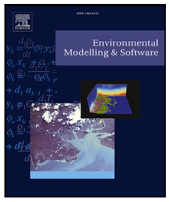
Environmental Modelling and Software 172 (2024) 105931

Contents lists available at ScienceDirect

Environmental Modelling and Software

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



Automatic classification of land cover from LUCAS in-situ landscape photos using semantic segmentation and a Random Forest model

Laura Martinez-Sanchez [a,*], Linda See [b], Momchil Yordanov [a], Astrid Verhegghen [a], Neija Elvekjaer [a], Davide Muraro [a], Raphaël d'Andrimont [a], Marijn van der Velde [a]

[a] European Commission, Joint Research Centre (JRC), Ispra, Italy
[b] International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

ARTICLE INFO

Dataset link: http://data.europa.eu/89h/c6166c60-5221-437b-87ed-3aaec123801f

Keywords:
Land cover
LUCAS
Semantic segmentation
Random Forests
Deep learning
Landscape elements

ABSTRACT

Spatially explicit information on land cover (LC) is commonly derived using remote sensing, but the lack of training data still remains a major challenge for producing accurate LC products. Here, we develop a computer vision methodology to extract LC information from photos from the Land Use-Land Cover Area Frame Survey (LUCAS). Given the large number of photographs available and the comprehensive spatial coverage, the objective is to show how the automatic classification of photos could be used to develop reference data sets for training and validation of LC products as well as other purposes. We first selected a representative sample of 1120 photos covering eight major LC types across the European Union. We then applied semantic segmentation to these photos using a neural network (Deeplabv3+) trained with the ADE20k dataset. For each photo, we extracted the original LC identified by the LUCAS surveyor, the segmented objects, and the pixel count for each ADE20k class. Using the latter as input features, we then trained a Random Forest model to classify the LC of the photo. Examining the relationship between the objects/features extracted by Deeplabv3+ and the LC labels provided by the LUCAS surveyors demonstrated how the LC classes can be decomposed into multiple objects, highlighting the complexity of LC classification from photographs. The results of the classification show a mean F1 Score of 89%, increasing to 93% when the Wetland class is not considered. Based on these results, this approach holds promise for the automated retrieval of LC information from the rich source of LUCAS photographs as well as the increasing number of geo-referenced photos now becoming available through social media and sites like Mapillary or Google Street View.

1. Introduction

Over the past decade, there has been a proliferation of satellite based land cover (LC) maps produced, from global products with multiple classes (Bontemps et al., 2013; Buchhorn et al., 2020) to binary time series focused on different thematic areas such as tree cover, cropland, water bodies and built-up surfaces (Corbane et al., 2019; Hansen et al., 2013; Pekel et al., 2016; Potapov et al., 2022). These advances have been largely driven by the opening of the Landsat archive, the availability of new high-resolution satellite imagery (e.g. from Sentinel), as well as the new cloud-based computing environments and machine learning routines.

A fundamental input to LC map production is the reference data needed to both train the classification algorithms, and validate the resulting layers using statistically robust accuracy assessment (Stehman and Foody, 2019). In the past, LC maps were trained and validated using in-situ or field-based data, but there are substantial costs involved in the data collection, particularly for mapping large areas (Szantoi et al., 2020). Moreover, machine learning algorithms, and in particular, newer deep learning methods, need large quantities of high-quality training data regardless of the specific algorithm used (Maxwell et al., 2018). The use of various semi-supervised learning algorithms has been one approach to addressing the lack of training data (Padmanaba et al., 2013) but this method still requires a good set of basic reference data to train the algorithms. Transfer learning algorithms can use data from different domains to compensate for a lack of training data from the direct domain of interest (Weiss et al., 2016) and have shown promise in land cover classification by using existing pre-trained deep learning networks (Alem and Kumar, 2022). An entirely different approach has been towards increasing the reference database using visual interpretation of satellite imagery, e.g., from very high-resolution imagery available from Google Earth and Microsoft Bing Maps (Waldner et al., 2019), using crowdsourcing through applications

* Corresponding author.
E-mail address: laura.martinez-sanchez@ec.europa.eu (L. Martinez-Sanchez).

https://doi.org/10.1016/j.envsoft.2023.105931
Received 7 March 2023; Received in revised form 11 December 2023; Accepted 12 December 2023
Available online 16 December 2023

such as Geo-Wiki (See et al., 2022; Radoux et al., 2014). However, there is uncertainty related to the imagery dates, geolocation errors, the quality of the crowdsourced data, and the accuracy of the LC maps sampled using these approaches.

More recently, street level imagery (SLI) from Google StreetView (GSV), Mapillary, Baidu, etc., as well as geo-tagged photographs from photo sharing sites such as Flickr, are being used as sources of reference data in many different types of applications. Some of these have involved the implementation of virtual street level surveys, e.g., to complement field-based street tree surveys (Berland and Lange, 2017), as a potential source of in-situ data for crop monitoring (d'Andrimont et al., 2018), to audit neighborhood and built environments (Rundle et al., 2011; Kelly et al., 2013), as inputs to models of mobility patterns in cities (Goel et al., 2018) and to fill in missing sidewalks in aerial images (Ning et al., 2022). However, more recent applications have generally entailed the use of some type of automated approach to first classify the photographs, extract information, and then apply other machine learning algorithms to predict specific features of interest, e.g., the prediction of house prices using features extracted from Flickr photographs (Chen et al., 2022b), the prediction of various socio-economic characteristics such as income and voting patterns from GSV imagery (Gebru et al., 2017), and the development of a visual walkability index using features extracted from Baidu Map Street View (Zhou et al., 2019a). There have also been studies in which combinations of RGB bands have been used to create GSV difference images in order to extract the area of vegetation and calculate a Green View Index (Li et al., 2015), which has then been combined with a survey on physical activity to understand the influence of street greenery (Lu, 2019).

Two main developments have aided the research in this area. The first comes from the field of computer vision where there have been rapid advances in extracting information from photographs using approaches such as scene recognition for a single class, object detection for extraction of individual objects in a photograph, and semantic segmentation, which assigns a class to each pixel (Neuhold et al., 2017). As these approaches require large training datasets, the second main development has been in the availability of substantial labeled image data sets, e.g., ImageNet (Russakovsky et al., 2015), COCO (Lin et al., 2014), ADE20K (Zhou et al., 2019b) and Mapillary's Vistas dataset (Neuhold et al., 2017), as well as deep learning networks that have already been trained on these datasets such as VGG (Chen et al., 2017), ResNet (He et al., 2016) and DeepLabv3+ (Chen et al., 2017), or more recently, visual transformers (Chen et al., 2022a) or InternImage, a new large-scale CNN-based network (Wang et al., 2022), among others.

In the area of land use mapping, a number of studies have used pre-trained networks with existing image databases to classify SLI for building types or LU (rather than LC), often demonstrated on small urban areas. For example, Kang et al. (2018) used VGG16 trained on the Places2 dataset to remove images from GSV that did not contain buildings of interest. They then used four different networks (AlexNet, ResNet18, ResNet34, VGG16) pre-trained on ImageNet to classify different building types, achieving high accuracy for some classes such as office buildings (85%) and garages (99%). Zhu and Newsam (2015) used a linear support vector machine (SVM) fed with all pixels from Flickr photographs to predict 8 LU classes, achieving an overall accuracy of 76%. In a subsequent paper, they extended this approach to 45 classes for the city of San Francisco using a scene and object-based recognition approach, but they achieved a lower accuracy of 46.7% for this finer-grained solution (Zhu et al., 2019). Cao et al. (2018) combined aerial and SLI to produce a LU map with 11 classes in an area of New York City. They used the Places-CNN trained on the Places365 dataset (with 10 million images) to extract features from four GSV images at each location, reduced the data using Principal Components Analysis and then combined these features with aerial images using SegNet. They achieved an overall accuracy of 78% and showed that the use of SLI improved the classification over using aerial imagery alone. Similarly, overhead imagery in combination with SLI improved the identification of urban objects from SLI taken from OpenStreetMap (Srivastava et al., 2019, 2020). Using VGG16 to extract feature vectors from each photograph, they then applied these to CNN models to predict the land use type, obtaining accuracies of between 41 to 62%. In a slightly different application, features were extracted from Google Street View images using both Places-CNN and DeepLab-v3+ trained with the Places365 and Cityscapes data sets, respectively, which were then input to a model to map Local Climate Zones, which have ten detailed urban land use classes (Cao et al., 2023).

Other than urban land use types, little other work has been carried out for other LC classes or related variables. One exception is the study by Xu et al. (2017), where the pre-trained CNN model Inception-v3 was trained and validated using photographs from the Global Geo-Referenced Field Photo Library to identify 11 different LC types. The overall accuracy varied from 48.4 to 73.6% depending on the probability threshold chosen. Other notable exceptions include the prediction of crop type and phenology with Mobilenetv2 trained with bespoke SLI collected in the Netherlands (d'Andrimont et al., 2022), the mapping of cherry blossoms using Mapillary images classified using YOLOv4 (Funada and Tsutsumida, 2022) and a tree cover index (for urban streets in the city of Cardiff, UK), in which vegetation was first identified in GSV images using thresholding followed by semantic segmentation using the PSPNet. Here the use of semantic segmentation was critical for providing context and reducing the mismatch that occurs with the use of object detection (Stubbings et al., 2019). A similar approach was used to segment SLI for assessing street greenery in the city of Nanjing, China, using multiple indicators including a green view index, NDVI and an indicator related to species diversity (Tong et al., 2020).

Another valuable database of georeferenced photographs is from LUCAS (Land Use/Cover Area Frame Survey), which takes place every three years across European Union (EU) member states since 2006 (d'Andrimont et al., 2020). Expert surveyors document each location, which is a systematic sample, using harmonized LC and LU protocols and take a set of in-situ photographs of each location. LUCAS was designed for monitoring changes in LC and LU related to EU policy but has also been applied in combination with Corine Land Cover (CLC) to provide unbiased area estimates (Gallego and Bamps, 2008) and for monitoring landscape diversity across Europe (Palmieri et al., 2011). In terms of remote sensing, LUCAS has been used to develop a high resolution LC and LU map for Germany (Mack et al., 2017), validate the Greek national LC map (Karydas et al., 2015) and three global LC products over Europe (Gao et al., 2020), and has been used as both training and validation for a pan-European Landsat-based LC map (Pflugmacher et al., 2019) and for Sentinel-2 LC classification (Weigand et al., 2020). More recently, LUCAS photographs were classified using a deep learning approach in combination with Sentinel 1 and 2 imagery to develop a grassland management intensity map (Saadeldin et al., 2022) while LUCAS and Flickr photos were used to verify two land cover products, where the images were first classified using the Nature Scene Image Classification model, based on the GoogLeNet Inception network (Cui et al., 2022).

One of the advantages of LUCAS photographs for LC classification is that they have been taken at each point location and in the four cardinal directions away from the point by a trained surveyor, who has labeled the LC and LU of the point location; hence, they represent high quality ground truth information. Moreover, they are not restricted to streets because the sample has been systematically generated and therefore does not have the same bias as SLI. At the same time, visual interpretation of LC from photographs is challenging because the LC at the location of the observer is not necessarily the LC shown in the photograph (for example, objects may obstruct the view). Similarly, the estimation of openness or closeness of the LC depends on the view seen from the landscape photo, which is made from the surface plane and viewed from an oblique angle with a focus on the often-distant horizon.

