is as close as possible to the classes occurrences in the training dataset. Therefore, the autocorrelation problem is minimized while maintaining a good representativeness of the crop type occurrences in the validation dataset.

For the training datasets, 1000 RGB images per class with size $244 \times 244 \times 3$ have been selected, where the centres of the images were placed at 1000 randomly selected points within the crop fields. Note that the number of input images is lower than the usual number of samples used to train CNNs (Ian Goodfellow, Yoshua Bengio, 2016). This is consistent with the transfer learning approach, where only a reduced number of new information (related to the new problem taken into account) are needed for the fine tuning of the model.

Due to the strong imbalance in the distribution of the fields, as already noted from Table 1 and Table 2, the number of training images for each class should be selected carefully. Indeed, if choosing the number of images in a proportional way with respect to the data distribution, a high risk of overfitting is expected during the training phase (Buda et al., 2018). Imposing the same number of training samples per class improves the network ability to perform an unbiased discrimination among all classes rather than focusing on the most popular crop types only. On the other hand, for the validation dataset, a number of images proportional to the real distribution is chosen in order to reflect the real output of the model. For instance, the number of validation images for the Malawi campaign is 84 for cassava, 83 for groundnut, 542 for maize, 2216 for other, 12 for sweet potatoes, and 64 for tobacco. In this way, the performances evaluation will reflect the real class distribution and provide a consistent and reliable description of the results.

4.5. Data transformation and augmentation

Image samples depicting different crop types have been extracted automatically using the reference dataset. As already introduced in Section 3.1, each crop class is divided into cultivated field, post-harvested field, and sparse planting. To obtain more reliable results, only cultivated fields are associated to the crop class during the testing and validation process. Post-harvested and sparse planting fields are merged with the minority crop class to form the “other” class.

Before inputting image samples to the DNNs, a data augmentation approach has been used. Each image training sample was transformed using random values of all the considered operations: horizontal reflection (50% of chance), rotation ($0-360^\circ$), translation (up to ± 30 pixels) and scaling (0.9 to 1.1). This kind of augmentation was applied to overcome the problems related to low number of polygons, especially for minority classes. In this way, the risk that the classifier will distinguish a crop type only basing on geometrical features (like the angle of plant rows, shadow orientation, and other features related to the specific acquisition campaign taken into account) is minimized, enhancing the chance to distinguish the crop type looking at more general information such as the spectral characteristics (Yu et al., 2017).

5. Results

5.1. Definition of the scores and implementation setup

We conducted a set of experiments on presented method to assess its reliability using different evaluation metrics and varying number of frozen layers. The macro-averaged evaluations of overall accuracy, precision, recall and F1-score have been used to compare the results.

Given N input classes and the generic confusion matrix with elements $c_{i,j}$ (element in the i -th row and j -th column), the total number of input sample for the i -th class is defined as summation along the columns, i.e.

$$T_{c,i} = \sum_{j=1}^N c_{j,i},$$

whereas the total number of classified samples for the i -th class is defined as summation along the rows, i.e.

$$T_{r,i} = \sum_{j=1}^N c_{i,j}.$$

Consequently, the total number of samples is

$$T = \sum_{i=1}^N T_{c,i} = \sum_{i=1}^N T_{r,i}.$$

Finally, the precision P_i and the recall R_i for the i -th class are evaluated as

$$P_i = \frac{c_{i,i}}{T_{r,i}}, R_i = \frac{c_{i,i}}{T_{c,i}}.$$

Note that, in remote sensing applications, precision and recall are

often referred to as user’s accuracy and producer’s accuracy, respectively (Barsi et al., 2018). Basing on these basic definitions, the formulation of the macro-averaged evaluations of overall accuracy, precision, recall and F1-score are reported in Table 3. The column related to the macro averaged evaluation of the scores simply reports the arithmetic mean of the scores given by the single classes. On the contrary, the weighted evaluation of the scores is related to a weighted average of the scores per single class, thus considering the relative occurrences of the classes within the overall input dataset.

The approach presented in the paper was implemented in Matlab® 2020 environment. Deep Learning Toolbox was used to download and fine-tune pre-trained networks, while Mapping toolbox was used to handle geo-referenced data.

5.2. Results for the Malawi dataset using GoogLeNet

The GoogLeNet results for the Malawi dataset are reported on the left of Fig. 7, where the number of layers is consistent with the layers numbering in Fig. 5. First, one can see that the results generally get worse increasing the number of frozen layers. When no layer is frozen, all the pre-trained weights in every layer are used as initial guess for the fine-tuning process and all of them are adapted to the crop detection problem. For this campaign, a noteworthy difference in weighted and macro-averaged scores is observed, which implies that the imbalanced distribution of the classes reported in Table 1 and Table 2 strongly affects the final results. In particular, this result suggests that a relevant inhomogeneity characterizes the input dataset. It is quite relevant to note that the weighted precision scores in Fig. 7 almost remains constant varying the number of frozen layers. The high value of weighted precision means that a small percentage of false negative is observed, even though the lower weighted recall values suggest that, even choosing the best values of frozen layers, around 25% of input values are not properly recognized. The confusion matrix reporting the recall values for the crop types of the Malawi campaign is shown on the left of Fig. 9. The values related to the model with highest overall accuracy have been chosen, which correspond to the one with 2 frozen layers. As can be seen, groundnut represents the most problematic crop type as it is often confused with sweet potatoes and tobacco.

