Ecological Informatics 50 (2019) 220–233

Contents lists available at ScienceDirect

Ecological Informatics

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

Automatic detection of woody vegetation in repeat landscape photographs using a convolutional neural network

Ulrike Bayr*, Oskar Puschmann

Norwegian Institute of Bioeconomy Research (NIBIO), Division of Survey and Statistics, Department of Landscape Monitoring, P.O. Box 115, 1431 Ås, Norway

A R T I C L E I N F O

Keywords:
Repeat photography
Photo monitoring
Landscape monitoring
Landscape change
Vegetation succession
Machine learning

A B S T R A C T

Repeat photography is an efficient method for documenting long-term landscape changes. So far, the usage of repeat photographs for quantitative analyses is limited to approaches based on manual classification. In this paper, we demonstrate the application of a convolutional neural network (CNN) for the automatic detection and classification of woody regrowth vegetation in repeat landscape photographs. We also tested if the classification results based on the automatic approach can be used for quantifying changes in woody vegetation cover between image pairs. The CNN was trained with 50 × 50 pixel tiles of woody vegetation and non-woody vegetation. We then tested the classifier on 17 pairs of repeat photographs to assess the model performance on unseen data. Results show that the CNN performed well in differentiating woody vegetation from non-woody vegetation (accuracy = 87.7%), but accuracy varied strongly between individual images. The very similar appearance of woody vegetation and herbaceous species in photographs made this a much more challenging task compared to the classification of vegetation as a single class (accuracy = 95.2%). In this regard, image quality was identified as one important factor influencing classification accuracy. Although the automatic classification provided good individual results on most of the 34 test photographs, change statistics based on the automatic approach deviated from actual changes. Nevertheless, the automatic approach was capable of identifying clear trends in increasing or decreasing woody vegetation in repeat photographs. Generally, the use of repeat photography in landscape monitoring represents a significant added value to other quantitative data retrieved from remote sensing and field measurements. Moreover, these photographs are able to raise awareness on landscape change among policy makers and public as well as they provide clear feedback on the effects of land management.

1. Introduction

It is in the nature of cultural landscapes to undergo continuous change in close interaction with the way we use them. Unfortunately, not all landscape changes move in the desired direction. In Norway as in many other countries, secondary forest succession on semi-natural and marginal land as consequence of abandonment is at present one of the most prevailing trends. The result has negative impacts on farmland and grazing resources, landscape diversity, cultural values and biological diversity (Amici et al., 2017; Fjellstad and Dramstad, 1999; Lasanta et al., 2015; Sang et al., 2014). Landscape monitoring is an efficient way to identify ongoing trends at an early stage, which enables us to steer landscape changes towards a more sustainable land management.

Today, landscape monitoring is mainly based on remote sensing imagery retrieved from aerial or satellite platforms. However, also ground-based photographs have been acknowledged as an important data source for documenting the state and change of landscapes and ecosystems for a long time (Pickard, 2002). Repeat photography, often referred as photo monitoring, is a method where ground-level photographs are taken from exactly the same location at different points in time. In the case of landscapes, the time steps between the images are usually several years or decades, sometimes even up to a whole century. These “then and now” images are effective in communicating long-term landscape changes to a broad audience (Klett, 2010). However, with the technological advances in aerial and satellite remote sensing, ground-based photographs have lost most of their relevance in modern landscape monitoring. One of the main reasons for this development is that the retrieval of quantitative information from photographs and their use for spatial analyses is limited. Nonetheless, these photographs tell a detailed story on how the landscape has changed and therefore, landscape monitoring can greatly benefit from the integration of this rich data source.

With *The Changing Mile*, Hastings and Turner (1965) laid the

* Corresponding author.
E-mail address: ulrike.bayr@nibio.no (U. Bayr).

https://doi.org/10.1016/j.ecoinf.2019.01.012
Received 10 December 2018; Received in revised form 28 January 2019; Accepted 28 January 2019
Available online 06 February 2019

foundation for systematic repeat photography as a research method. Klett et al. (1984) supplemented this work with their experiences from the photographic project *Second View*. Since then, repeat photography has been used for the monitoring of glacier retreat (Molnia, 2010; Wiesmann et al., 2012), geomorphological processes (Khan et al., 2013, Frankl et al., 2011, Conedara et al., 2013), tree line changes (Roush et al., 2007; Van Bogaert et al., 2011), vegetation cover (Herrero et al., 2017, Masubelele et al., 2015, Rhemtulla et al., 2002, Hendrick and Copenheaver, 2009, Manier and Laven, 2002), costal habitats (Reimers et al., 2014), plant phenology (Julitta et al., 2014; Luo et al., 2018; Moore et al., 2016; Snyder et al., 2016), accuracy assessment (Kolecka et al., 2015) and for the study of general landscape changes (Kaim, 2017; Kull, 2005; Nüsser, 2001; Puschmann et al., 2006; Sanseverino et al., 2016). For a comprehensive overview of the broad application of repeat photography in the natural science, we refer to the work of Webb et al. (2010).

The majority of these studies use repeat photographs as qualitative data, often with the intention to support the results from field measurements or other remote sensing data. One reason is that missing geographical information, highly variable image content and scale as well as perspective issues make it difficult to perform quantitative and spatial analysis on oblique photographs. In general, ground-based photographs suffer from perspective distortion, high interclass variation, varying illumination and background clutter (Clark and Hardegree, 2005; Kull, 2005).

Despite these limitations, there have been attempts to retrieve quantitative information directly from photographs. Hall (2001) and Roush et al. (2007) applied a rectangular grid on top of the photographs to calculate vegetation cover percentages. Clark and Hardegree (2005) used point sampling along randomly placed horizontal transects through the image. They classified each image manually into cover types and introduced image cover as a quantitative measure. Some ecological studies combined repeat photographs with in-situ field measurements to achieve quantitative results (Hoffmann and Todd, 2010; Masubelele et al., 2015; McClaran et al., 2010). More recently, the use of monoplotting software offered the possibility to assigning real world coordinates to photographs in order to use them for geospatial analysis (Bozzini et al., 2012; Conedera et al., 2013; Stockdale et al., 2015). Fortin et al. (2018) retrieved class-specific land cover estimates from repeat photographs taken within the *Mountain Legacy Project* and compared results with Landsat classifications. Common to all these studies is that the classification step is performed manually, usually by drawing polygons around specific landscape elements followed by a visual interpretation. Although there exists a wide range of automatic segmentation and classification approaches for aerial and satellite imagery (Blaschke, 2010; Tewkesbury et al., 2015), these methods do not work in the same way for ground-based photographs due to the very oblique perspective.

Machine learning, and particularly deep learning, has evolved into the most commonly used approaches for the automatic classification of digital images (LeCun et al., 2015). The major advantage of deep learning is that the time-consuming and complex step of previous feature extraction becomes unnecessary. Instead, the model learns and extracts the relevant features itself during the training process. The major drawback of deep learning is that large amounts of labeled training data are required (Kamilaris and Prenafeta-Boldú, 2018). Among the deep learning architectures, convolutional neural networks (CNN) are particularly suitable for image analysis due to their ability to extract spatial features. CNNs have proven to be quite powerful in performing different tasks such as object detection (Everingham et al., 2010; Tompson et al., 2014), classification (Traore et al., 2018, Xu et al., 2017, Amara et al., 2017, Lu et al., 2017, Han et al., 2018) and semantic segmentation (Chen et al., 2014; Shelhamer et al., 2016).