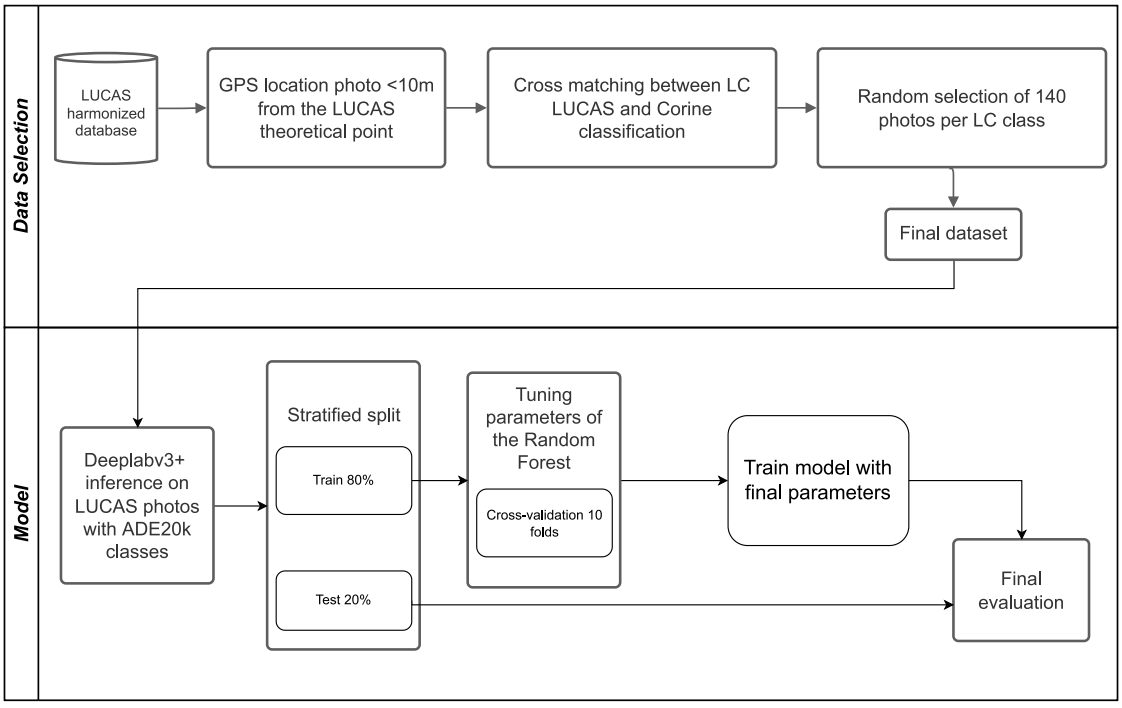


Fig. 1. Schematic of the overall workflow of the study including data selection and model development.

Other issues to consider include the scale of the observation, the complexity of landscapes and the elements that they comprise, as well as the required granularity of the LC classification. These challenges may be why there has been little research to date in extracting information directly from LUCAS photographs. Using techniques from computer vision such as semantic segmentation to classify visible elements, we can determine which of these are important for LC classification, and can develop LC classification models. Importantly, in such a process, the strengths and limitations of this type of in-situ based classification, and the eventual synergies with EO-based classification, can be explored.

Hence, the overall aim of this study is to assess the extent to which semantic segmentation of in-situ LUCAS photos to extract relevant LC-related variables can be used to train a machine learning algorithm to classify LC. We first select a representative sample of photographs from the EU-wide LUCAS survey and apply semantic segmentation using DeepLabv3+ (Chen et al., 2017) trained with the ADE20K dataset (Zhou et al., 2019b). We then compare the dominant segmented objects predicted in each photograph with the original LUCAS LC classification to ensure the feasibility of this approach. Finally, we use a Random Forest (RF) classifier to predict LUCAS LC classes using the segmented objects as input features, with a discussion of the limitations and potential of such an approach for LC mapping in the future.

1.1. Objectives

This study assesses to what extent semantic segmentation of relevant LC related variables can be combined with RF to classify LC from in-situ LUCAS landscape photos for the purpose of generating reference data for training and validation as well as other applications. The detailed objectives are:

1. To select a representative set of photos from the EU-wide LUCAS survey.
2. To semantically segment the LUCAS photos with DeepLabv3+ (Chen et al., 2017) using the ADE20K dataset (Zhou et al., 2017, 2019b).
3. To evaluate the association between the dominant DeepLab/ADE20K predicted class in each photo and the original LUCAS LC classification.
4. To predict LUCAS LC using RF and the distribution of segmented objects on the LUCAS photos.
5. To discuss the limitations and the potential of the proposed methodology.

2. Materials and methods

In this section, we describe (1) the LUCAS survey and the sampling of photos for this study; (2) the DeepLab/ADE20K driven semantic segmentation; (3) the subsequent classification via the RF model; and (4) the accuracy metrics used to evaluate the classification.

The selection of the LUCAS point photos used in this study and the overall workflow are provided in Fig. 1.

2.1. LUCAS survey

Our reference in-situ data set is based on the data and photos collected during the LUCAS 2018 survey (Eurostat, 2018a). At each surveyed LUCAS point, observations have been made on LC and other variables, and photographs have been taken. LUCAS is a two phase sample survey, where the first phase draws theoretical points on a grid with a 2-km systematic grid covering the whole of the EU's territory. Each theoretical point is then classified into one of 10 LC types from the C3-Classification (Eurostat, 2018a) using visual interpretation of orthophotos or satellite images (Eurostat, 2018b). The second phase is then carried out by sampling from those points identified in the first phase in order to obtain a statistically representative spatial distribution of the main LC/LU (land use) classes, which have been surveyed in-situ during the 2018 LUCAS campaign. When each point was then reached by the surveyor, the LC/LU and other variables were observed and recorded using the C3-Classification (Eurostat, 2018a). Photographs were then taken in the four cardinal directions away from the point, while a

fifth photo was taken of the location of the LUCAS theoretical point for which the observations were made, which represents the LC/LU recorded. The photos of the point must also facilitate the process of finding the exact location of the observed point in the next LUCAS survey. Therefore, a LUCAS marker (e.g., an identifiable object such a frisbee, flag, etc.) is placed at the exact location of the point, and if possible, the photo should contain stable field elements (e.g., a house, barn, track or any other "quasi-stable" landmark). In case the point is not reachable, the LUCAS marker should not be used but the photo should be taken in the direction of the point. Moreover, the photo must be taken in a non-tilted landscape format with the point in the center, when possible, and the horizon should be about 5/6 of the way up the viewfinder. The location of the position from which the photos are taken is geolocated via a GNSS receiver. Note that these coordinates may differ from the coordinates of the theoretical location because of accessibility issues or location measurement inaccuracies.

Since a point has neither width nor length and considering that the LC must be classified at the theoretical point by the surveyor, the observation is actually made for a circle with a 1.5 m radius, representing an area of about 7 m^2. For the vast majority of points, this definition is easily applied since the area to classify is fully homogeneous in terms of LC/LU. Nevertheless, there may be situations in which the location of the point and/or the observation of the LC/LU are ambiguous. In these cases, the observation window is extended to 20 m, and the LUCAS photo taken by the surveyors at that location should represent the LC/LU observed.

The LUCAS 2018 survey sampled a total of 337,845 LUCAS points pictures, out of which approximately 240,000 points were visited in the field by surveyors.

2.1.1. Selection of LUCAS point photographs for this study

The identification of the LC class is undertaken using the first level (or most generic LC description) from the LUCAS legend, A: Artificial Land, B: Cropland, C: Woodland, D: Shrubland, E: Grassland, F: Bare soil and Lichens, G: Water Areas, H: Wetlands; see Fig. 2 for examples of photographs from each of these classes. To maximize the representativeness of the point photos in covering the observed LC classes, a selection of the LUCAS point photos was made following the cascading specifications set out in the LUCAS 2018 protocol. These specifications prescribe the type of observation that can be made (Eurostat, 2018a); here we consider only the first specification in which the in situ observation is made at a distance < 100 m from the theoretical LUCAS point.

Since the LC information associated with each photo is the reference for this analysis, we implemented several steps to ensure that the photos accurately represent the LC classified by the surveyor. First, we only used LUCAS point photos from in-situ observations that were made at a distance of less than 100 m from the theoretical point to ensure that the photos accurately represent the LC/LU classification. Secondly, we extracted the CLC (European Union, Copernicus Land Monitoring Service 2018. European Environment Agency , EEA) classification at each point. The CLC is based on detailed ortho-imagery and should be representative for a surface area of 100 m^2. By crossing the LUCAS and CLC classes (see Suppl. Fig. 1), and by selecting those points where the LC classes matched, we want to increase the spatial representativeness and verify the thematic LC information embedded in the LUCAS point photo. In a third step, from the photos that remained, we randomly selected and verified 140 LUCAS photos for each LC class. In total this was 1120 photos, which was empirically estimated as a good number as it is still possible to assess this number visually but nevertheless large enough to draw conclusions. Photos were not selected if they were (1) obstructed by an object (e.g., a lamppost); (2) covered by large patches that were anonymized; and (3) not corresponding to the LUCAS LC classification (e.g., a photo of a lawn for a point classified as artificial). Examples of these different cases are shown in Fig. 3. The 140 selected LUCAS point photos for each LC class were then semantically segmented with DeepLab/ADE20k.

2.2. Semantic image segmentation inference and evaluation

To classify the LC from the LUCAS point photos, we first applied the semantic segmentation to divide the picture into segments where each image pixel is mapped to an object. Training a semantic segmentation model requires a large amount of images with each pixel labeled with a class. Since this process is very time consuming, we opted to use an already pre-trained semantic segmentation model. We chose DeepLabV3 (Chen et al., 2017), which is a semantic segmentation architecture where the encoder is composed of a ResNet to extract features and Atrous Spatial Pyramid Pooling (ASPP) to extract feature information at multiple scales. This enhances the prediction accuracy and boundary information of the semantic segmentation. The decoder is implemented using a combination of low-level and high-level features. High-level features are first up-sampled by a factor of 4 and concatenated with the low-level feature from the ResNet structure. Before the concatenation, the channels of the low-level features are reduced with a convolutional layer of 1×1. To obtain the final segmentation, the features are refined with 3×3 convolutions and a final bilinear up-sampling, again by a factor of 4 (see Fig. 4). In our case, we used a Split-Attention Network, which is a new ResNet variant that significantly boosts the performance of this model (Zhang et al., 2020).

In this paper, we used an implementation of the DeepLabv3+ done by Gluoncv (Guo et al., 2020). Gluoncv is a toolkit based on Apache MXNet for deep learning processes. Gluoncv allows the use of a specific architecture and pre-trained model on a specific dataset to obtain the weights of the already trained model to do the inference. Our interest here is to describe the full landscape to subsequently extract LC information. The most suitable dataset for this task is ADE20k (Zhou et al., 2017, 2019b). ADE20k is a semantic segmentation dataset containing more than 20K scene-centric images, extracted from public databases like SUN or Places, exhaustively annotated with pixel-level objects and object part labels. There are 150 semantic categories, which include classes relevant for this study such as field, grass, tree, earth, etc. Hence, we used, in inference mode, Deeplabv3+ pre-trained on ADE20k to extract classes that describe the landscape in the LUCAS photos. Therefore, after applying semantic segmentation to the LUCAS photos selected, each pixel of each photo is mapped to a legend value belonging to one ADE20k class. Since no actual training was done using the LUCAS photos, no metrics were extracted at this point.