5.3. Results for the Malawi dataset using VGG16

The VGG16 results for the Malawi dataset are reported on the right of Fig. 7. It can be noted that the vertical axes of the two plots in Fig. 7 cover the same interval and VGG16 presents a similar trend to the one given by GoogLeNet.

As noted in Section 5.2, also in this case results suggest that freezing a small number of the layers (or even no freezing at all) is a better solution instead of freezing all of them except the last fully-connected layer. Indeed, when the number of frozen layers is 0 or 1, the F1 score is greater than 0.85, indicating a very good classification ability of the VGG16 model. The confusion matrix for the VGG16 model is shown on the right of Fig. 9. The values related to the model with highest overall

Table 3
Definition of the evaluation scores.

	Macro averaged evaluation	Weighted evaluation
Overall accuracy	$OA = \frac{\sum_{i=1}^N C_{ii}}{T}$	N.A.
Precision	$P = \sum_{i=1}^N P_i$	$P_w = \frac{\sum_{i=1}^N (T_{ci}P_i)}{T}$
Recall	$R = \sum_{i=1}^N R_i$	$R_w = \frac{\sum_{i=1}^N (T_{ci}R_i)}{T} = \frac{\sum_{i=1}^N C_{ii}}{T} = OA$
F1	$F1 = 2 \frac{PR}{P+R}$	$F1_w = 2 \frac{P_w R_w}{P_w + R_w}$

accuracies have been chosen, which correspond to the one with no frozen layers. On average, the recall values in this case are a little bit better than the ones obtained with GoogLeNet.

5.4. Results for the Mozambique dataset using GoogLeNet

The GoogLeNet results for the Mozambique dataset are reported on the left of Fig. 8. In this case, the difference between macro-averaged and weighted results is pretty much reduced with respect to the previous Malawi study case. This result suggests that the input classes are better distributed in the input dataset. The best results are obtained when no layers are frozen, reaching almost 85% in the weighted F1 score. The recall values for each crop type are reported in the confusion matrix shown on the left of Fig. 10. As before, the values related to the model with highest overall accuracies have been chosen, which correspond to the one with no frozen layers. The results obtained in this case are quite better than the ones reported for the Malawi campaign, and this is consistent with the previous comments concerning Fig. 8.

5.5. Results for the Mozambique dataset using VGG16

The VGG16 results for the Mozambique dataset are reported on the right of Fig. 8. The results seem generally similar to the ones provided by the GoogLeNet model, even though in this case more oscillations are denoted changing the number of frozen layers. Nonetheless, the performances tend to decrease when the number of frozen layers increases. The recall values for each crop type are reported in the confusion matrix shown on the right of Fig. 10. The values related to the model with highest overall accuracies have been chosen, which correspond to the one with no frozen layers.

5.6. Results without transfer learning

An important question that is worth to address is to which extent results presented in the previous subsections are obtained due to knowledge transferred from computer vision. To answer this question, we trained DNN from scratch. The same architectures, namely GoogLeNet and VGG16, were used and trained using the same learning hyper-parameters to avoid any bias in the methodology. The only difference was that the networks were initialized with random weights. Results obtained in this case were always random considering Malawi and Mozambique datasets. Inspection of learning curves reveals that the networks were not able to learn features during all the training epochs.

6. Discussion

The results show that the proposed approach is meaningful and reliable while simplifying the analysis design since pre-existing models are used. This research demonstrates that transfer learning from computer vision works when applied to crop type mapping basing on drone images. This means, that some knowledge learned for classification of computer vision images can be useful for discrimination of crop types in drone data. Moreover, having shown that using no frozen layers leads to the best classification results, future analyses can take this finding as input to further simplify the overall classification effort.

It is noteworthy that no parametric analysis has been performed for the choice of the hyper-parameters such as the learning rate as it was beyond the scope of this work. However, opportunely changing these parameters can lead to further improvements of the classification results. It is also interesting that the reported results are in line with other similar ones reported in literature, as one can see comparing current findings with (Chew et al., 2020).