However, many automatic detection and classification tasks in computer vision focus on rather easily distinguishable objects like humans, roads or vehicles (Krizhevsky et al., 2017). While automatic vegetation classification is a common task in remote sensing (Längkvist et al., 2016; Zhang et al., 2018), there are few studies based on terrestrial RGB photographs. Harbaš et al. (2018) used a fully convolutional network (FCN) to detect and segment roadside vegetation for the navigation of autonomous vehicles. Buscombe and Ritchie (2018) combined a CNN with conditional random fields (CFR) to classify and segment landscape classes in images from different data sources, among them also oblique photographs. However, most of these studies handle vegetation as a single class, while less emphasis has been put on distinguishing between different vegetation types in landscape photographs. In contrast to aerial and satellite remote sensing, landscape repeat photography operates only in the visible spectrum, which strongly limits the possibilities to differentiate between different species (Harbaš et al., 2018). Nevertheless, the automatic identification of single plant species in simple photographs (e.g. taken with mobile phones) has made great progress in recent years (Wäldchen et al., 2018). Unfortunately, these advanced classifiers are limited to detailed close-up photographs and thus, not yet suitable for landscape photographs.

More species-specific classifications based on landscape photographs can be found in precision agriculture, for example, for weed detection or ground cover estimations (Kamilaris and Prenafeta-Boldú, 2018; Milioto et al., 2017; Skovsen et al., 2017). Some studies in computer vision dealing with ‘scene parsing’ use separate classes for grasses and trees (Farabet et al., 2013). For example, Bosch et al. (2007) tried to separate tree vegetation from grass vegetation in natural images using a probabilistic classifier. Zhang et al. (2016) applied spatial contextual superpixel models to differentiate between different vegetation types along roads.

While research is striving towards new aerial and satellite remote sensing technologies, we are sitting on enormous quantities of valuable photographic material from the ground perspective. The potential for quantitative research represented by this material remains largely unexploited. Thanks to their potential for high level of detail, repeat photographs provide extensive information on landscape change ranging from a broad landscape level down to the species level. By supplementing conventional remote sensing imagery and field measurements with ground-level repeat photography, we are able to gain a more holistic view of landscapes. However, the retrieval of quantitative information from landscape photographs is still challenging. The automatic classification of image content presents a first step to use large quantities of photographs more efficiently and to analyze them in a standardized way.

Forest regrowth is a central topic in landscape monitoring and one of the predominant processes covered by repeat photography. Hence, this study will focus on the detection of typical regrowth vegetation in photographs. The gradual transition from open grassland to forest makes the drawing of distinct lines between vegetation classes difficult. To our knowledge, no study has tested an automatic approach to distinguish between woody vegetation and non-woody vegetation in repeat landscape photographs.

The main objective of our study is to test the application of a CNN for the automatic recognition of woody regrowth vegetation in repeat landscape photographs. This can be useful for analyzing large quantities of photographs, quantifying changes between repeat photographs and help to identify general trends of landscape change. To reach this objective, we define the following subordinate objectives:

1. We evaluate the performance of the trained CNN for the automatic classification of woody vegetation in repeat landscape photographs.
2. We test how the automatic classification performs in quantifying changes in woody vegetation compared to the manual classification.

2. Material and methods

2.1. Taking repeat photographs

There exist a large number of local and regional repeat photography projects, but quality and precision of the used photographic method is varying considerably. Repeat photography as a scientific approach is more than just taking a second snapshot in time. It requires careful preparation and a special consideration of the physical conditions. To achieve comparable image pairs, lighting, weather and seasonal conditions should be as similar as possible. In older photographs, where season or time is unknown, this information can be roughly estimated through visual interpretation of shadow cast and plant phenology. Unfortunately, weather and lighting conditions are in practice difficult to replicate since field work needs to be planned beforehand, budgets are limited and schedules are usually tight (Puschmann and Dramstad, 2002).

More important than the physical aspects is that repeat photographs have to be taken from exactly the same location in order to avoid shifts in the perspective. This point from where a picture is taken is called the vantage point (Klett et al., 1984). Mislocating the vantage point by as little as half a meter can cause visible mismatch between repeat photographs. Depending on the origin of the initial photograph, the reconstruction of the original photo location can be more or less challenging. Historical photographs retrieved from private and public archives often contain only scarce information on date and place, which makes the identification of the original location time-consuming, if not impossible. However, distinct landmarks and local knowledge may help to identify the original location. In contrast, newer initial photographs, taken with the intention of repeating them in the future, already contain information on GPS-location, date and time, which allows revisiting the location easily (Puschmann et al., 2018).

Even if coordinates are known, the spatial error of mobile GPS devices does require some effort to find the exact location. In order to get as close as possible to the original vantage point, we use a method which has been progressively developed through the *"Tilbakeblikk"*-project by Puschmann et al. (2006). This method is based on overlaying the initial photograph with a rectangular grid that corresponds to the internal grid of the camera (Fig. 1).

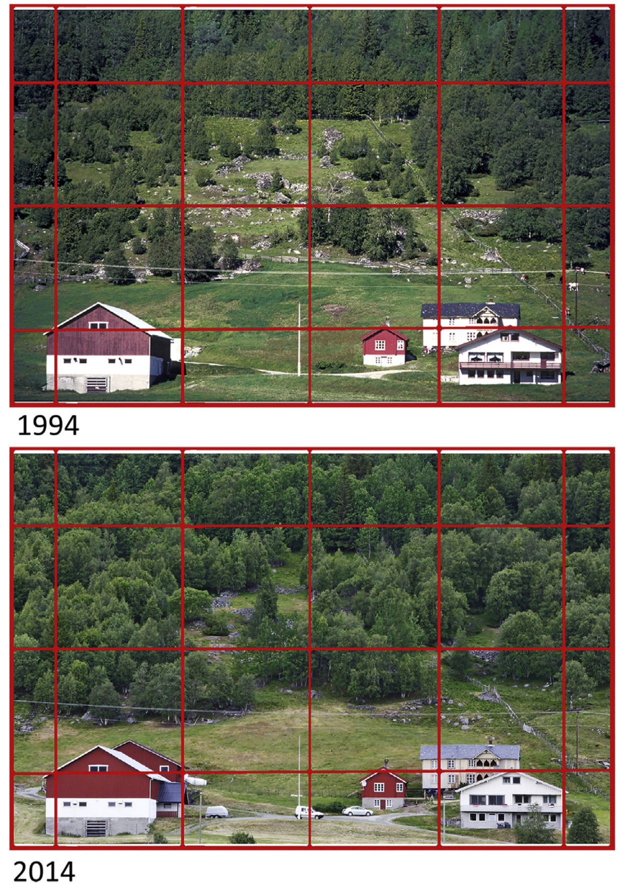

Fig. 1. Vantage point adjustment method. The initial photograph is overlaid with a rectangular grid corresponding to the internal grid of the camera. The grid allows checking crossing points and lines with the current view to obtain repeat photographs of high concordance. The photographs are taken in Hemsedal, Buskerud County, Norway showing forest regrowth on extensively grazed pastureland.

Lines and crossing points allow checking the current view against the spatial arrangement of objects and landscape elements in the old photograph. The camera position is adjusted accordingly until all check points in the current angle of view match the initial photograph. This method allows us to reconstruct the original photo location with high precision. As a result we achieve repeat photographs with a high degree of concordance, which in turn is an essential prerequisite for further overlay analyses. Until now, the grid has been applied manually on top of each photograph using Adobe Photoshop™ software. During this study, we developed a program which is able to perform this task automatically. In the field, we use a paper printout of the overlaid initial photograph together with a mobile GPS-device and a compass.

After the correct vantage point has been identified, we record GPS-coordinates, focal length, viewing direction, time and date as well as further important characteristics describing the vantage point. Photographs are always taken using a tripod-mounted single-lens reflex camera (SLR). The focal length is chosen in accordance with that one used for the older image. However, using a slightly smaller focal length than in the original photo, provides more room for cropping operations during post-processing.