A first level evaluation of the semantic segmentation was done by extracting the presence/absence of all ADE20k classes for each photo. The cumulative addition of the number of pixels belonging to the same ADE20k class in a photo will be referred to henceforth as the *features*. We then deleted the classes from ADE20k that were not represented and therefore absent in the LUCAS dataset and assessed the distribution of the dominant features against the in-situ classified LUCAS LC classes (see Fig. 7). Based on this, we were able to visually assess if there is a clear correspondence between the ADE20k features and the LUCAS LC classes.

2.3. LC prediction with a random forest model

RF are a general term for ensemble methods that use tree-type classifiers and are known to be robust against multi-collinearity and overfitting (Breiman, 2001). The RF algorithm creates multiple decision trees (i.e., the number of estimators) where each tree in the ensemble is built from the original training data or from a bootstrapped sample. In each decision tree, the best node-splitting is done with a random subset of the features. RF achieve a reduced variance by combining diverse trees, sometimes at the cost of a slight increase in the bias. In practice, the variance reduction is often significant and overall it yields a better model. To improve the accuracy of the RF model, we ran a grid search

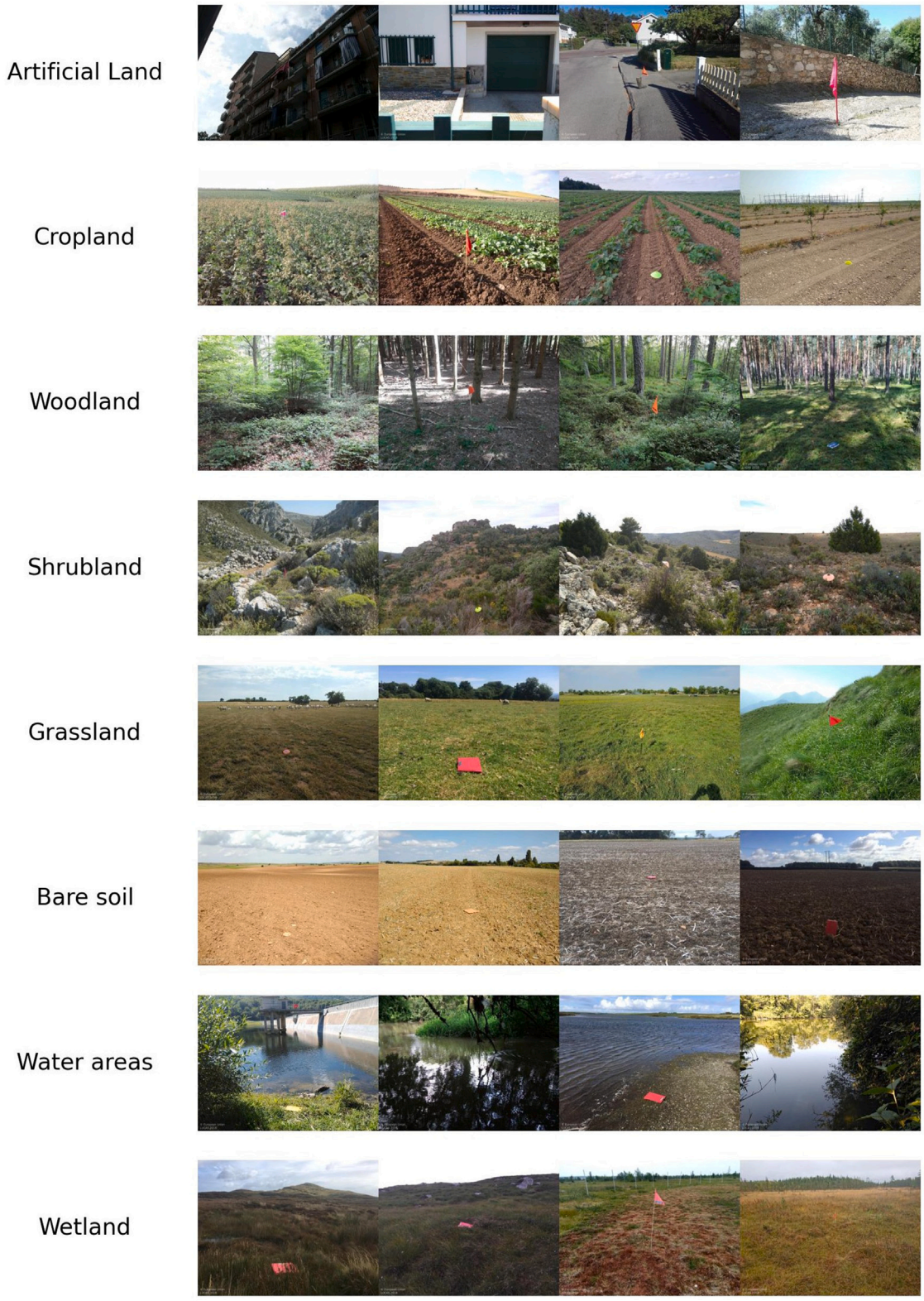
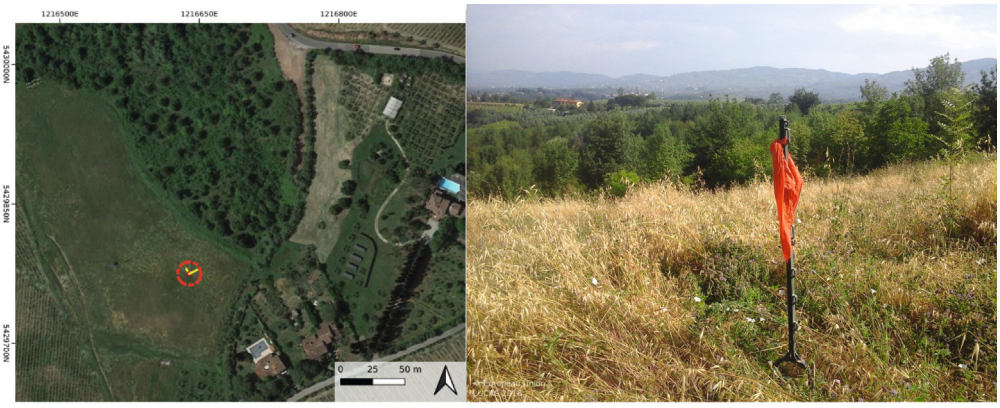


Fig. 2. Random examples of LUCAS point photos for all level-1 LUCAS classes, which are indicated on the left.

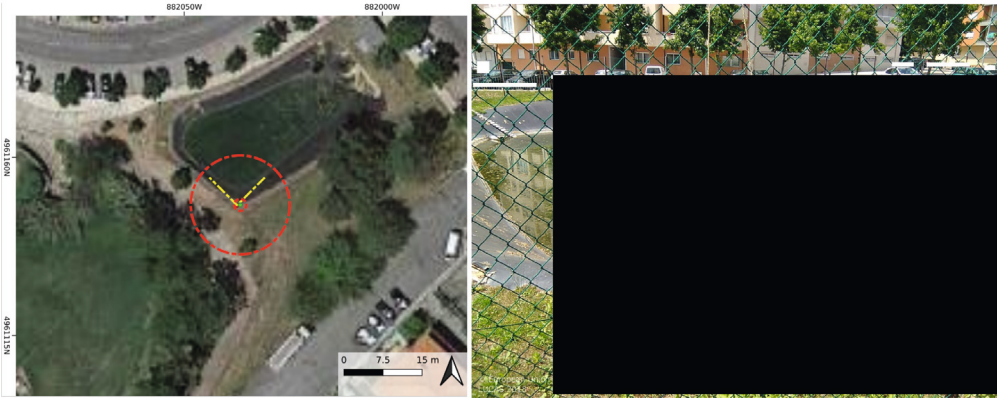
of seven hyperparameters with a 5-fold cross-validation on the training set, with 20% left out for the final metrics, to find the best combinations of hyper-parameter values, leading to a total of 28,800 fits. We ran one grid search for each depth of ResNet. See Supplementary Table 1 for the hyperparameters used.

The input features to train the RF are the total number of pixels in each semantically segmented mask belonging to each ADE20k class. Since the photos we used to segment do not have labeled segmented delineation references, we selected the depth of the ResNet with the highest accuracy on cross validation of the RF.



(a) LC: Cropland, heterogeneous field

(b) LC: Artificial Land, field of view is obstructed



(c) LC: Artificial Land, large patch of the photo is removed for anonymisation

Fig. 3. Examples of LUCAS point photos that were discarded during the selection process (on the right) along with their location (on the left): (a) Heterogeneous field of view on a sloped terrain, (b) the field of view is obstructed and the land cover is not represented on the photo, and (c) anonymization removes most of the visual information on the photo. On the left are very high resolution satellite images marked with: green point = theoretical point; red point = GPS point; yellow lines = field of view; red circles = minimum (1.5 m) and maximum radius (20 m) that were taken into account for the LC classification.

To measure the performance of the model and extract the final metrics, we used the test set. The RF classification algorithms and the hyperparameter tuning were implemented using Python's Scikit-learn packages RandomForestClassifier and GridSearchCV (Pedregosa et al., 2011). In contrast to the original publication (Breiman, 2001), the Scikit-learn implementation combines classifiers by averaging their probabilistic prediction instead of letting each classifier vote for a single class.

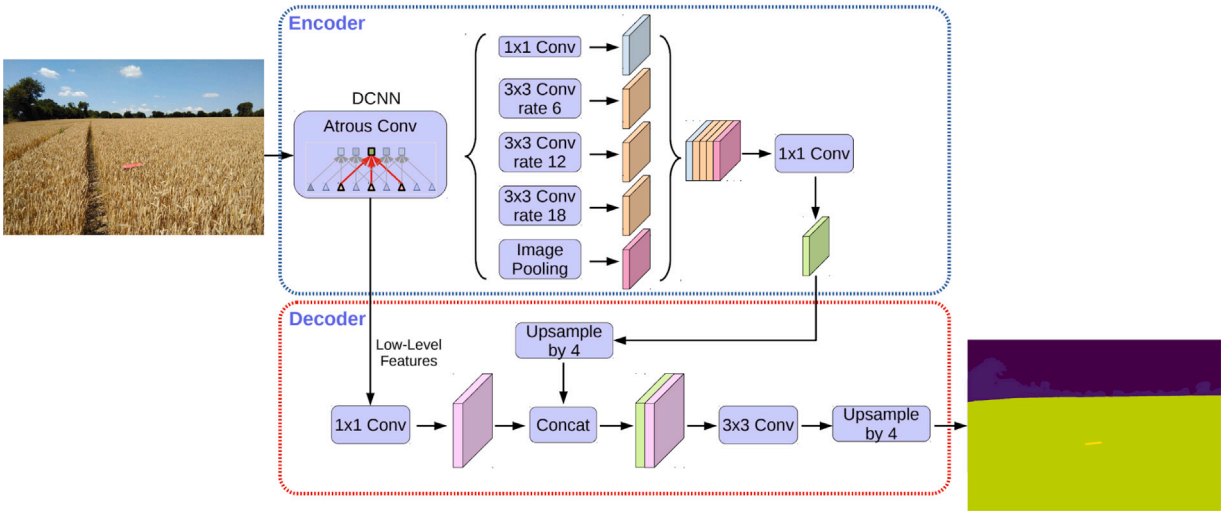


Fig. 4. Architecture of the Deeplabv3+ adapted from the original image in the paper by Chen et al. (2017).

2.4. Evaluation of segmentation and random forest results

The following metrics were used to assess the classification performance:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1score = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \quad (4)$$

$$Producer's\ accuracy = 1 - \frac{FN}{FN + TP} \quad (5)$$

$$User's\ accuracy = 1 - \frac{FP}{FP + TP} \quad (6)$$

where TP is the number of true positive LC classes predicted, TN is the number of true negative, FP is the number of false positive while FN is the number of false negatives LC classes. These values are derived from a confusion matrix that tabulates the LC class predictions against the LUCAS reference. In addition, the F1 Score, Producer's and User's accuracy, precision and recall were calculated for each individual class.