The critical comparison between GoogLeNet and VGG16 is beyond the goal of this paper. Using fixed hyper parameters, the results demonstrate similar performances for the two models, which is most probably associated with the deep structure of both GoogLeNet and

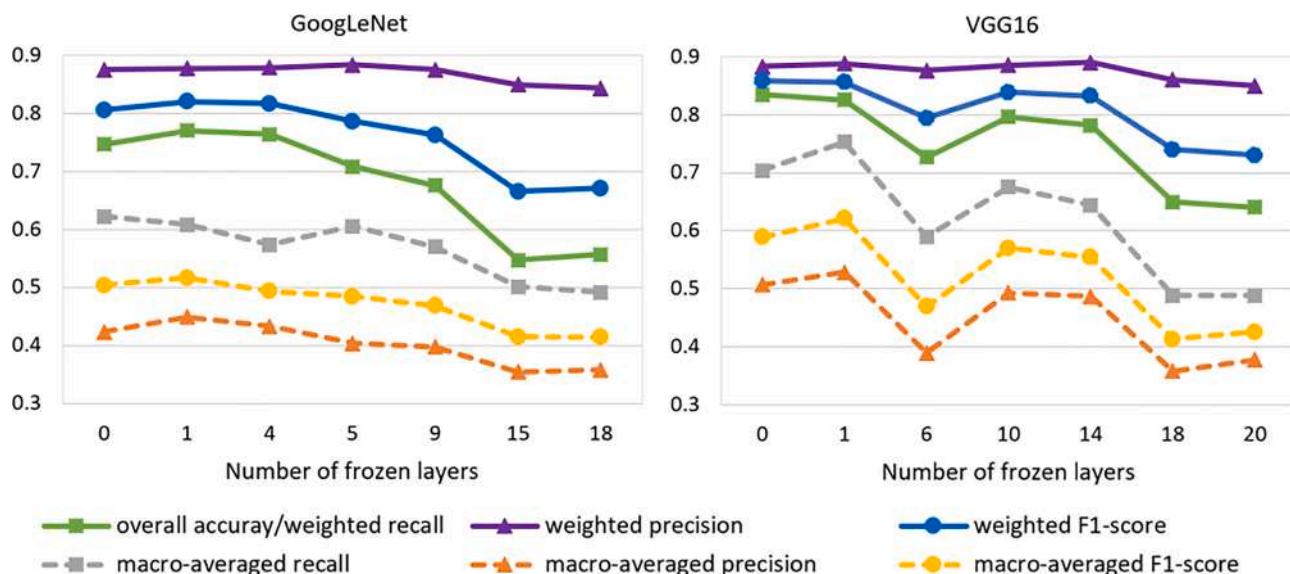


Fig. 7. Validation results vs. number of frozen layers for the Malawi dataset and two considered architectures: GoogLeNet (left) and VGG16 (right).

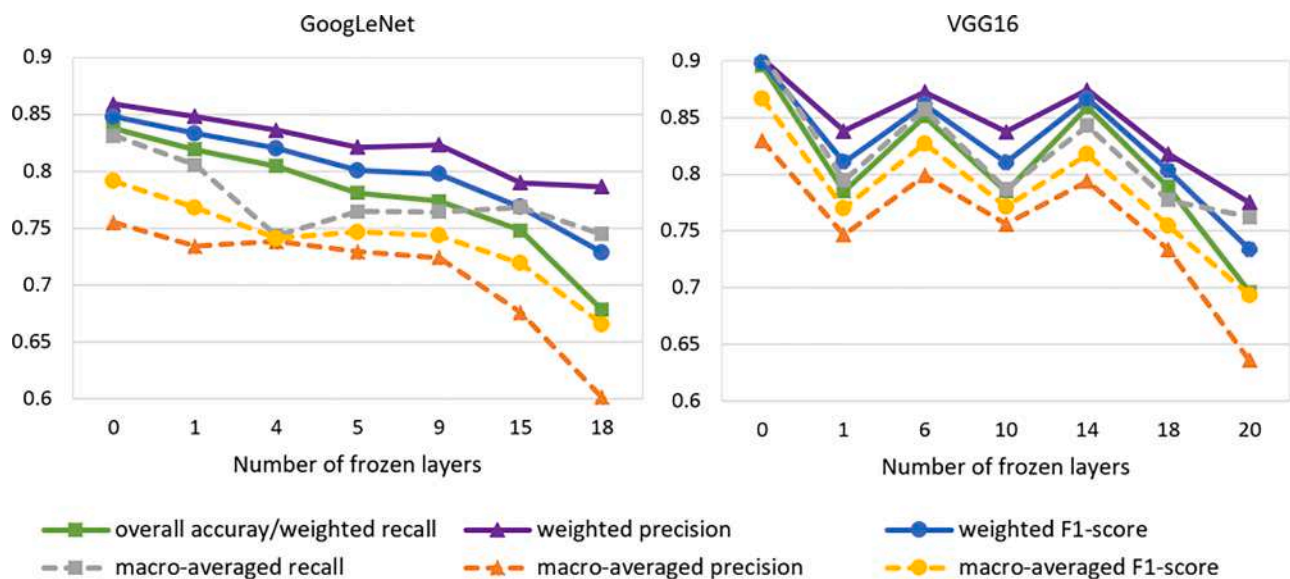


Fig. 8. Validation results vs. number of frozen layers for the Mozambique dataset and two considered architectures: GoogLeNet (left) and VGG16 (right).