In the field, we often face the problem that shrubs and trees have grown up in front of the vantage point, which then block the view partially or completely. We still aim to retake the photograph as these images convey a message. Another simple reason is that future revisits may reveal that the same area has been reopened again. Back in office, the raw images are edited as little as possible to maintain the natural representation of the landscape. Precise image matching of the initial photograph and the new photograph is performed through semi-transparent overlay in Adobe Photoshop™. The final image pairs are arranged vertically to allow easy comparison (Fig. 2).

2.2. Data material

As part of the Norwegian monitoring program for agricultural landscapes, the Norwegian Institute of Bioeconomy Research (NIBIO) has been working with repeat photography since 1998 (Dramstad et al., 2002; Puschmann and Dramstad, 2002). The related project *"Tilbakeblikk – Norwegian landscapes in retrospect"* supplements the program by retaking landscape photographs from all over Norway dating back until the mid-19th century. A selection of repeat photographs taken within the *Tilbakeblikk*-project is publicly available at www.tilbakeblikk.no (NIBIO, 2018). At present, NIBIOs photo archive contains > 3500 repeat photographs illustrating changes in the Norwegian cultural landscape. Given the limited spectral information that can be captured from black-and-white photographs, we restricted the study to RGB color photographs.

For our study, we used two independent sets of photographs. The first set contained 50 single landscape photographs with varying image content, scale and illumination taken by the authors. These photographs were used to collect a broad range of training samples. The second set consisted of seventeen image pairs (= 34 repeat photographs), which were used to test how the trained classifier performs on unseen data (Fig. 2). We had to limit the number of image pairs to seventeen, because their manual classification for model evaluation is

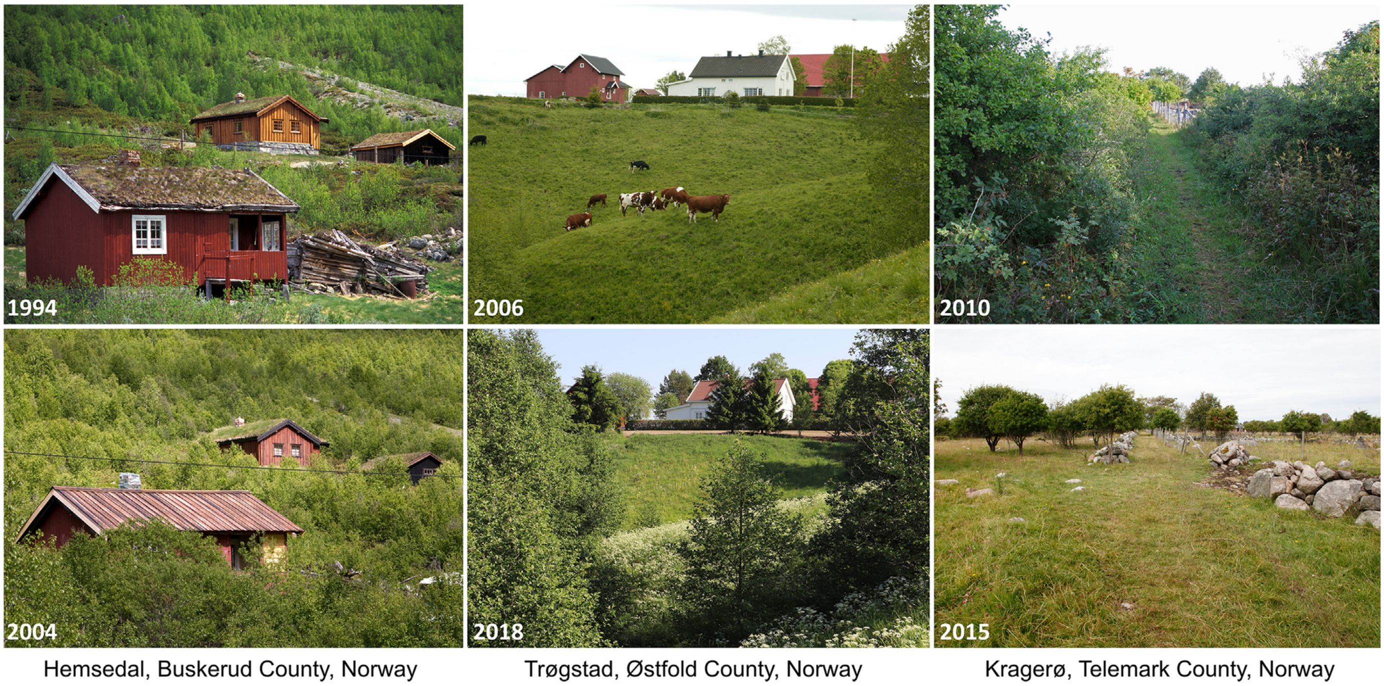

Fig. 2. Examples of the repeat photographs used in this study. Forest regrowth is one of the predominant ecological processes in repeat landscape photographs (left and middle). However, occasionally photographs capture also the reopening of regrown land (right).

time-consuming. Except for two initial images originating from private archives, all repeat photographs used in this study were taken by Oskar Puschmann in the period 1994 to 2018.

2.3. Data preparation

From the first set of training photographs, we extracted a total of 57,960 samples with a tile size of 50 × 50 pixels and labeled them manually. Since our focus lies on woody vegetation, we decided to perform a binary classification with woody vegetation as the positive class. The negative class contains all other content such as grasses, herbs, open soil, buildings, humans, animals, sky, water, stone and asphalt. The data set is nearly class-balanced comprising 28,080 samples (48.4%) of the positive class and 29,880 samples (51.6%) of the negative class. Example tiles from the training set are shown in Fig. 3, which illustrates the high intra-class variability.

To prevent spatial autocorrelation between neighboring tiles, the tiles retrieved from all 50 training photographs were collected in one folder and shuffled, before they were split into training, validation and test set. In this regard, the validation set is used during training for the fine tuning of hyperparameters and model selection (Hastie et al., 2009). Only after the best model setup has been found, its prediction error is assessed once again on the test set. For splitting the whole set of samples into the three sets, we first took a random sample of 10% from the total number of tiles as validation set and then a further sample of 10% as test set. The remaining 80% of the tiles were used as training set. To further increase the number of sample tiles, we performed data augmentation on the training set. Data augmentation is a common practice in machine learning to artificially increase the number of training samples by applying slight transformation on the original data, e.g. horizontal flip, rotation, scaling, brightness changes, shearing or zooming.

The second set of photographs was used for testing the model performance on new and more realistic data. When whole images needs to be classified with CNN, there arise two problems. First, classical CNNs assign only a single category on the input data and do not provide a classification with distinct boundaries. Second, the CNN takes only a fixed input size. To overcome these limitations, we split each photograph into 50 × 50 pixel tiles to match the size of the training samples. Before splitting, all photographs were resized to a maximum width of 2500 pixels to reduce computing time.

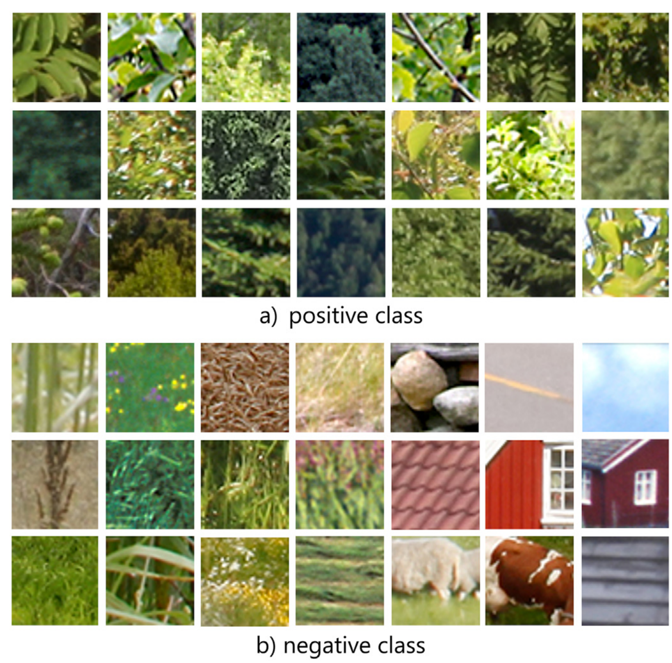

Fig. 3. Examples of the 50 × 50 pixel sample tiles for training. Woody vegetation is used as positive class (a), the negative class (b) contains grass, open soil, buildings, human, animals, sky, water, stone and asphalt. Note that many tiles of grassy and herbaceous plants look very similar to tiles of woody plants.