We also undertook an analysis based on feature importance, which is a measure of the predictive power or relevance of each input feature in the model's decision-making process. It is typically calculated based on the decrease in impurity (such as Gini impurity or entropy) that each feature contributes when used for splitting nodes in the decision trees within the forest. The idea is that if a feature consistently leads to a substantial reduction in impurity across different trees, it is likely to be more important in making accurate predictions.

3. Results

3.1. LUCAS point photo selection

As mentioned in the methodology, we first selected the LUCAS points that were not farther than 100 m away from the theoretical point to ensure that the photo captured the LC reported by the surveyor. This step resulted in a total of 160,064 points. The resulting points were then filtered by cross-matching the LUCAS LC with the CLC classification; see the crossmatching matrix in Supplementary Figure 1. Finally, to undertake a meaningful comparison between the two LC nomenclatures, we had to determine the correspondence between the LC classes of LUCAS and the CLC, which is summarized in Table 1.

Table 1
The relationship between the LUCAS and the Corine LC classes.

LUCAS LC Classes	CORINE LC Classes
A: Artificial Land	1: Artificial surfaces
B: Cropland	21: Arable land 22: Permanent crops 24: Heterogeneous agricultural areas
C: Woodland	31: Forests
D: Shrubland	322: Moors and heathland 323: Sclerophyllous vegetation 324: Transitional woodland-shrub
E: Grassland	23: Pastures 321: Natural grasslands
F: Bare soil Lichen	211: Non-irrigated arable land 331: Beaches, dunes, sands 332: Bare rocks 333: Sparsely vegetated areas 334: Burnt areas
G: Water areas	5: Water bodies
H: Wetlands	4: Wetlands

For Cropland, we matched the B Classes from LUCAS LC to the C2 classes from CLC. However, CLC subclass 231 (pastures) was not considered as Cropland but rather as Grassland. For the class Bare soil (F in LUCAS LC), we also considered photos that intersected with CLC class 211: Non-irrigated arable land, to ensure that any photos with bare soil were not omitted.

This crossmatching step resulted in a total of 102,371 LUCAS points and photos across the EU. The selection of photos was undertaken randomly, ensuring that only high-quality images were selected for the segmentation task, excluding examples like that shown in Fig. 3. This step resulted in 140 photos per class where possible. The 140-photo threshold was reached for the bare soil, arable land, shrubland, grassland, and woodland classes. However, since the distribution across the different LC classes is unequal, only 136 and 137 LUCAS points were selected for the water areas and wetland LC classes, respectively. The spatial distribution of these LUCAS points across the EU is shown in Fig. 5.

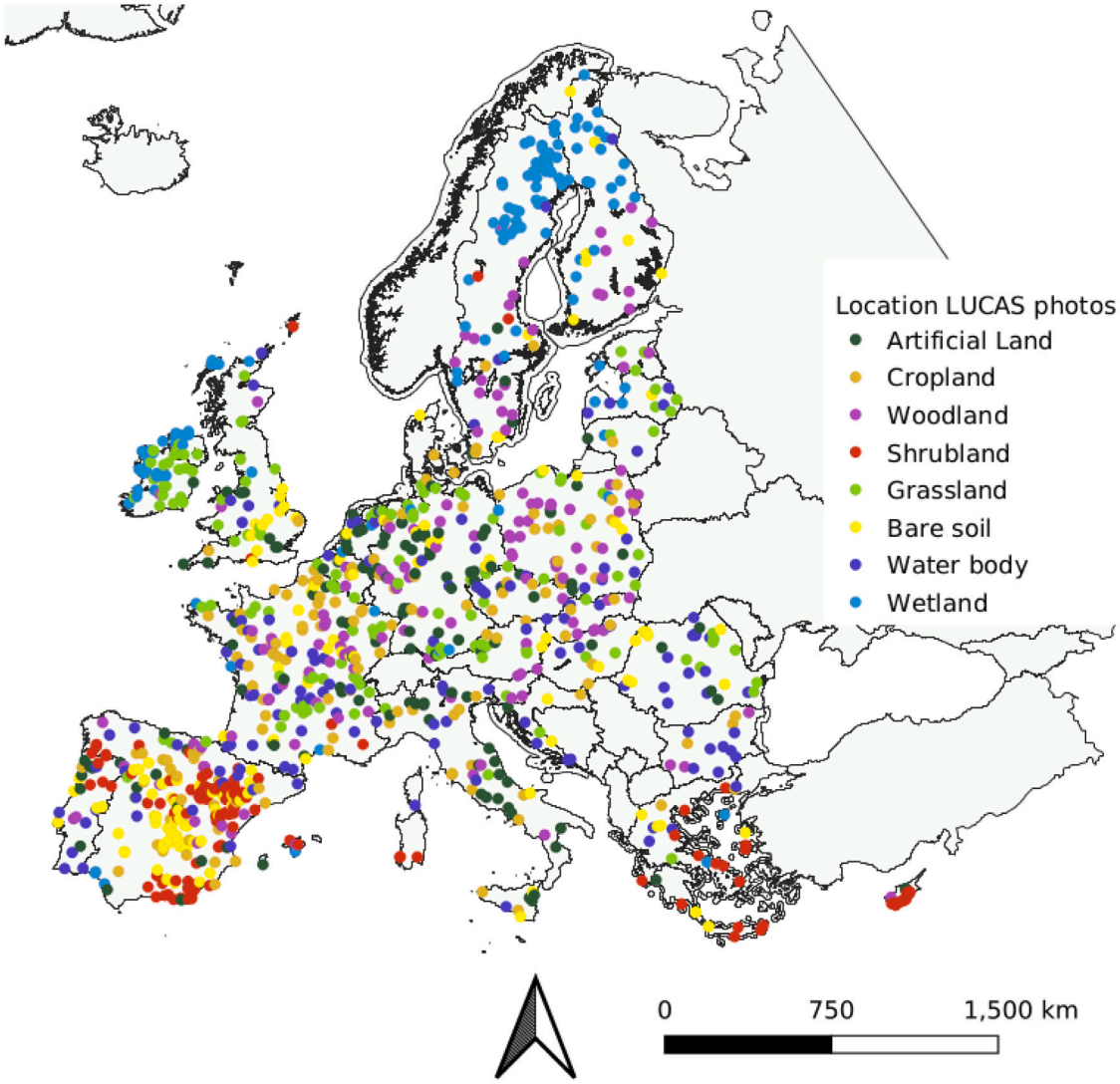


Fig. 5. The distribution of the point photos selected from the LUCAS 2018 data set shown with the level 1 LC classification legend.

3.2. Semantic segmentation

After the segmentation of all images, we evaluated the relationship between the segmented features and the LUCAS LC class. A total of 153 classes from ADE20k were found in the 1,112 images segmented. Fig. 6 shows the semantic segmentation output from Deeplabv3+ for a selection of photos with LUCAS LC labeled as (a) cropland, (b) water areas, (c) bare soil, and (d) grassland. The segmentation done by the network in Fig. 6(a) and (b) are good as they match the actual objects on the photos. In the case of Fig. 6(c) and (d), problems related to the identification of the objects are evident. Specifically, in Fig. 6(c), some plants have been segmented as trees, while in Fig. 6(d), the rocky terrain has been segmented as a wall object and the lichen on the rocks as a tree class. This is due to a domain shift between the ADE20k images used to train the model and the LUCAS photos used during the inference.

The resulting data set, including the original images plus the full segmentation, is available at http://data.europa.eu/89h/c6166c60-5221-437b-87ed-3aaec123801f (European Commission, Joint Research Centre (JRC), 2018).

Fig. 7 displays the relation between all features detected in the LUCAS photos by Deeplabv3+ for each LUCAS LC class, i.e., the number of ADE20k classes present and their total pixel area in a photo and the LC class. From this, we can observe several expected and clear relational patterns, e.g., the LC class Artificial Land has the majority of features belonging to artificial human-made objects. The class Woodland is mainly formed by trees, earth/soil, and plants, while sky is not present as the trees tend to obstruct it in these photos. In addition, the class Cropland is composed of plant, field, and sky, etc. Hence, we can see that there are clear patterns between the features detected by Deeplabv3+ trained with ADE20k and the LC LUCAS classes observed by the surveyor. At the same time, it is also apparent that each LC is composed of multiple elements that may occur across all the other LC types, which demonstrates the complexity of LC classification from photographs.

3.3. LC classification with RF

We ran a hyperparameter tuning on the RF for each backbone depth (50, 101, 200, 269); see Supplementary Table 1 for the hyperparameters used. Table 2 summarizes the performance results for each backbone after the hyperparameter tuning of the RF. ResNet101 had the highest cross-validation mean accuracy in the RF hypertuning stage; thus, it was selected as representative. However, the difference in the accuracy between the backbones is minimal. See Supplementary Table 2 for the final hyperparameters for each backbone.

ResNet101 with the best set of RF hyperparameters was then applied to the test dataset to obtain the confusion matrix and the performance

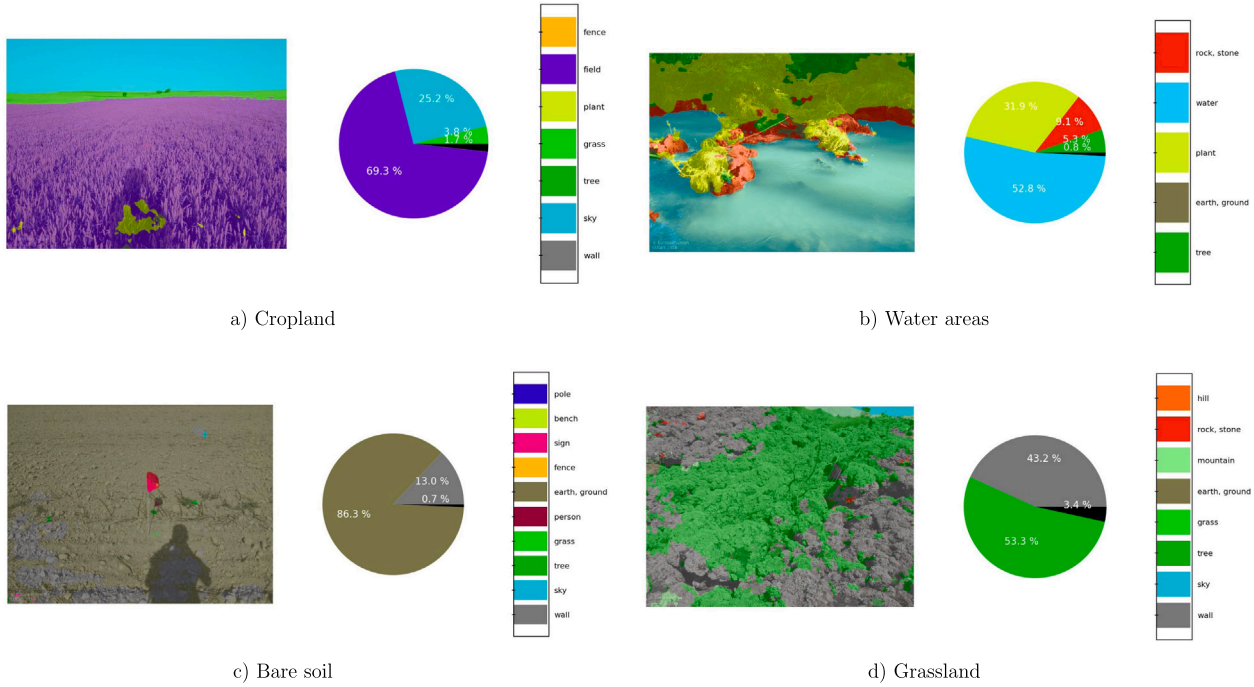


Fig. 6. Semantic segmentation with Deeplabv3+ trained with the ADE20k dataset. Examples (a) and (b) illustrate near perfect segmentation of objects, while (c) and (d) provide examples of incorrect segmentation.