VGG16. The two models can successfully transfer the knowledge acquired from computer vision tasks to crop-type mapping, needing only a fine-tuning of the parameters to adapt to the new problem. However, there is no evidence that varying the values of the hyper-parameters would not change the resulting validation scores. Consequently, the results only suggest that using common values of the hyper-parameters the two proposed models work similarly.

This study demonstrates that one of the usual limitations of remote sensing, i.e. the availability of reduced datasets, can be overcome by using transfer learning. Indeed, reduced datasets cannot often be used with deep learning approaches since overfitting problem arise. However, using the fine-tuning optimization of the weights associated to the transfer learning approach seems to solve this problem. Only a reduced number of optimization epochs are required to properly tune the weights to the crop type mapping problem. Overfitting is limited because the weights are locally optimized around the guess values provided by the pre-elaborated tuning performed on the huge computer vision dataset. As can be noted comparing the Malawi and Mozambique

results, the input data quality and distribution strongly affect the classification ability of the models. Referring to Fig. 3, the recommendation for future drone campaigns is to be very careful on guarantying uniform lighting conditions, i.e. avoid mixing images captured during the morning and during the afternoon. Moreover, the distribution of samples per class should be as uniform as possible so that the classifier can achieve good performances for each of the classes. Indeed, even though the training dataset has been composed with the same number of samples for each class, samples coming from most represented classes are more heterogeneous than samples coming from less represented classes. This means that the former contains a greater variety of different samples within class examples than the latter. Hence, the generalization ability of the model is greater for most represented classes (such as maize) than for less represented classes (such as sweet potatoes for the Malawi campaign), as demonstrated by the results obtained on the test dataset on both campaigns. This problem can only be slightly overcome by the data augmentation strategy, as it can only be fully solved with new independent input data containing higher variety of fields for

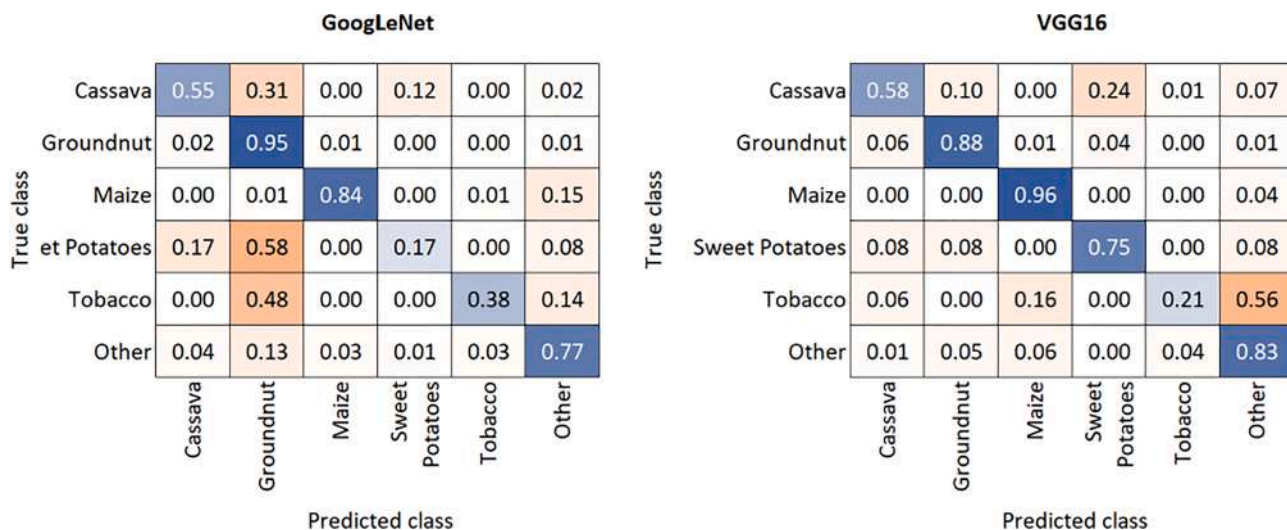


Fig. 9. Confusion matrices for the Malawi dataset and the two considered architectures: GoogLeNet (left) and VGG16 (right).