2.4. Convolutional neural network setup and training

CNNs are composed of three main components: convolutional

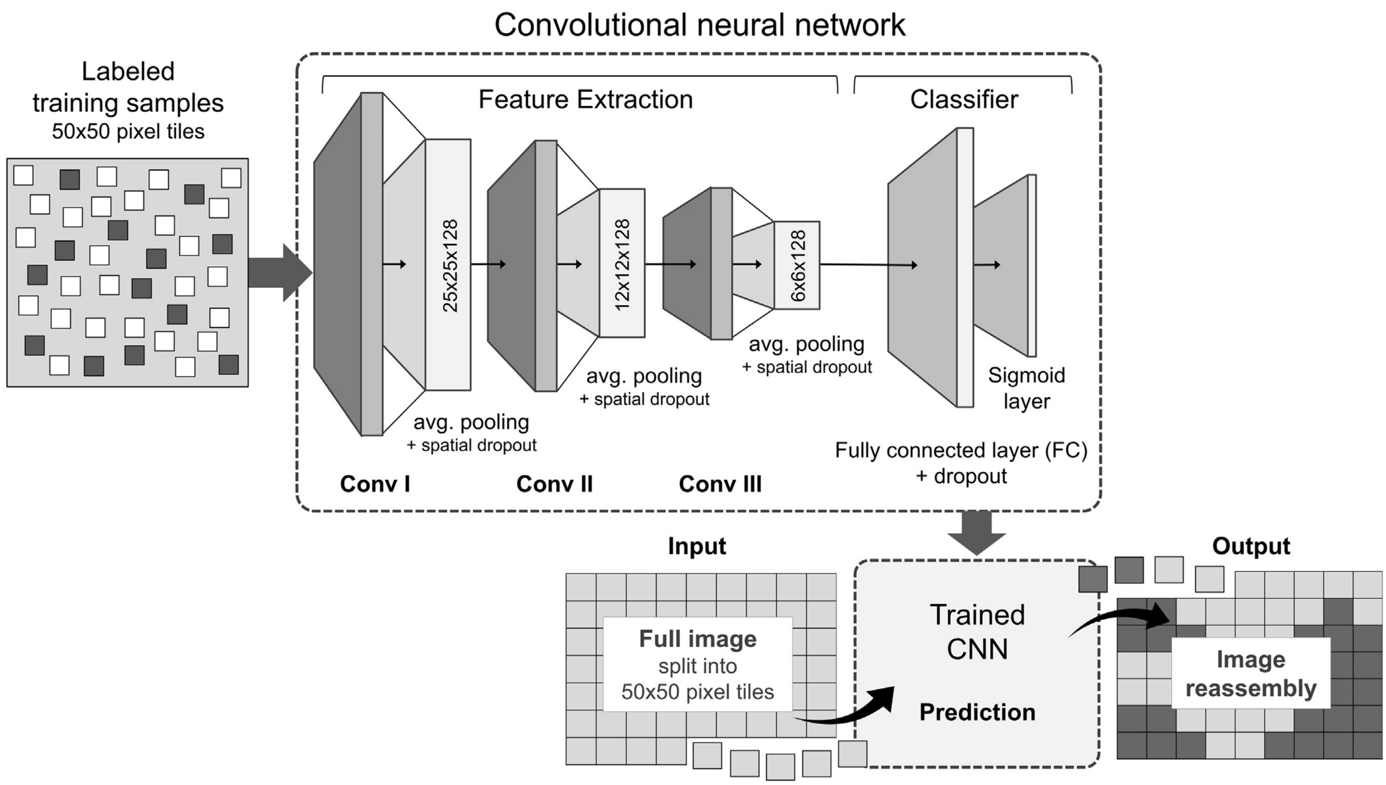

Fig. 4. Illustration of the CNN architecture and the classification process. The network was trained with manually labeled samples with a tile size of 50 × 50 pixel. The trained classifier was then applied on whole repeat photographs, which were also split into 50 × 50 pixel tiles. The classifier predicts the output for each single tile and reassembles them to the original image size.

layers, pooling layers and fully connected layers (Voulodimos et al., 2018). The first two components are responsible for automatic feature extraction by applying a large number of different filters on the input data. This process of feature extraction is performed on multiple levels, whereby the output of each level is the input to the following. From level to level, the extracted features increase in complexity - from rather simple features (e.g. edges) on the lowest level to more complex features on the highest level (Gu et al., 2018). By passing large quantities of labeled training data through the network, the model successively learns to recognize the relevant features, which are necessary to distinguish between classes.

For the automatic classification of woody vegetation, we developed a CNN consisting of three convolutional layers and one fully connected layer (Fig. 4). Each of the three convolutional layers was filtered with 128 kernels of size 3 × 3. Average pooling with a 2 × 2 filter was performed on each convolutional layer. Besides the conventional dropout on the fully connected layer (dropout rate = 0.7), we additionally applied spatial dropout (Tompson et al., 2014) on each convolutional layer (dropout rate = 0.3). This was in our case more successful in preventing overfitting and improving generalization. Dropout was applied on the training set only. We added a sigmoid function to the final layer, which is responsible for the binary classification. The optimal structure of the CNN was determined through a heuristic trial-and-error process. We implemented the CNN using the R-package "R Interface to keras" (Chollet and Allaire, 2017) and TensorFlow backend. The R code for model setup is available as supplementary material (Supplemental 1).

Although the use of pre-trained models has proven to be advantageous for many image recognition tasks (Weiss et al., 2016), we decided to train the model from scratch. We recognized that our specific classification problem was too different from the pre-trained networks. Moreover, we considered the number of labeled samples available (nearly 30,000 per class) as large enough for training our own network.

We trained the CNN using ReLU activation for the convolutional layers and an exponential linear unit (ELU) activation on the fully connected layer (Clevert et al., 2015). As optimizer we chose the adaptive ADAM (Adaptive Moment Estimation) with a learning rate of $lr = 0.0001$. Training was performed with a batch size of 256 over 16,290 iterations (= 90 epochs).

In addition to our model for the classification of woody vegetation, we trained a second model (hereafter called CNN_{veg}) for the classification of all vegetation (woody, herbaceous and grassy vegetation). For this purpose, we aggregated the sample tiles for woody vegetation with samples for grassy and herbaceous vegetation into the positive class. The sample data for the negative class included now open soil, buildings, humans, animals, sky, water, stone and asphalt. Since grassy samples accounted for a large proportion of the negative class, the sample number for positive and negative classes was heavily imbalanced after redistribution (49,489 vegetation vs. 8404 non-vegetation samples). Since vegetation is overrepresented in most photographs, this imbalance is considered to be of limited importance.

2.5. Binary classification of whole photographs

In the next step, we tested how the trained CNN performs on new and more realistic data. For this purpose, we applied the classifier on the second set of photographs, containing 34 repeat photographs split into tiles. All 50 × 50 pixel tiles of the photographs are predicted by the trained classifier in a fixed order and then reassembled into a raster grid of the original size showing the classification result (compare Fig. 4). This grid-based approach is not only more suitable for the CNN, which requires identical input sizes during training and prediction, it is also more robust against slight differences between the image pairs, which may have occurred due to camera distortion.