Table 2
Accuracy metrics for the training dataset with a 5-fold cross-validation of the hyperparameter tuning on each backbone.

Model	Accuracy
deeplab_resnet50	0.83
deeplab_resnet101	0.85
deeplab_resnet200	0.81
deeplab_resnet269	0.80

Table 3
Performance metrics after the hyperparameter tuning for the best backbone, resnet101, averaging the metrics obtained by LC class.

	Precision	Recall	F1- score
All LC classes	0.89	0.89	0.89
Without the Wetlands class	0.94	0.93	0.93
	+0.05	+0.04	+0.04

Table 4
User's and producer's accuracy for LC classification with and without the Wetland class.

	With Wetlands		Without Wetlands	
	User	Producer	User	Producer
A-Artificial Land	1	1	1	1
B-Cropland	0.97	0.82	1	0.82
C-Woodland	0.93	0.93	0.96	0.96
D-Shrubland	0.93	0.89	0.93	0.93
E-Grassland	0.93	0.89	0.92	1
F-Bare soil, Lichens	0.73	0.82	0.81	0.86
G-Water Areas	0.92	0.96	0.92	1
H-Wetland	0.72	0.85		

metrics for each class (Figs. 8 and 9 and Tables 3 and 4). As we can see, the most difficult class for the RF to classify is Wetlands, see Table 3. This is expected because wetlands can have a similar appearance to grassland and shrubland and is often mixed in with these LC types. Wetlands are ecosystems that arise when inundation by water produces soils dominated by anaerobic and aerobic processes (Council et al., 1995) but that tend to be dominated by grassy vegetation. Shrublands and grasslands are defined as areas with a dominance of specific plant species. Because of this, our system had problems with wetlands. Hence, we have tested it without wetlands to demonstrate its value on other land cover classes (Figs. 8 and 9). This increases all the performance metrics, resulting in a final averaged F1 Score of 93% (see Table 3).

In both performance metrics, RF is able to perfectly discriminate Artificial land, with an F1 Score of 1. There is an improvement in the F1 Score for the classes Woodland, Shrubland and Grassland compared with the model trained with the Wetland class. This is due to the exclusion of the LC class Wetland resulting in a more robust RF model, which is better able to discriminate the interconnection between these LC classes (Fig. 8). The confusion of the RF model trained without Wetlands can be see in the classification of Bare Soil (class F), shown in Fig. 10, where examples (a) is arguably identified as Cropland which is considered as a misclassification since the LC is coded as Bare soil while example (c) shows incorrect predictions by the RF.

3.4. Evaluating the dominant segmented variable

From the final selected RF model with no Wetland class, we extracted the feature importance and corresponding standard deviations (see Fig. 11), where features with an importance score lower than 0.5 were filtered out for clarity. Higher values indicate greater importance, implying that the corresponding features play a more significant role in the classification task. Among the retained features, 'tree' exhibited the highest importance with a score of 0.091, followed by 'water' (0.088), 'grass' (0.082), and 'sky' (0.073). Features such as 'rock', 'field', 'building', 'earth/ground', and 'plant' had importance scores ranging from 0.051 to 0.063.

These findings provide insights into the relative importance of different features in the RF classifier's decision-making process. By understanding the significance of each feature, we gain a better understanding of the underlying patterns to derive landscape components for different LC classes. These patterns include not only the relationships between features but also the proportions of those features present

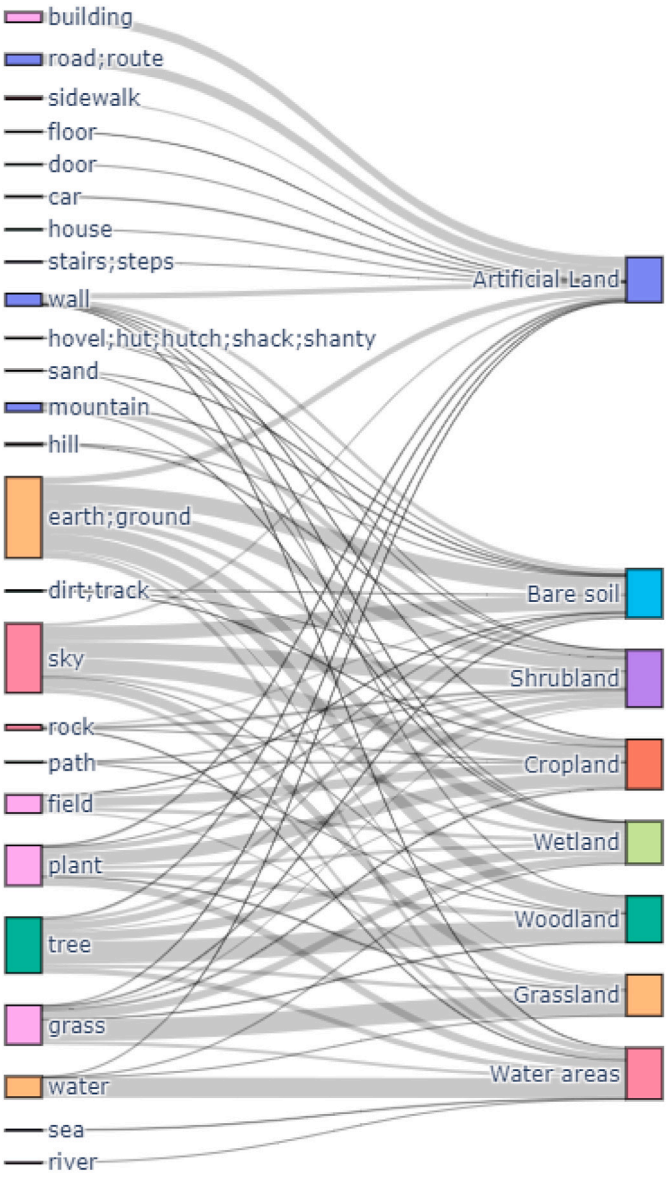


Fig. 7. Sankey diagram showing the relation between the features detected by Deeplabv3+ trained with ADE20k and the LUCAS LC class; the thickness of the links shows the amount of features linked to each LC class.

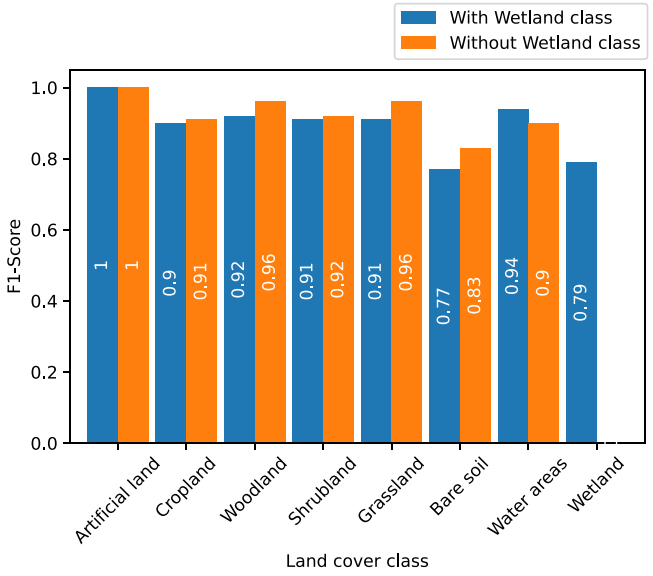


Fig. 8. Random Forest F1 Score metrics for each LC LUCAS class.

within the image. As illustrated in Fig. 12, the distinction between the Grassland and Woodland LC classes relies heavily on the proportion of pixels associated with each feature, since both LC classes share the same features. Additionally, the Artificial Land class is the only one that exhibits the presence of the 'building' feature, highlighting an example where the mere presence of a specific feature can lead to a distinct class separation. Both the Water and Wetland classes have 'water' but the amount is much smaller in Wetland while the presence of 'field' is much larger.

By leveraging this understanding of feature importance and considering the proportional representation of features, it becomes possible to differentiate between various LC classes more effectively. These insights also contribute to the broader goal of accurately characterizing and mapping landscape components based on their distinct feature compositions.

4. Discussion

Image classification with deep learning has proven to be useful for LC and LU classification of geo-tagged images (Leung and Newsam, 2015; Xing et al., 2018; ElQadi et al., 2020; Saadeldin et al., 2022). However, compared with previous studies in the literature, we have presented a different approach to the classification of LC from photographs by first segmenting the image into relevant landscape elements and then analyzing the relationship between these ADE20k classes and the LC class using a RF classifier. This two-stage approach is the key to understanding the links between the landscape in an image and the LC class that is predicted. As we have shown in Fig. 7, the landscape elements segmented by Deeplabv3+ with ADE20k classes and the LC classes exhibit clear links, showing that some elements are more prevalent in certain LC classes (e.g., the ADE20k earth/ground class is associated with the LUCAS Bare soil LC class). This semantic segmentation is critical as it provides context that would otherwise not be possible using a more targeted solution such as image classification. Similar conclusions were made by Stubbings et al. (2019), who used semantic segmentation in the development of an Urban Street Tree Vegetation Index using Google Street View imagery. Moreover, RF shows potential in classifying LC types that have the same type of ADE20k classes in the photos but with different proportions (i.e., different pixel areas). For example, both the Grasslands and Woodlands LC classes tend to have the ADE20k 'grass' and 'tree' classes present in the picture but in different proportions (see Fig. 12).