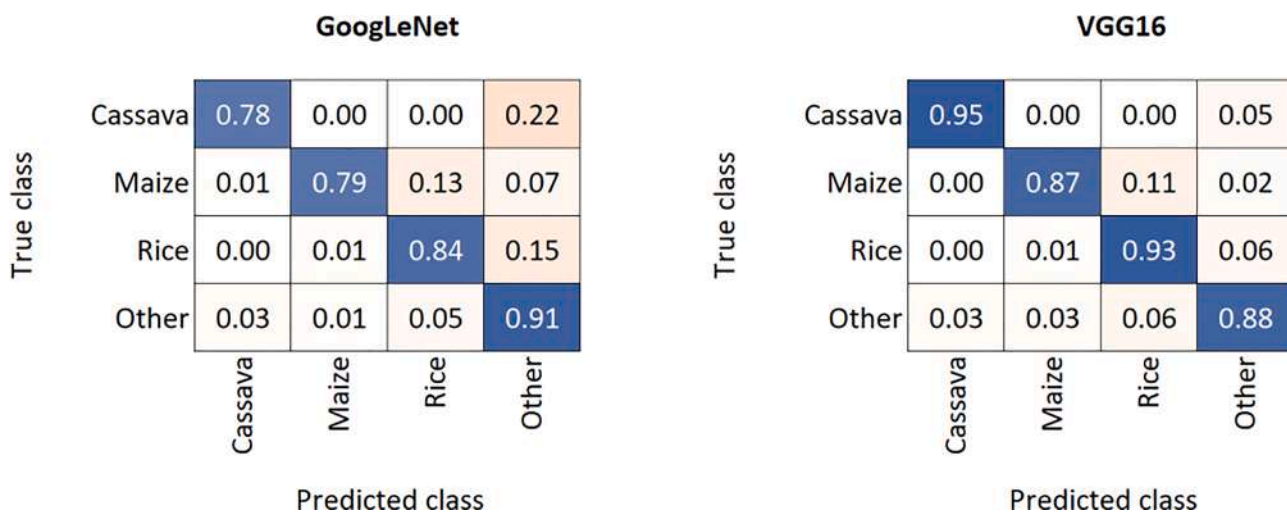


Fig. 10. Confusion matrices for the Mozambique dataset and two considered architectures: GoogLeNet (left) and VGG16 (right).

minority classes.

To conclude, it is relevant to comment on one of the most significant aspects when dealing with SML and crop mapping, which is the model’s generalization ability to adapt to new, previously unseen data. Indeed, it is of the utmost interest to know how the proposed TL approach classifies new data related to different geographical areas or temporal intervals. Differently from other classification problems, this concern is not fully addressed via classical (e.g. stratified) training, validation, and testing splitting. Not even a standard k-fold cross validation would properly answer to the question. As other classification problems in the remote sensing domain, crop mapping is affected by the spatial correlation of the data used for training and testing. The same crop type can have different spectral signature when looking at far away regions or fixed regions over different time intervals (Pohjankukka et al., 2017).

In this paper, a fixed test dataset has been chosen in order to have a distribution of the features similar to the one of the training dataset. This is what is commonly done in machine learning problems, where you usually want the test and the train dataset to have similar distributions of the features. Moreover, to make comparable experiments for different number of frozen layers, the same distributions of features in train and test datasets is needed. This way, it is easier to avoid better results due to better model performances for some specific classes which can be over-represented in testing set. Nonetheless, the test dataset has been

identified with four mosaics which are spatially separated from the train dataset, thus reducing the risk of over-estimating the model performances due to spatial correlation. Better analysis can be carried out to tackle in detail the spatial correlation related to crop mapping (Airolo et al., 2019), but this goes beyond the present objective of this work.

7. Conclusion

This research focused on the application of transfer learning from computer vision data to crop type mapping using drone images as input data. The paper reports and discusses the successful application of the proposed approach. The values of overall accuracy and weighted F1, both greater than 0.85 for some numerical experiments, prove that the proposed approach is promising and can be used for practical application concerning crop type mapping taking advantages of the benefits of deep CNN. The important aspect denoted by the authors is that non-expert users can use the discussed transfer learning approach without needing a deep insight into the properties and characteristics of the predicting model.