For evaluation purpose, the seventeen image pairs were also classified manually. This has been done by overlaying each photograph in

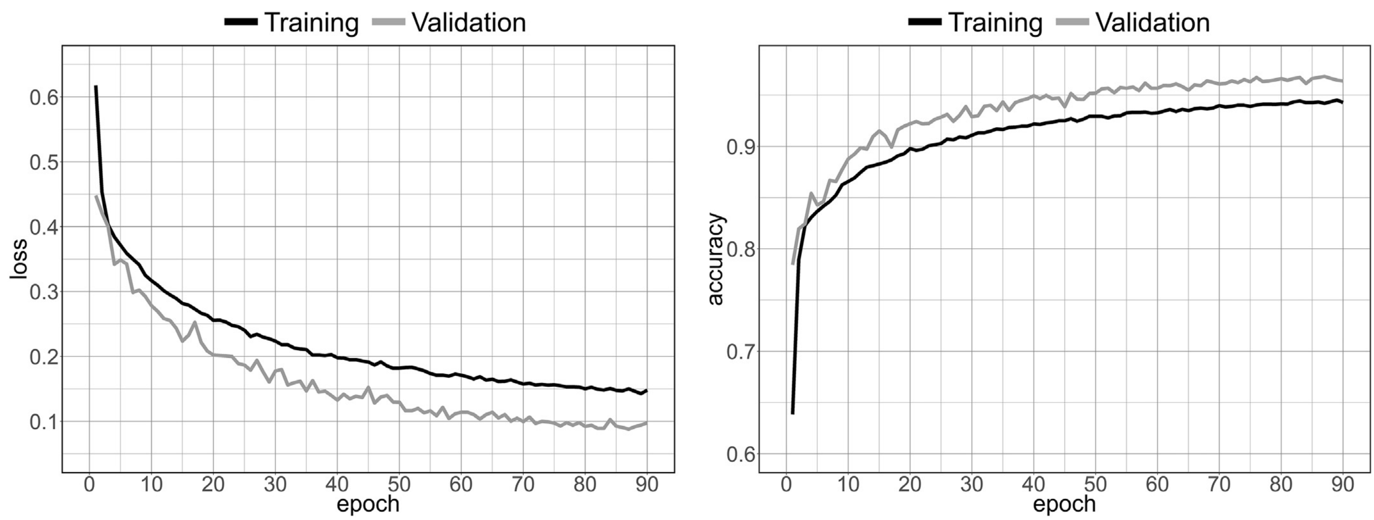

Fig. 5. Learning process for loss (left) and model accuracy (right) over 90 epochs. The gap between training and validation accuracy is caused by using dropout on the training data only.

Esri ArcMap (10.5) with a grid consisting of 50 × 50 pixel cells, which corresponds to the tile size from the automatic approach. The high resolution of the photographs allowed to visually distinguish between classes. Thus, the results from the manual approach can be considered as reference data. For most tiles, the class was unquestionable. Tiles containing both classes were classified as positive class, if woody vegetation covered at least 50% of the cell. When it comes to quantification, 2D ground photographs cannot provide real area coverage. Therefore, we used image cover as a quantitative measure to analyze changes between image pairs. According to Clark and Hardegree (2005) image cover is defined as percentage of the pixel count of specific cover types in a landscape photograph.

2.6. Model evaluation

All statistics were performed in R (version 3.3.1). The best model was chosen based on the two parameters accuracy and loss, whereby loss serves as a measure on how far model predictions differ from the actual class. Model accuracy and loss were calculated for both training and validation set. We tested the performance of the final model on two different data sets: 1) on individual tiles and 2) on whole repeat photographs. Prediction accuracy on individual tiles was calculated using the 5796 tiles from the test set (= 10%), which has been separated from the total number of samples before training. We evaluated the accuracy on whole repeat photographs based on the image pairs of the second set of photographs. The classification results for each of these 34 images were compared to the corresponding manual classification (reference data). A confusion matrix was prepared for each photograph individually. The confusion matrix consists of pixel numbers for true positives (*TP*), true negatives (*TN*), false positives (*FP*) and false negatives (*FN*). As accuracy metrics, we use Overall Accuracy (*OA*), Precision (*P*), Recall (*R*), F1 score (*F1*) and the Matthews correlation coefficient (*MCC*). Metrics are calculated as follows:

$$R = \frac{TP}{TP + FN} \tag{1}$$

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$F1 = 2 * \frac{P * R}{P + R} \tag{3}$$

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}} \tag{5}$$

The *MCC* measures the correlation between observed and predicted values (range −1 to 1) and is used to evaluate the quality of binary classifications (Vihinen, 2012). In contrast to *R, P* and *F1* measures, the *MCC* takes into consideration all four numbers in the confusion matrix. To prove that the model performs better than by chance, we compared our results with the accuracy reached by classifying the whole image as the majority class (No Information Rate, *NIR*) and by a class-weighted random guess (*wRG*).

In order to evaluate the overall performance, we averaged the mentioned metrics over all 34 images. To explain differences in classification accuracy between images, we divided them into two groups according to image quality (0 = low, 1 = high). The grouping was performed manually through visual assessment of lighting conditions and image resolution. We assume unequal variances and thus, use *Welch's t*-test for comparing the means of both groups.

To assess the usability of the automatic prediction for change analysis between single photographs, we compare the automatically measured change in image cover of woody vegetation with the actual change (based on manual approach). The difference between automatically measured change and actual change is reported as absolute error (percentage difference). We rank the seventeen image pairs according to the absolute error and calculate Spearman's rank correlation coefficient.

3. Results

3.1. Training and model performance on individual tiles

We stopped the learning process after 90 epochs, where the learning curve converged and loss values did not longer decrease. Fig. 5 illustrates the improvement in loss and accuracy during the training process. The fact that validation accuracy exceeds training accuracy is a common effect when dropout and data augmentation is applied on training data only. Consequently, the training set contains more difficult tiles than the validation set. Based on the validation set, the best model reached a maximum accuracy of 96.4% (*loss* = 0.13). Additionally, we evaluated the performance of the final model on a separate test dataset containing n = 5796 individual tiles, which were not used during the process of model selection. On these individual tiles, the model reached an accuracy of 96.7% (*loss* = 0.08).