Another important feature of this approach was the use of an 'off-the-shelf' DL model, avoiding the need for manual segmentation, which is a very time-consuming task. Here we used Deeplabv3+ trained on ADE20k classes in inference mode to extract segmented information from the LUCAS photos. This was possible due to the domain similarities between the ADE20k images used to train the network and the LUCAS photos. Other studies have also used pre-trained networks to segment images, e.g., Cao et al. (2018, 2023) but largely in an urban context. In fact, it should be noted that the ADE20k and COCO datasets are mostly geared towards urban and sub-urban environments, and thus they are not meant to cope with the heterogeneity of complex landscapes across all LC types. For example, the COCO dataset includes 'plant' for all vegetation types. The ADE20k dataset includes 'field', which has proven to be useful here, but a field may be bare, contain shrubs, plants, crops, trees, or grassland, and hence there

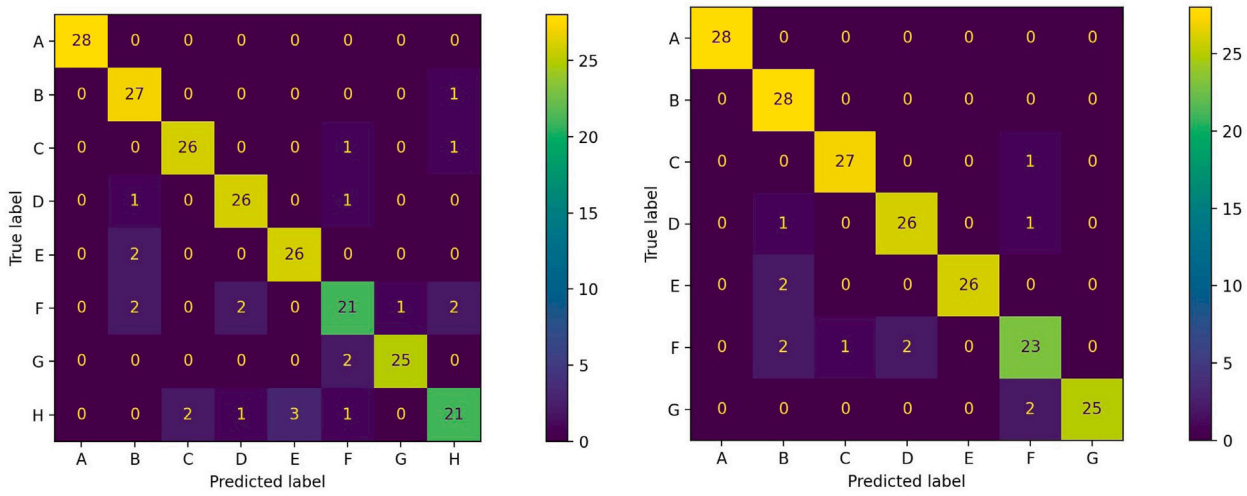


Fig. 9. Confusion matrix from the RF classifier with and without the class Wetlands. The classes are A: Artificial Land, B: Cropland, C: Woodland, D: Shrubland, E: Grassland, F: Bare soil and Lichens, G: Water Areas, H: Wetlands.

are limitations in using this dataset. For this reason, the RF model had high precision and recall on detecting Artificial Land since the neural network used was mainly trained on urban areas. As shown in Fig. 7, this LC class is the only class that relies heavily on artificial human-made objects, which makes it easier for the RF to discriminate the LUCAS Artificial Land class from others. However, we have also shown that RF can deal with the heterogeneity of landscapes, especially for ones that are managed, like grasslands or croplands, where the landscape can vary depending on the intensity of the LU or the size of the parcel. For those classes, we achieved an F1 Score of 0.96 and 0.91, respectively (see Fig. 8 and Table 3).

In terms of performance overall, this two-stage approach yielded an average F1 Score of 89% when including the Wetland class and 93% without (Fig. 8 and Table 3). Removing the Wetland class also improved the F1 Score, recall and precision of some other classes where there was confusion. In our upcoming research, we intend to improve the wetland classification procedure. This could entail the introduction of filters reliant on geographical coordinates or the integration of more images to account for the temporal dimensions of wetland changes.

Compared to other studies that have classified geo-tagged photos for LC, the overall accuracies in these studies are generally lower, ranging between 46.7 to 76% (Zhu and Newsam, 2015; Zhu et al., 2019), 41 to 62% (Srivastava et al., 2019, 2020), 78% (Cao et al., 2018) and 48.4 to 73.6% (Xu et al., 2017). However, the majority of these papers dealt primarily with urban classes where the classifier in this study performed very well on artificial surfaces. In contrast, the Wetland class was mapped by Xu et al. (2017), who achieved user's and producer's accuracies ranging between 0.58 to 0.85 and 0.65 to 0.89, respectively, depending on the probability threshold chosen. In this study, no filtering was applied and values of 0.72 to 1 for user's and 0.82 to 1 for producer's accuracies were achieved, which shows that these metrics performed in line with this other study's results, see Table 4. Hence, overall, the approach presented here shows promise for the automated classification of LC from geo-tagged photos.

In terms of the performance related to the segmentation inference with the four backbones available (ResNet 50/101/200/269), we found that ResNet with 101 layers performed best, although the differences in performance with the other depths was minimal; see Table 2. One hypothesis that may support this finding is that shallower networks tend to overfit less on the training data compared to deeper nets (Bejani and Ghatee, 2021). Adding more layers will extract more features but can also result in an overfitted model with a lower capacity to generalize than shallower models. This might be a possible argument to explain the higher accuracy of our results for resnet50 and 101 compared with 200 and 269.

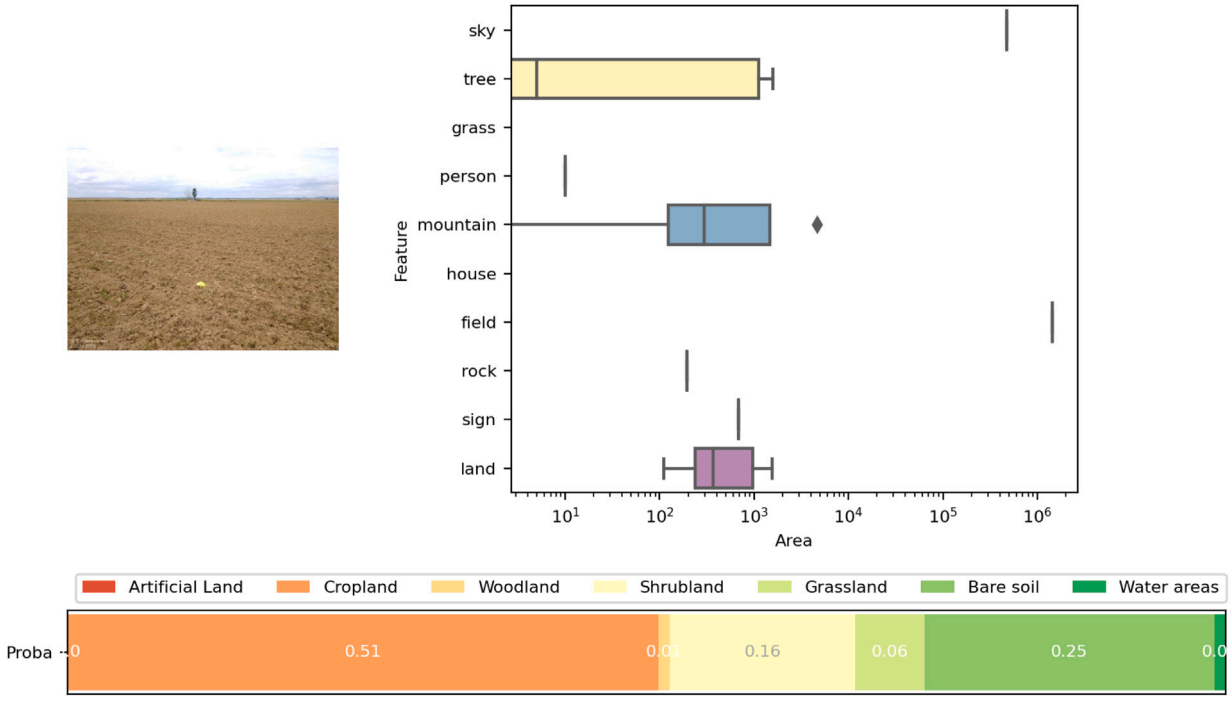
In addition to the limitations related to the ability to recognize the Wetland class, another issue relates to the selection of point photos from LUCAS. The LC class at each location is initially determined through aerial photo-interpretation 'from the top' on an area of 7 m^2 while the landscape photos provide an oblique view. The field of view captured by the LUCAS photos may therefore be different from the classified LC, even after cross-matching of the LUCAS point locations with the CLC maps (Fig. 3). This difference in perspectives demonstrates the challenges of LC classification with geo-tagged photos, but it also shows that they have considerable potential for producing high temporal and spatial resolution reference data sets for training, validation and other applications.

5. Conclusions

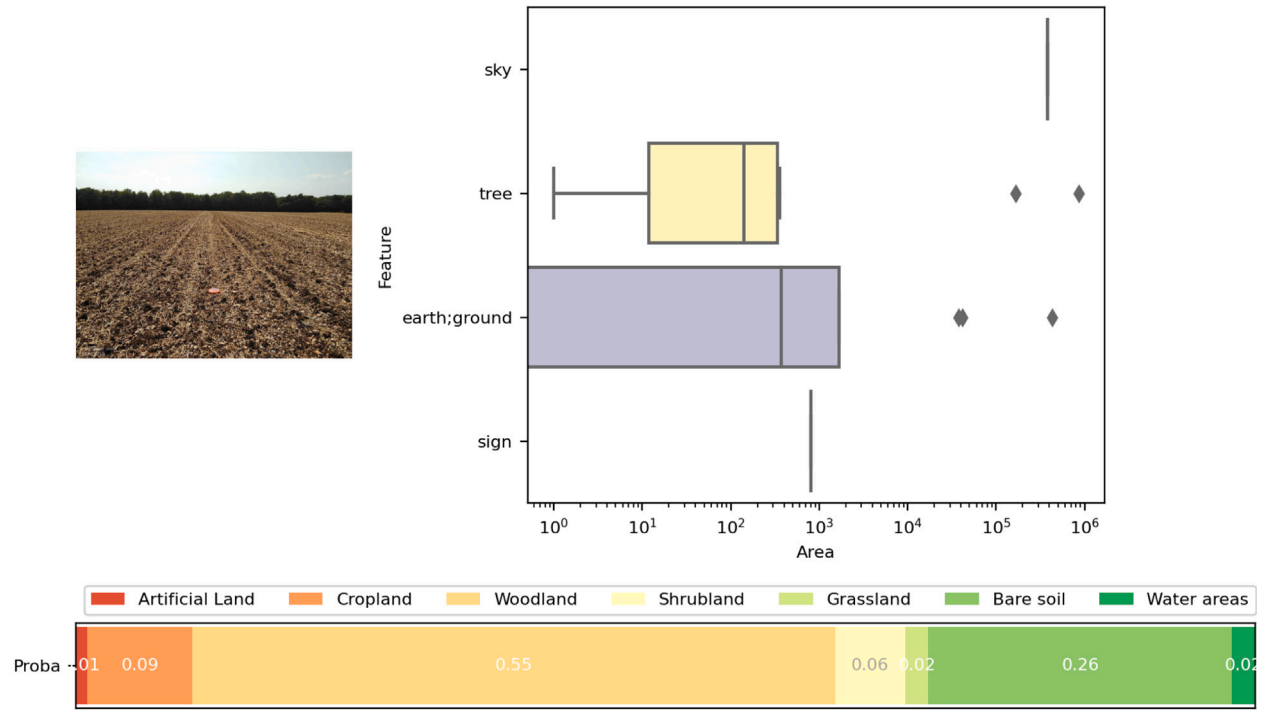
In this paper, we proposed a two-stage approach for LC classification using LUCAS photos. First, we segmented the photos with an already trained neural network to extract the relevant LC related classes present in the photos. Secondly, we used the cumulative addition of these pixel-wise detected classes in each photo as input features to a RF classifier. Overall the results were good in terms of the future applicability of this approach to automatically classifying geo-referenced landscape images more generally. However, the Wetland class created confusion with other classes and hence further work in improving the identification of wetlands from photographs is still required.

We have also shown that by segmenting the in-situ landscape imagery, the different elements that comprise a landscape in all its complexity can be captured, and that by combining these elements via a classifier, we can derive meaningful information for LC classification. Although we used an 'off-the-shelf' DL model instead of training one from scratch, specific annotated semantic segmentation training sets should be developed, which embed more variables that represent the natural environment and can better differentiate between LC classes. This will help to overcome some of the errors in the segmentation that we highlighted in Fig. 6, but will also contribute to better characterization of other elements in the landscape as well as the ability to recognize more detailed LC classes. Additionally, future applications of this work using other sources of SLI like Google Street View, Flickr, or other datasets could help improve the temporal and spatial analysis of this study.