CRedit authorship contribution statement

Artur Nowakowski: Conceptualization, Methodology, Software,

Validation, Investigation, Writing - original draft, Writing - review & editing. **John Mrziglod**: Conceptualization, Methodology, Investigation. **Dario Spiller**: Formal analysis, Writing - original draft, Writing - review & editing. **Rogério Bonifacio**: Resources, Supervision, Writing - review & editing. **Irene Ferrari**: Resources, Data curation. **Pierre Philippe Mathieu**: Supervision. **Manuel Garcia-Herranz**: Resources, Supervision. **Do-Hyung Kim**: Resources, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Abdullahi, H.S., Mahieddine, F., Sheriff, R.E., 2015. Technology impact on agricultural productivity: A review of precision agriculture using unmanned aerial vehicles. Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICTS. https://doi.org/10.1007/978-3-319-25479-1_29.
- Airola, A., Pohjankukka, J., Torppa, J., Middleton, M., Nykänen, V., Heikkonen, J., Pahikkala, T., 2019. The spatial leave-pair-out cross-validation method for reliable AUC estimation of spatial classifiers. *Data Min. Knowl. Disc.* <https://doi.org/10.1007/s10618-018-00607-x>.
- Bargoti, S., Underwood, J., 2017. Deep fruit detection in orchards. *Proceedings - IEEE International Conference on Robotics and Automation*. <https://doi.org/10.1109/ICRA.2017.7989417>.
- Barsi, Kugler, Z., László, I., Szabó, G., Abdulmutalib, H.M., 2018. Accuracy dimensions in remote sensing. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. <https://doi.org/10.5194/isprs-archives-XLII-3-61-2018>.
- Berni, J.A.J., Zarco-Tejada, P.J., Suárez, L., Fereres, E., 2009. Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle. *IEEE Trans. Geosci. Remote Sens.* <https://doi.org/10.1109/TGRS.2008.2010457>.
- Bosilj, P., Aptoula, E., Duckett, T., Cielnik, G., 2019. Transfer learning between crop types for semantic segmentation of crops versus weeds in precision agriculture. *J. Field Rob.* <https://doi.org/10.1002/rob.21869>.
- Boulent, J., Foucher, S., Théau, J., St-Charles, P.L., 2019. Convolutional neural networks for the automatic identification of plant diseases. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2019.00941>.
- Buda, M., Maki, A., Mazurowski, M.A., 2018. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*. <https://doi.org/10.1016/j.neunet.2018.07.011>.
- Swain, C.K., Uz Zaman, Q., 2012. Rice crop monitoring with unmanned helicopter remote sensing images. In: *Remote Sensing of Biomass - Principles and Applications*. InTech. <https://doi.org/10.5772/18004>.
- Chew, R., Rineer, J., Beach, R., O'Neil, M., Ujeneza, N., Lapidus, D., Miano, T., Hegarty-Craver, M., Polly, J., Temple, D.S., 2020. Deep neural networks and transfer learning for food crop identification in UAV images. *Drones*. <https://doi.org/10.3390/drones4010007>.
- De Bie, C.A., Khan, M.R., Toxopeus, A.G., Venus, V., Skidmore, A.K., 2008. Hypertemporal image analysis for crop mapping and change detection. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, Li Fei-Fei, 2010. ImageNet: A large-scale hierarchical image database. 248–255. <https://doi.org/10.1109/cvpr.2009.5206848>.
- Deng, L., 2012. The MNIST database of handwritten digit images for machine learning research. *IEEE Signal Process. Mag.* <https://doi.org/10.1109/MSP.2012.2211477>.
- di Gennaro, S.F., Battiston, E., di Marco, S., Facini, O., Matese, A., Nocentini, M., Palliotti, A., Mugnai, L., 2016. Unmanned Aerial Vehicle (UAV)-based remote sensing to monitor grapevine leaf stripe disease within a vineyard affected by esca complex. *Phytopathologia Mediterranea*. https://doi.org/10.14601/Phytopathol_Mediterr-18312.
- Dijkstra, K., van de Loosdrecht, J., Schomaker, L.R.B., Wiering, M.A., 2019. Centroidnet: A deep neural network for joint object localization and counting. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-030-10997-4_36.
- Du, P., Bai, X., Tan, K., Xue, Z., Samat, A., Xia, J., Liu, W., 2020. Advances of four machine learning methods for spatial data handling: a review. *J. Geovisualization and Spatial Anal.* 4, 13.
- Dyrmann, M., Mortensen, A.K., Midtby, H.S., Jørgensen, R.N., 2016. Pixel-wise classification of weeds and crops in images by using a Fully Convolutional neural network. *International Conference on Agricultural Engineering*.
- Fan, Z., Lu, J., Gong, M., Xie, H., Goodman, E.D., 2018. Automatic tobacco plant detection in UAV images via deep neural networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* <https://doi.org/10.1109/JSTARS.2018.2793849>.
- Foerster, S., Kaden, K., Foerster, M., Itzerott, S., 2012. Crop type mapping using spectral-temporal profiles and phenological information. *Comput. Electron. Agric.* <https://doi.org/10.1016/j.compag.2012.07.015>.
- Gallego, J., Carfagna, E., Baruth, B., 2010. Accuracy, objectivity and efficiency of remote sensing for agricultural statistics. In: *Agricultural Survey Methods*. <https://doi.org/10.1002/9780470665480.ch12>.