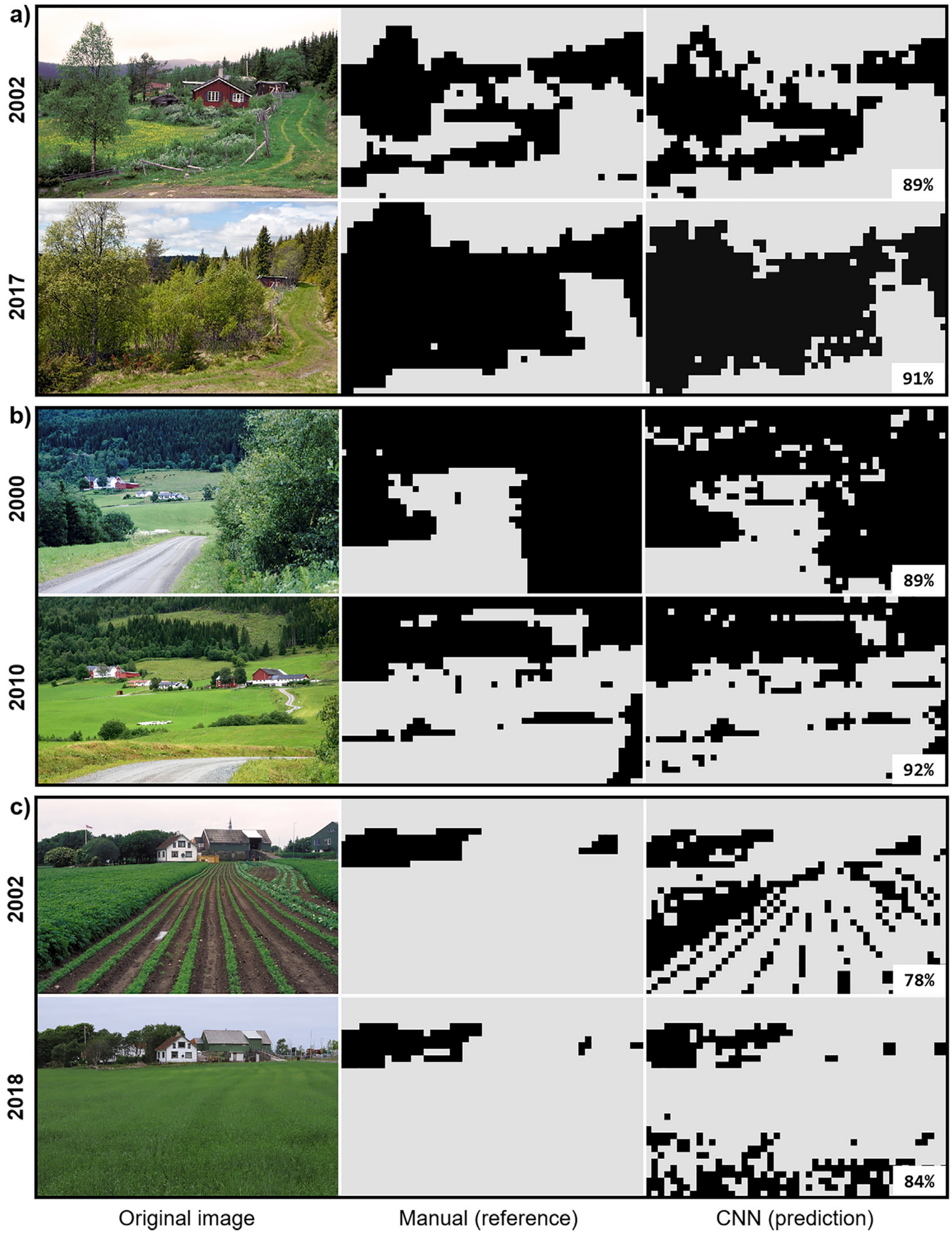

Fig. 6. Classification results for three image pairs. The illustration compares the classification of woody vegetation using a manual approach (middle) and the automatic CNN-based approach (right). Examples a) and b) show image pairs with good predictions, while c) illustrates an example with poor prediction. The broad-leafed vegetables in c) appear for the CNN like woody vegetation. Percentages represent the corresponding overall accuracies.

Table 1
Mean accuracy metrics for different classifiers when applied on whole repeat photographs (wRG = weighted random-guess, NIR = No Information Rate/majority-class, CNN_{woody} = woody vegetation, CNN_{veg} = all vegetation).

Classifier	Overall accuracy (%)	Min (%)	Max (%)	Recall	Precision	F1	MCC
wRG	64.3	49.0	98.2	–	–	–	–
NIR	73.1	54.3	99.5	–	–	–	–
CNN_{woody}	**87.7**	73.5	96.4	0.82	0.78	78.2	0.66
CNN_{veg}	**95.2**	86.7	99.5	0.96	0.97	96.5	0.87

3.2. Classification accuracy on whole photographs

After testing the model performance on individual tiles, the trained CNN was used to classify the tiles of whole photographs. Examined individually, overall accuracies varied considerably between images from *min* = 73.5% to max = 96.4%. Of the 34 images, 15 reached accuracies above 90%, another 15 between 80% and 90% and four images reached accuracies below 80%. *F1* scores ranged from 0.42 to 0.98. A full list of results for all 34 images is provided in Table A1, Appendix. Fig. 6 shows three image pairs in direct comparison to the manually classified photographs. In order to evaluate the overall performance of the CNN on the test images, we calculated averaged metrics from the individual results. Averaged over all 34 images, the model reached an overall accuracy of 87.7% and a F1 score of 78.2% (Table 1). With this, the model lies above the accuracy achieved by weighted random guess (wRG = 64.3%) and a majority class classification (NIR = 73.1%). Matthews correlation coefficient reached a value of 0.66. Most confusion existed between woody vegetation and other similar vegetation types such as grasses and herbaceous species. Some further confusion appeared between woody vegetation and wooden houses or cabins, which are very common landscape elements in Norway. The automatic classification results for all test images are available as supplemental material (Supplemental 2).

For comparison purposes, we trained the same network a second time but with vegetation as one class. In this case, overall accuracy for recognizing vegetation reached a mean of 95.2%, which is an increase of 7.8% compared to the classification of only woody vegetation (Table 1). Also mean F1 scores were considerably higher (96.5%) as were mean MCC (0.87). For individual photographs, accuracy ranged from min = 86.7% to max = 99.5%. Only four photographs reached accuracies below 90%, while nine photographs resulted in an almost perfect prediction with ≥98%. Particularly photographs with rather poor predictions for woody vegetation, reached considerably higher accuracies when vegetation is classified as one class. For example, the worst prediction (B14-2004) improved from OA = 73.5% to OA = 98.8%, which is a difference of 25%. Confusion matrices for both classification tasks are shown in Table 2 (woody vegetation) and Table 3 (all vegetation). Two examples for results achieved by the CNN_{veg} are shown in Fig. 7.

3.3. Influence of photo quality on classification accuracy

Since accuracies for classifying woody vegetation varied strongly between images, we grouped the images according to their image quality (image resolution and lighting). The grouping resulted in two unequal sample sizes with n = 13 for low quality images and n = 21 for high quality images. *Welch's t*-test showed that the means of both groups are significantly different on the 95% confidence interval (p = .045). The group of high quality images reached a mean accuracy of 89.3%, images with low quality 85.0% (Fig. 8a). We compared also means for classification accuracies when vegetation was scored as one class. In this case, the influence of image quality on accuracy was not significant (Fig. 8b).

3.4. Quantifying changes in image cover between repeat photographs

We compared the manually classified image pairs with each other and used image cover of woody vegetation as a quantitative measurement. The same was done with the automatically classified images. Fig. 9 shows the absolute error (%) between actual changes (manual approach) and changes measured with the CNN-based approach. Only six image pairs classified by the CNN were within the acceptable tolerance range of ± 5% from the reference. Twelve images were within a range of ± 10% difference. Due to one outlier with an error of 34.8%, the mean (μ = 9.17, sd = 8.17) diverged considerably from the median (M = 6.80, MAD = 5.34). By comparing the absolute error with classification accuracies of the single images, we found that the absolute error was highest for image pairs where the prediction was very different between the two images. Absolute error was also high for image pairs, where both images had poor classification accuracies. To test for correlation between predicted change and actual change, we ranked the seventeen image pairs corresponding to the absolute error. Spearman's rank correlation coefficient was rho = 0.917 (p < 0.001) (Fig. 10).

Table 2
Confusion matrix for the CNN_{woody} showing mean class accuracies for tiles from whole photographs.

CNN_{woody}	Woody	Non-woody
woody	**86.1**	11.5
non-woody	13.9	**88.5**

Table 3
Confusion matrix for the alternative CNN_{veg} showing mean class accuracies for tiles from whole photographs.