The extraction of LC information from SLI offers valuable data for the validation and improvement of LC maps. By comparing the LC information derived from SLI with the corresponding LC classes in pre-existing maps, discrepancies and inaccuracies can be identified and



a)True LC: Bare soil, RF result: Cropland



b) True LC: Bare soil, RF result: Woodland

Fig. 10. The LUCAS point photo distribution by pixel area of the ADE20k objects segmented and the probability output by the Random Forest for each LC class.

corrected. Furthermore, this method can serve as a ground truth for future LC classification techniques in Europe using remote sensing data, specially for unsupervised methods (Paris et al., 2022).

Finally, beyond the provision of reference data for training and validation of LC maps, extracting landscape elements from in-situ imagery can be valuable for other applications such as better quantifying

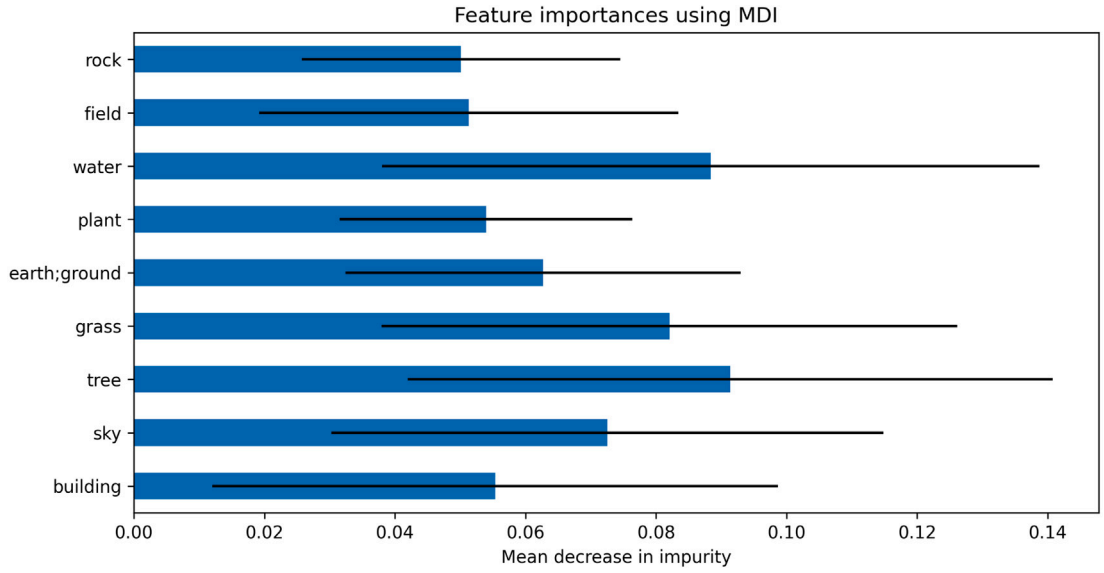


Fig. 11. Features Importance for features with a mean decrease of impurity bigger than 0.05.

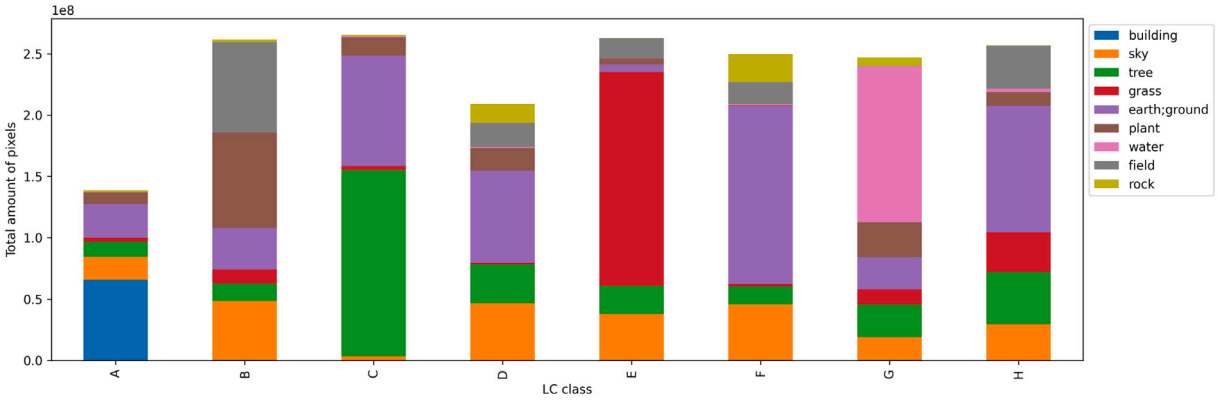


Fig. 12. Distribution of the amount of pixels for each feature in each LC class. The classes are A: Artificial Land, B: Cropland, C: Woodland, D: Shrubland, E: Grassland, F: Bare soil and Lichens, G: Water Areas, H: Wetlands.

landscape complexity (Ode et al., 2010), deriving metrics on landscape structure, composition, and heterogeneity as a facilitator for e.g. biodiversity (Fahrig et al., 2011), or as shown in Zhao et al. (2022) as spatial characteristics of the soundscape ecology in urban areas. Moreover, we have explored how image segmentation can provide information on landscape openness, by quantifying the distance to landscape elements making up the horizon (Martinez-Sanchez et al., 2022). Other applications include the monitoring of habitats using photographs and for verification purposes, e.g., to confirm declarations related to the Common Agricultural Policy in the EU. Hence, this shows the potential not only for LC classification purposes but also for the extraction of landscape elements from the increasing volume of georeferenced photographs.

CRediT authorship contribution statement

Laura Martinez-Sanchez: Conceptualization, Data curation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **Linda See:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Momchil Yordanov:** Data curation, Methodology, Writing – review & editing. **Astrid Verhegghen:** Formal analysis, Writing – original draft, Writing – review & editing. **Neija Elvekjaer:** Data curation, Writing – review & editing. **Davide Muraro:** Writing – original draft, Writing – review & editing. **Raphaël d'Andrimont:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Marijn van der Velde:** Conceptualization, Data curation, Formal analysis, Methodology, Supervision, Validation, Writing – original draft, 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.

Data availability

The data used in this study is available at http://data.europa.eu/89h/c6166c60-5221-437b-87ed-3aaec123801f (European Commission, Joint Research Centre (JRC), 2018)

Acknowledgments

We gratefully acknowledge the support of this research by the JRC Exploratory Research program through the Rural Refocus project (31280).

Software and data availability

Software name: LUCAS LC classifications
Developer: Laura Martinez-Sanchez
First year available: 2023
Program language: Python 3.X
License: GPL-3.0
Availability: https://github.com/MartinezLaura/LandCoverClassification.git
Used environment:
- CPU: Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20 GHz
- RAM: 25 GB
- GPU: NVIDIA GeForce GTX 1080 Ti

The dataset used in this study, including the images and masks, is available at http://data.europa.eu/89h/c6166c60-5221-437b-87ed-3aaec123801f (European Commission, Joint Research Centre (JRC), 2018).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.envsoft.2023.105931.

References

Alem, A., Kumar, S., 2022. Transfer learning models for land cover and land use classification in remote sensing image. Appl. Artif. Intell. (ISSN: 0883-9514) 36 (1), 2014192. http://dx.doi.org/10.1080/08839514.2021.2014192, iSSN: 1087-6545, URL https://www.tandfonline.com/doi/full/10.1080/08839514.2021.2014192.

Bejani, M.M., Ghatee, M., 2021. A systematic review on overfitting control in shallow and deep neural networks. Artif. Intell. Rev. 54 (8), 6391–6438.

Berland, A., Lange, D.A., 2017. Google street view shows promise for virtual street tree surveys. Urban For. Urban Green. 21, 11–15.

Bontemps, S., Defourny, P., Radoux, J., Van Bogaert, E., Lamarche, C., Achard, F., Mayaux, P., Boettcher, M., Brockmann, C., Kirches, G., et al., 2013. Consistent global land cover maps for climate modelling communities: Current achievements of the ESA's land cover CCI. In: Proceedings of the ESA Living Planet Symposium, Edimburgh, vol. 13, pp. 9–13.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Buchhorn, M., Lesiv, M., Tsendbazar, N.-E., Herold, M., Bertels, L., Smets, B., 2020. Copernicus global land cover layers—collection 2. Remote Sens. 12 (6), 1044.

Cao, R., Liao, C., Li, Q., Tu, W., Zhu, R., Luo, N., Qiu, G., Shi, W., 2023. Integrating satellite and street-level images for local climate zone mapping. Int. J. Appl. Earth Obs. Geoinf. (ISSN: 15698432) 119, 103323. http://dx.doi.org/10.1016/j.jag.2023.103323, URL https://linkinghub.elsevier.com/retrieve/pii/S1569843223001450.

Cao, R., Zhu, J., Tu, W., Li, Q., Cao, J., Liu, B., Zhang, Q., Qiu, G., 2018. Integrating aerial and street view images for urban land use classification. Remote Sens. 10 (10), 1553.

Chen, Z., Duan, Y., Wang, W., He, J., Lu, T., Dai, J., Qiao, Y., 2022a. Vision transformer adapter for dense predictions. arXiv preprint arXiv:2205.08534.

Chen, M., Liu, Y., Arribas-Bel, D., Singleton, A., 2022b. Assessing the value of user-generated images of urban surroundings for house price estimation. Landsc. Urban Plan. 226, 104486.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach. Intell. 40 (4), 834–848.

Corbane, C., Pesaresi, M., Kemper, T., Politis, P., Florczyk, A.J., Syrris, V., Melchiorri, M., Sabo, F., Soille, P., 2019. Automated global delineation of human settlements from 40 years of landsat satellite data archives. Big Earth Data 3 (2), 140–169.

Council, N.R., et al., 1995. Wetlands: Characteristics and Boundaries. National Academies Press.

Cui, L., Yang, H., Chu, L., He, Q., Xu, F., Qiao, Y., Yan, Z., Wang, R., Ci, H., 2022. The verification of land cover datasets with the geo-tagged natural scene images. ISPRS Int. J. Geo-Inf. 11 (11), 567.

d'Andrimont, R., Yordanov, M., Lemoine, G., Yoong, J., Nikel, K., Van der Velde, M., 2018. Crowdsourced street-level imagery as a potential source of in-situ data for crop monitoring. Land 7 (4), 127.

d'Andrimont, R., Yordanov, M., Martinez-Sanchez, L., Eiselt, B., Palmieri, A., Dominici, P., Gallego, J., Reuter, H.I., Joebges, C., Lemoine, G., et al., 2020. Harmonised LUCAS in-situ land cover and use database for field surveys from 2006 to 2018 in the European union. Sci. Data 7 (1), 1–15.

d'Andrimont, R., Yordanov, M., Martinez-Sanchez, L., Van der Velde, M., 2022. Monitoring crop phenology with street-level imagery using computer vision. Comput. Electron. Agric. 196, 106866.

ElQadi, M.M., Lesiv, M., Dyer, A.G., Dorin, A., 2020. Computer vision-enhanced selection of geo-tagged photos on social network sites for land cover classification. Environ. Model. Softw. 128, 104696.

European Commission, Joint Research Centre (JRC), 2018. Land Cover Computer Vision LUCAS. http://data.europa.eu/89h/c6166c60-5221-437b-87ed-3aaec123801f.

European Environment Agency (EEA)", f.ex. in 2018: © European Union, Copernicus Land Monitoring Service 2018. European Environment Agency (EEA). 2018. Corine land cover, copernicus land monitoring service. https://land.copernicus.eu/pan-european/corine-land-cover/clc2018.

Eurostat, 2018a. Technical reference document C-3: Classification. https://ec.europa.eu/eurostat/documents/205002/8072634/LUCAS2018-C3-Classification.pdf.