- Gumma, M.K., Tsusaka, T.W., Mohammed, I., Chavula, G., Ganga Rao, N.V.P.R., Okori, P., Ojiewo, C.O., Varshney, R., Siambi, M., Whitbread, A., 2019. Monitoring changes in the cultivation of pigeonpea and groundnut in Malawi using time series satellite imagery for sustainable food systems. *Remote Sens.* <https://doi.org/10.3390/rs11121475>.
- Hall, O., Dahlin, S., Marstorp, H., Archila Bustos, M., Öborn, I., Jirstrom, M., 2018. Classification of maize in complex smallholder farming systems using UAV imagery. *Drones*. <https://doi.org/10.3390/drones2030022>.
- Heng, L., Xiao, F., Chao, L., Longguo, L., Naiwen, L., Lei, M., 2018. Land use information quick mapping based on UAV low-altitude remote sensing technology and transfer learning. *Drones - Applications*. <https://doi.org/10.5772/intechopen.74475>.
- Huh, M., Agrawal, P., Efron, A.A., 2016. eprint arXiv:1608.08614. In: *What makes ImageNet good for transfer learning?*.
- Ian Goodfellow, Yoshua Bengio, A.C., 2016. *Deep Learning - Ian Goodfellow, Yoshua Bengio, Aaron Courville - Google Books*. In MIT Press. https://books.google.com.et/books?hl=en&lr=&id=omivDQAAQBAJ&oi=fnd&pg=PR5&dq=I.+Goodfellow,+Y.+Bengio,+and+A.+Courville,+Deep+Learning,+Cambridge,+MA,+USA:+MIT+Press,+2016.&ots=MMV7bonFTY&sig=daQijLM9rMProp8vtV9K9G9fE&redir_esc=y#v=onepage&q=L.Goodfell.
- Jensen, J.R., Garcia-Quijano, M., Hadley, B., Im, J., Wang, Z., Nel, A.L., Teixeira, E., Davis, B.A., 2006. Remote sensing agricultural crop type for sustainable development in South Africa. *Geocarto International*. <https://doi.org/10.1080/10106040608542378>.
- Ji, S., Zhang, C., Xu, A., Shi, Y., Duan, Y., 2018. 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sens.* <https://doi.org/10.3390/rs10010075>.
- Jin, S., Su, Y., Gao, S., Wu, F., Hu, T., Liu, J., Li, W., Wang, D., Chen, S., Jiang, Y., Pang, S., Guo, Q., 2018. Deep learning: Individual maize segmentation from terrestrial lidar data using faster R-CNN and regional growth algorithms. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2018.00866>.
- Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* <https://doi.org/10.1016/j.compag.2018.02.016>.
- Kogan, F., Kussul, N., Adamenko, T., Skakun, S., Kravchenko, O., Kryvobok, O., Shelestov, A., Kolotii, A., Kussul, O., Lavrenyuk, A., 2013. Winter wheat yield forecasting in Ukraine based on Earth observation, meteorological data and biophysical models. *Int. J. Appl. Earth Obs. Geoinf.* <https://doi.org/10.1016/j.jag.2013.01.002>.
- Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Mallocci, M., Duerig, T., Ferrari, V., 2018. The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale.
- Lee, S.H., Chan, C.S., Wilkin, P., Remagnino, P., 2015. Deep-plant: Plant identification with convolutional neural networks. *Proceedings - International Conference on Image Processing, ICIP*. <https://doi.org/10.1109/ICIP.2015.7350839>.
- Liakos, K.G., Busato, P., Moshou, D., Pearson, S., Bochtis, D., 2018. Machine learning in agriculture: A review. *Sensors (Switzerland)*. <https://doi.org/10.3390/s18082674>.
- Lottes, P., Khanna, R., Pfeifer, J., Siegwart, R., Stachniss, C., 2017. UAV-based crop and weed classification for smart farming. *IEEE International Conference on Robotics and Automation (ICRA) 2017*, 3024–3031. <https://doi.org/10.1109/ICRA.2017.7989347>.
- Maktab Dar Oghaz, M., Razaak, M., Kerdegari, H., Argyriou, V., Remagnino, P., 2019. Scene and environment monitoring using aerial imagery and deep learning. *Proceedings - 15th Annual International Conference on Distributed Computing in Sensor Systems, DCOSS 2019*. <https://doi.org/10.1109/DCOSS.2019.00078>.
- Mardanisamani, S., Maleki, F., Hosseinzadeh Kassani, S., Rajapaksa, S., Duddu, H., Wang, M., Shirliffe, S., Ryu, S., Josuttis, A., Zhang, T., Vail, S., Pozniak, C., Parkin, I., Stavness, I., Eramian, M., 2019. June. Crop Lodging Prediction From UAV-Acquired Images of Wheat and Canola Using a DCNN Augmented With Handcrafted Texture Features. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.
- McCabe, M.F., Houborg, R., Lucier, A., 2016. High-resolution sensing for precision agriculture: from Earth-observing satellites to unmanned aerial vehicles. *Remote Sensing for Agriculture, Ecosystems, and Hydrology XVIII*. <https://doi.org/10.1117/12.2241289>.
- Mehdipour Ghazi, M., Yanikoglu, B., Aptoula, E., 2017. Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing*. <https://doi.org/10.1016/j.neucom.2017.01.018>.
- Mohanty, S.P., Hughes, D.P., Salathé, M., 2016. Using deep learning for image-based plant disease detection. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2016.01419>.
- Nijhawan, R., Sharma, H., Sahni, H., Batra, A., 2018. A deep learning hybrid CNN framework approach for vegetation cover mapping using deep features. *Proceedings - 13th International Conference on Signal-Image Technology and Internet-Based Systems, SITIS 2017*. <https://doi.org/10.1109/SITIS.2017.41>.
- Pádua, L., Vanko, J., Hruška, J., Adão, T., Sousa, J.J., Peres, E., Morais, R., 2017. UAS, sensors, and data processing in agroforestry: a review towards practical applications. *Int. J. Remote Sens.* <https://doi.org/10.1080/01431161.2017.1297548>.
- Pan, S.J., Yang, Q., 2010. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* <https://doi.org/10.1109/TKDE.2009.191>.