CNN_{veg}	Vegetation	Non-vegetation
Vegetation	**96.2**	8.1
Non-vegetation	3.8	**91.9**

4. Discussion

4.1. Performance of the CNN for classifying woody vegetation in repeat photographs

We have tested the performance of a convolutional neural network in terms of classifying woody vegetation in repeat landscape photographs. With an accuracy of 96.7% on individual tiles and 87.7% on whole photographs our network performed well in detecting woody vegetation. In comparison, Zhang et al. (2016) reached an accuracy of 79.8% on the class “tree” in natural roadside photographs using a spatial contextual superpixel model. Similar to our approach, the authors tested the classifier also on individual tiles and achieved an accuracy of 94.8% for the tree-class. Byeon et al. (2015) reached a class accuracy of 64.2% for trees in ground photographs using a LSTM Recurrent Neural Network (RNN). Similarly, Shuai et al. (2016) combined a CNN with a directed acyclic graphic RNN (DAG-RNN) for scene labeling and achieved an accuracy of 82.5% for the tree-class. Although our results acquired better results than these studies, it is obvious that the model is far from perfect. Even though we aggregated all negative classes into one class, there were some obvious class-specific misclassifications. Most confusion appeared between woody vegetation and grassy or herbaceous vegetation. This is not surprising since regrowth vegetation is often characterized by a very gradual transition from perennial species to woody species (Harper et al., 2005). Even for experts it can be difficult to draw a distinct line between woody vegetation and other vegetation types only based on photographs.

The fact that classification accuracy on individual test tiles was considerably higher than on whole photographs is an expected effect of the sampling process. When we collected the sample data, we selected primarily tiles, which could be clearly assigned to the positive or negative class. Predicting the tiles of a whole photograph leads inevitably to the inclusion of uncertain tiles, for which a distinct assignment is not possible. With this in mind, we want to underline the importance of testing classifiers not only on standardized image sets, but also on more complex real-world images to get a realistic impression of how well the classifier performs in practice.

In this study we find the wide range of accuracies across individual photographs particularly notable. It is likely that the high variation in image content, resolution, scale and illumination affected the classification accuracy in this specific task. We identified image quality as an important influencing factor. The advances in camera technology have led to a strong increase in image quality over time. As a consequence, older RGB photographs are usually of poorer quality than newer ones, which in turn affects the performance of the automatic classification approach. Low resolution leads to the disappearance of the textural characteristics of woody vegetation so that it can be easily confused with grassy vegetation. Particularly in RGB photographs, the textural information is of high importance to distinguish between different vegetation types. Clark and Hardegree (2005) address the same issues

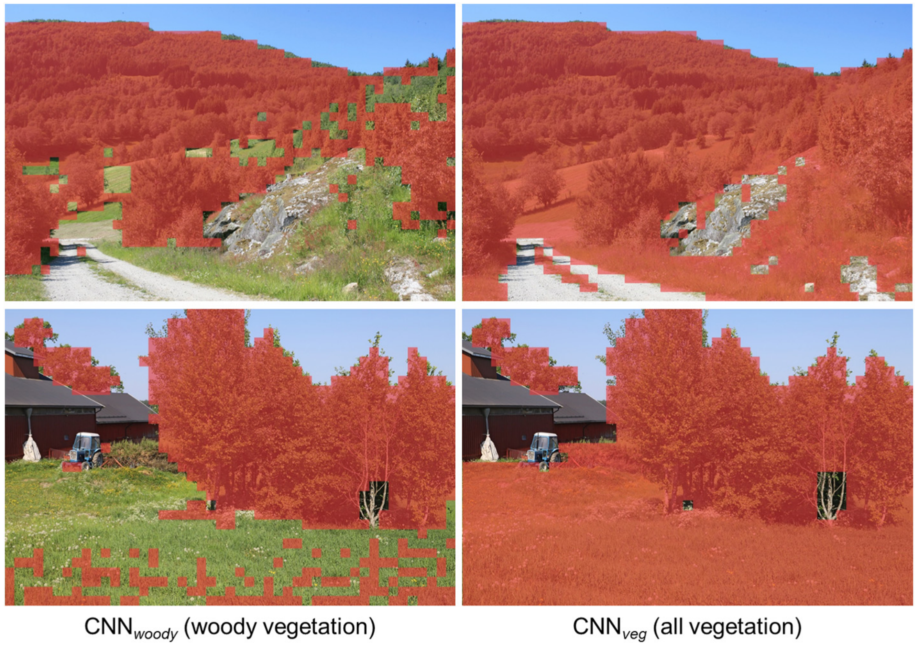

Fig. 7. Example of two photographs comparing the results from CNN$_{woody}$ and CNN$_{veg}$. In the left images, only woody vegetation is classified, the right images show classification of all vegetation.

regarding image quality when trying to quantify vegetation changes between repeat photographs using transect point sampling. Also Skovsen et al. (2017) experienced a higher rate of misclassification on blurred images, when they tried to distinguish clover from grasses and weeds.

Nevertheless, as our test with the alternative CNN$_{veg}$ demonstrated, image quality seems only to be relevant for distinguishing woody vegetation from other vegetation types. The CNN$_{veg}$ provided excellent results in the classification of total vegetation in nearly all test photographs (mean OA = 95.2%), even in photographs with low resolution. We assume that in this case color information alone was sufficient to successfully differentiate vegetation from non-vegetation. The high accuracy in classifying vegetation as a single class is in line with the results of Buscombe and Ritchie (2018), who reached accuracies between 89% and 96% in three data sets using 96 × 96 pixel tiles from RGB UAV images. Similarly, Harbaš et al. (2018) reached an accuracy of 96.3% using fully convolutional networks (FCN) to detect roadside vegetation. However, we need to emphasize, that our approach of assessing image quality only visually, is not ideal. To test the relationship between image quality and classification accuracy more systematically, an in-depth study of this particularly aspect should be carried out.

Besides image quality, we encountered further problems concerning the classification of woody vegetation in landscape photographs. The varying scale in photographs appeared to be a severe problem for the automatic classification approach. In particular, grassy and herbaceous vegetation close to the camera position was often wrongly classified as woody vegetation. Although the training data contained samples from both foreground and background, the network was not always able to distinguish properly between these two vegetation types. The most obvious reason for this is the strong visual similarity between woody species and herbaceous species. In this regard, broad-leafed crop plants are particularly challenging as Fig. 6c illustrates. Even for the human

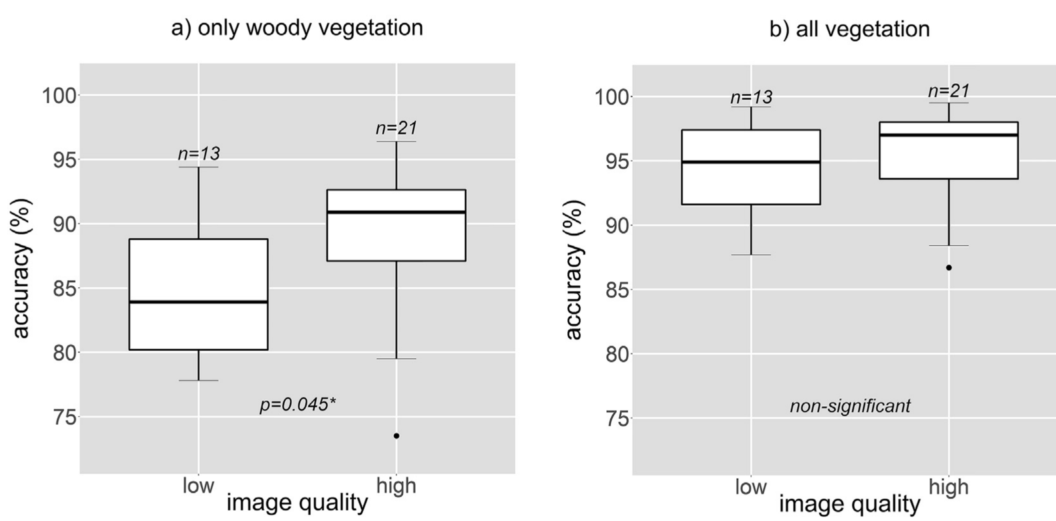

Fig. 8. Boxplot comparing accuracies reached on low quality and high quality photographs. Classification of woody vegetation (a) resulted in lower accuracies and a significant difference in means between low and high image quality. Classification of vegetation as one class (b) showed generally higher accuracies, but difference between image qualities was not statistically significant.