Eurostat, 2018b. Technical reference document S1: Stratification guidelines. https://ec.europa.eu/eurostat/documents/205002/7329820/LUCAS2018_S1-StratificationGuidelines_20160523.pdf.

Fahrig, L., Baudry, J., Brotons, L., Burel, F.G., Crist, T.O., Fuller, R.J., Sirami, C., Siriwardena, G.M., Martin, J.-L., 2011. Functional landscape heterogeneity and animal biodiversity in agricultural landscapes. Ecol. Lett. 14 (2), 101–112.

Funada, S., Tsutsumida, N., 2022. Mapping cherry blossoms from geotagged street-level photos. bioRxiv.

Gallego, J., Bamps, C., 2008. Using CORINE land cover and the point survey LUCAS for area estimation. Int. J. Appl. Earth Obs. Geoinf. 10 (4), 467–475.

Gao, Y., Liu, L., Zhang, X., Chen, X., Mi, J., Xie, S., 2020. Consistency analysis and accuracy assessment of three global 30-m land-cover products over the European union using the LUCAS dataset. Remote Sens. 12, 3479.

Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E.L., Fei-Fei, L., 2017. Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the United States. Proc. Natl. Acad. Sci. 114 (50), 13108–13113.

Goel, R., Garcia, L.M., Goodman, A., Johnson, R., Aldred, R., Murugesan, M., Brage, S., Bhalla, K., Woodcock, J., 2018. Estimating city-level travel patterns using street imagery: A case study of using Google street view in Britain. PLoS One 13 (5), e0196521.

Guo, J., He, H., He, T., Lausen, L., Li, M., Lin, H., Shi, X., Wang, C., Xie, J., Zha, S., et al., 2020. GluonCV and GluonNLP: Deep learning in computer vision and natural language processing. J. Mach. Learn. Res. 21 (23), 1–7.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., et al., 2013. High-resolution global maps of 21st-century forest cover change. science 342 (6160), 850–853.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Kang, J., Körner, M., Wang, Y., Taubenböck, H., Zhu, X.X., 2018. Building instance classification using street view images. ISPRS J. Photogramm. Remote Sens. 145, 44–59.

Karydas, C., Gitas, I., Kuntz, S., Minakou, C., 2015. Use of LUCAS LC point database for validating country-scale land cover maps. Remote Sens. 7, 5012–5041.

Kelly, C.M., Wilson, J.S., Baker, E.A., Miller, D.K., Schootman, M., 2013. Using Google street view to audit the built environment: Inter-rater reliability results. Ann. Behav. Med. 45 (suppl_1), S108–S112.

Leung, D., Newsam, S., 2015. Land cover classification using geo-referenced photos. Multimedia Tools Appl. 74 (24), 11741–11761.

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., Zhang, W., 2015. Assessing street-level urban greenery using Google street view and a modified green view index. Urban For. Urban Green. 14 (3), 675–685.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft coco: Common objects in context. In: European Conference on Computer Vision. Springer, pp. 740–755.

Lu, Y., 2019. Using Google street view to investigate the association between street greenery and physical activity. Landsc. Urban Plan. 191, 103435.

Mack, B., Leinenkugel, P., Kuenzer, C., Dech, S., 2017. A semi-automated approach for the generation of a new land use and land cover product for Germany based on landsat time-series and lucasin-situ data. Remote Sens. Lett. 8, 244–253.

Martinez-Sanchez, L., Borio, D., d'Andrimont, R., van der Velde, M., 2022. Skyline variations allow estimating distance to trees on landscape photos using semantic segmentation. Ecol. Inform. 70, 101757.

Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: An applied review. Int. J. Remote Sens. 39 (9), 2784–2817.

Neuhold, G., Ollmann, T., Rota Bulo, S., Kontschieder, P., 2017. The mapillary vistas dataset for semantic understanding of street scenes. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 4990–4999.

Ning, H., Ye, X., Chen, Z., Liu, T., Cao, T., 2022. Sidewalk extraction using aerial and street view images. Environ. Plan. B: Urban Anal. City Sci. (ISSN: 2399-8083) 49 (1), 7–22. http://dx.doi.org/10.1177/2399808321995817, ISSN: 2399-8091, URL http://journals.sagepub.com/doi/10.1177/2399808321995817.

Ode, Å., Hagerhall, C.M., Sang, N., 2010. Analysing visual landscape complexity: Theory and application. Landscape Res. 35 (1), 111–131.

Padmanaba, M., Sheil, D., Basuki, I., Liswanti, N., 2013. Accessing local knowledge to identify where species of conservation concern occur in a tropical forest landscape. Environ. Manag. (ISSN: 0364-152X) 52 (2), 348–359. http://dx.doi.org/10.1007/s00267-013-0051-7, ISSN: 1432-1009. Number: 2. URL http://link.springer.com/article/10.1007/s00267-013-0051-7.

Palmieri, A., Martino, L., Dominici, P., Kasanko, M., 2011. Land cover and land use diversity indicators in LUCAS 2009 data. Land Qual. Land Use Inf. Eur. Union 59–68.

Paris, C., Gasparella, L., Bruzzone, L., 2022. A scalable high-performance unsupervised system for producing large-scale HR land cover maps: The Italian country case study. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15, 9146–9159.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. Nature 540 (7633), 418–422.

Pflugmacher, D., Rabe, A., Peters, M., Hostert, P., 2019. Mapping pan-European land cover using landsat spectral-temporal metrics and the European LUCAS survey. Remote Sens. Environ. 221, 583–595.

Potapov, P., Turubanova, S., Hansen, M.C., Tyukavina, A., Zalles, V., Khan, A., Song, X.-P., Pickens, A., Shen, Q., Cortez, J., 2022. Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century. Nature Food 3 (1), 19–28.

Radoux, J., Lamarche, C., Van Bogaert, E., Bontemps, S., Brockmann, C., Defourny, P., 2014. Automated training sample extraction for global land cover mapping. Remote Sens. 6 (5), 3965–3987.

Rundle, A.G., Bader, M.D., Richards, C.A., Neckerman, K.M., Teitler, J.O., 2011. Using Google street view to audit neighborhood environments. Am. J. Prevent. Med. 40 (1), 94–100.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al., 2015. Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. 115 (3), 211–252.

Saadeldin, M., O'Hara, R., Zimmermann, J., Mac Namee, B., Green, S., 2022. Using deep learning to classify grassland management intensity in ground-level photographs for more automated production of satellite land use maps. Remote Sens. Appl.: Soc. Environ. 26, 100741.

See, L., Bayas, J.C.L., Lesiv, M., Schepaschenko, D., Danylo, O., McCallum, I., Dürauer, M., Georgieva, I., Domian, D., Fraisl, D., et al., 2022. Lessons learned in developing reference data sets with the contribution of citizens: The Geo-Wiki experience. Environ. Res. Lett. 17 (6), 065003.

Srivastava, S., Vargas Munoz, J.E., Lobry, S., Tuia, D., 2020. Fine-grained landuse characterization using ground-based pictures: A deep learning solution based on globally available data. Int. J. Geogr. Inf. Sci. 34 (6), 1117–1136.

Srivastava, S., Vargas-Munoz, J.E., Tuia, D., 2019. Understanding urban landuse from the above and ground perspectives: A deep learning, multimodal solution. Remote Sens. Environ. 228, 129–143.

Stehman, S.V., Foody, G.M., 2019. Key issues in rigorous accuracy assessment of land cover products. Remote Sens. Environ. 231, 111199.

Stubbings, P., Peskett, J., Rowe, F., Arribas-Bel, D., 2019. A hierarchical urban forest index using street-level imagery and deep learning. Remote Sens. 11 (12), 1395.

Szantoi, Z., Geller, G.N., Tsendbazar, N.-E., See, L., Griffiths, P., Fritz, S., Gong, P., Herold, M., Mora, B., Obregón, A., 2020. Addressing the need for improved land cover map products for policy support. Environ. Sci. Policy 112, 28–35.

Tong, M., She, J., Tan, J., Li, M., Ge, R., Gao, Y., 2020. Evaluating street greenery by multiple indicators using street-level imagery and satellite images: A case study in Nanjing, China. Forests (ISSN: 1999-4907) 11 (12), 1347. http://dx.doi.org/10.3390/f11121347, URL https://www.mdpi.com/1999-4907/11/12/1347.

Waldner, F., Schucknecht, A., Lesiv, M., Gallego, J., See, L., Pérez-Hoyos, A., d'Andrimont, R., De Maet, T., Bayas, J.C.L., Fritz, S., et al., 2019. Conflation of expert and crowd reference data to validate global binary thematic maps. Remote Sens. Environ. 221, 235–246.

Wang, W., Dai, J., Chen, Z., Huang, Z., Li, Z., Zhu, X., Hu, X., Lu, T., Lu, L., Li, H., et al., 2022. Internimage: Exploring large-scale vision foundation models with deformable convolutions. arXiv preprint arXiv:2211.05778.

Weigand, M., Staab, J., Wurm, M., Taubenböck, H., 2020. Spatial and semantic effects of LUCAS samples on fully automated land use/land cover classification in high-resolution sentinel-2 data. Int. J. Appl. Earth Obs. Geoinf. 88, 102065.

Weiss, K., Khoshgoftaar, T.M., Wang, D., 2016. A survey of transfer learning. J. Big Data (ISSN: 2196-1115) 3 (1), 9. http://dx.doi.org/10.1186/s40537-016-0043-6, URL http://journalofbigdata.springeropen.com/articles/10.1186/s40537-016-0043-6.

Xing, H., Meng, Y., Wang, Z., Fan, K., Hou, D., 2018. Exploring geo-tagged photos for land cover validation with deep learning. ISPRS J. Photogramm. Remote Sens. 141, 237–251.

Xu, G., Zhu, X., Fu, D., Dong, J., Xiao, X., 2017. Automatic land cover classification of geo-tagged field photos by deep learning. Environ. Model. Softw. 91, 127–134.

Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Zhang, Z., Lin, H., Sun, Y., He, T., Muller, J., Manmatha, R., Li, M., Smola, A., 2020. ResNeSt: Split-attention networks. arXiv preprint arXiv:2004.08955.

Zhao, Y., Xu, S., Huang, Z., Fang, W., Huang, S., Huang, P., Zheng, D., Dong, J., Chen, Z., Yan, C., Zhong, Y., Fu, W., 2022. Temporal and spatial characteristics of soundscape ecology in urban forest areas and its landscape spatial influencing factors. Forests (ISSN: 1999-4907) 13 (11), 1751. http://dx.doi.org/10.3390/f13111751, URL https://www.mdpi.com/1999-4907/13/11/1751.

Zhou, H., He, S., Cai, Y., Wang, M., Su, S., 2019a. Social inequalities in neighborhood visual walkability: Using street view imagery and deep learning technologies to facilitate healthy city planning. Sustain. Cities Soc. 50, 101605.

Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., Torralba, A., 2017. Scene parsing through ade20k dataset. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 633–641.

Zhou, B., Zhao, H., Puig, X., Xiao, T., Fidler, S., Barriuso, A., Torralba, A., 2019b. Semantic understanding of scenes through the ade20k dataset. Int. J. Comput. Vis. 127 (3), 302–321.

Zhu, Y., Deng, X., Newsam, S., 2019. Fine-grained land use classification at the city scale using ground-level images. IEEE Trans. Multimed. 21 (7), 1825–1838.

Zhu, Y., Newsam, S., 2015. Land use classification using convolutional neural networks applied to ground-level images. In: Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 1–4.