- Peña, J.M., Torres-Sánchez, J., de Castro, A.I., Kelly, M., López-Granados, F., 2013. Weed mapping in early-season maize fields using object-based analysis of unmanned aerial vehicle (UAV) images. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0077151>.
- Pohjankukka, J., Pahikkala, T., Nevalainen, P., Heikkonen, J., 2017. Estimating the prediction performance of spatial models via spatial k-fold cross validation. *Int. J. Geographical Inf. Sci.* <https://doi.org/10.1080/13658816.2017.1346255>.
- Rahnemoonfar, M., Sheppard, C., 2017. Deep count: Fruit counting based on deep simulated learning. *Sensors (Switzerland)*. <https://doi.org/10.3390/s17040905>.
- Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., Hughes, D.P., 2017. Deep learning for image-based cassava disease detection. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2017.01852>.
- Rebetez, J., Satizábal, H.F., Mota, M., Noll, D., Büchi, L., Wendling, M., Cannelle, B., Perez-Uribe, A., Burgos, S., 2016. Augmenting a convolutional neural network with local histograms - A case study in crop classification from high-resolution UAV imagery. *ESANN 2016–24th European Symposium on Artificial Neural Networks*.
- Reyes, A.K., Caicedo, J.C., Camargo, J.E., 2015. Fine-tuning deep convolutional networks for plant recognition. *CEUR Workshop Proceedings*.
- Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., McCool, C., 2016. Deepfruits: A fruit detection system using deep neural networks. *Sensors (Switzerland)*. <https://doi.org/10.3390/s16081222>.
- Senthilnath, J., Kandukuri, M., Dokania, A., Ramesh, K.N., 2017. Application of UAV imaging platform for vegetation analysis based on spectral-spatial methods. *Comput. Electron. Agric.* <https://doi.org/10.1016/j.compag.2017.05.027>.
- Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. <https://doi.org/10.1109/CVPR.2015.7298594>.
- Tri, N.C., Van Hoai, T., Duong, H.N., Trong, N.T., Van Vinh, V., Snasel, V., 2017. A novel framework based on deep learning and unmanned aerial vehicles to assess the quality of rice fields. *Adv. Intelligent Syst. Comput.* https://doi.org/10.1007/978-3-319-49073-1_11.
- Valente, J., Doldersum, M., Roers, C., Kooistra, L., 2019. Detecting rumex obtusifolius weed plants in grasslands from UAV RGB imagery using deep learning. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. <https://doi.org/10.5194/isprs-annals-IV-2-W5-179-2019>.
- Wahab, I., Hall, O., Jirstrom, M., 2018. Remote sensing of yields: application of UAV imagery-derived NDVI for estimating maize vigor and yields in complex farming systems in Sub-Saharan Africa. *Drones*. <https://doi.org/10.3390/drones2030028>.
- Wójtowicz, M., Wójtowicz, A., Piekarczyk, J., 2016. Application of remote sensing methods in agriculture. *Communications in Biometry and Crop Science*.
- Xie, M., Jean, N., Burke, M., Lobell, D., Ermon, S., 2016. Transfer learning from deep features for remote sensing and poverty mapping. *30th AAAI Conference on Artificial Intelligence, AAAI 2016*.
- Yalcin, H., 2017. Plant phenology recognition using deep learning: Deep-Pheno. *2017 6th International Conference on Agro-Geoinformatics, Agro-Geoinformatics 2017*. <https://doi.org/10.1109/Agro-Geoinformatics.2017.8046996>.
- Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., Xu, B., Yang, X., Zhu, D., Zhang, X., Zhang, R., Feng, H., Zhao, X., Li, Z., Li, H., Yang, H., 2017. Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2017.01111>.
- Yin, Z., Williams, T., 1997. Obtaining spatial and temporal vegetation data from landsat MSS and AVHRR/NOAA satellite images for a hydrologic model. *Photogramm. Eng. Remote Sens.* 63.
- Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? *Adv. Neural Inf. Process. Syst.*
- Yu, X., Wu, X., Luo, C., Ren, P., 2017. Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework. *GIScience Remote Sens.* <https://doi.org/10.1080/15481603.2017.1323377>.