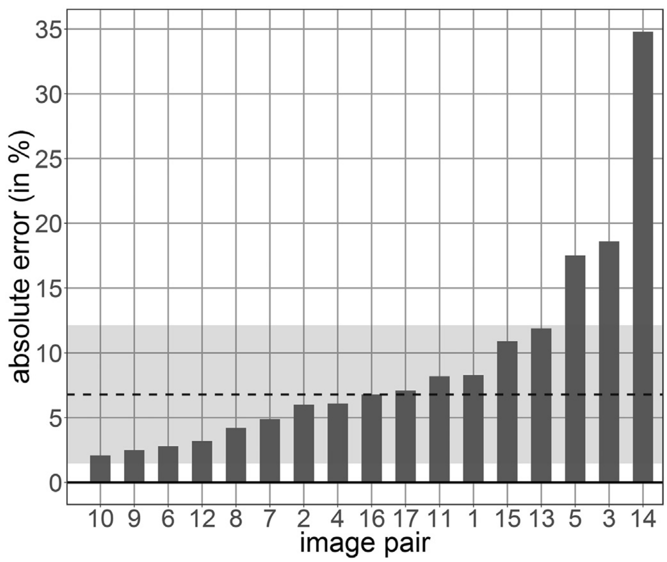

Fig. 9. Absolute error (%) in image cover change, which is calculated by subtracting predicted change (CNN-based classification) from actual change (manual classification). The dashed line and the grey area marks the median ± mean absolute deviation MAD (*median* = 6.8, MAD_{median} = 5.1).

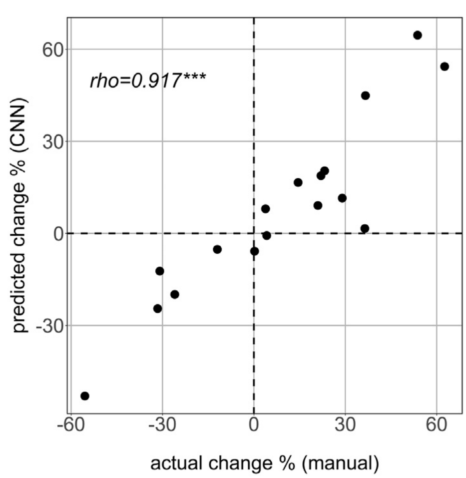

Fig. 10. Plot comparing change measured on basis of the automatic classification and actual change based on the manual classification. Spearman's rho of 0.917 indicates a relationship between the two variables.

eye the vegetables in this example may look like the leaves of woody plants. However, in contrast to the network, it is easy for us to understand the spatial context and to recognize the area as cultivated land. To solve this issue, it would be necessary that also the network understands the spatial context in order to classify vegetation accordingly. This might be reached by including more information from the whole image during training, but would also require another approach than the tiling method we used in our study. For example, Tang et al. (2015) were able to improve classification accuracy by including GPS coordinates of the location from where each image was taken. Through this information, the network was able to make predictions based on the geospatial context. Since also most repeat photographs contain GPS information, it should be possible to link these photographs to spatial information from existing land cover maps.

Regarding the background problem, one opportunity would be to limit the classification to the middle and foreground, while excluding the background. This requires a concise definition of where exactly the background begins. To retain the benefits of machine learning, an automatic approach would be preferable for this task. However, due to the high variability in image content, image quality, viewing angle and scale, it is complicated to develop a universal rule for defining the background in landscape photographs.

Also photographs with strong shadow cast posed to be a problem for the automatic approach. Although we could not examine this incident systematically due to the limited number of test images, we discovered in some of them that shadowed areas were wrongly classified as woody vegetation. This is most likely because the training data for woody vegetation naturally contain a large number of very dark samples. Moreover, shadow cast on grassland makes it more difficult to recognize the typical texture of grassy vegetation. By increasing the number of training samples from shadowed areas, the networks might improve its classification accuracy for such areas.

4.2. Quantifying changes in image cover: Manual vs. automatic approach

We quantified changes in image cover of woody vegetation between repeat photographs based on the automatic CNN-based approach and on the manual classification. The comparison of change statistics from both approaches show that changes based on the automatic classification deviates considerably from the changes recorded through the manual approach. Although the CNN-based classification provided good individual results on the majority of the 34 photographs, we experienced some crucial limitations when changes were quantified on basis of the automatic classification. The most obvious problem is that the accuracy of change statistics strongly depends on classification accuracy of both images. Even if woody vegetation in one image is predicted perfectly, poor prediction in the second image causes considerable bias in the analysis. Moreover, the limitation to RGB images offsets one of the major advantages of repeat photography, namely capturing historical changes over much longer periods of time than remote sensing imagery.

A typical challenge for performing change analysis on photographs is the presence of non-static objects such as humans, animals or vehicles. Thanks to our contextual understanding, humans are in most cases able to guess what lies behind these objects, but for the network the same task is nearly impossible. The unintended classification of temporary objects in one image leads to discrepancies when classification results are compared with the second image.

Similar to raster analyses in remote sensing, we experienced the grid-based approach as most comfortable for analyzing changes between photographs. Although a smooth delineation of classes would result in a visually more appealing representation, it is rather inconvenient for overlay analysis. Moreover, a grid-based representation is more robust against slight perspective differences between photographs. If photographs are not taken from exactly the same location as the initial image, a sharp delineation would result in a large number of sliver polygons, which do not represent real image cover changes.

However, even if the automatic classification was not capable of providing accurate cover percentages of woody vegetation, it was quite successful in recognizing clear trends of increase or decrease. Depending on the research goal, approximate statements about changes between repeat photographs might be sufficient, especially when supplemented with other data sources. For example, in monitoring, the regrowth of woody vegetation on agricultural land is often used as an indicator for land abandonment. For this purpose, it would be sufficient to differentiate between strong, medium and weak increase (or decrease) of woody vegetation in order to draw appropriate conclusions. If these trends are registered by several repeat photographs of an area, we can assume that it is a general trend for this specific area. On this basis we can initiate further, more accurate investigations of extent and underlying causes. Nevertheless, in research questions, where accurate cover percentages are essential or historical black-and-white photographs needs to be analyzed, the manual classification is supposed to be superior.

With further improvements in the future, the automatic classification of woody vegetation might reach a level where it is able to provide acceptable results in shorter time. Possible improvements may be achieved by further increasing the sample size, using more photographs with high quality or by testing alternative CNN-architectures. Another possibility poses transfer learning, meaning that knowledge gained from one classification task is reused for another similar task (Weiss et al., 2016). While training time and computing requirements are considerably reduced, fine tuning a pre-trained model to fit the new task can be time-consuming, in particular when the input data are different. Nevertheless, for many classification tasks, pretrained models such as VGG16 (Simonyan and Zisserman, 2014) or ResNet50 (He et al., 2015) have proven to be beneficial, particularly in cases where only a small amount of training data is available.

Finally, we want to emphasize, that image cover as a quantitative measurement should be interpreted with caution, since it provides only relative changes between photographs instead of real area changes (Michel et al., 2010). Distant objects appear much smaller, while objects in the foreground are overrepresented. Also blocking through regrowth of vegetation or other objects in the foreground can give a skewed impression of actual landscape changes. A promising way to overcome many of the mentioned limitations of photographs is a technique called monoplotting, which allows relating the content of photographs to real-world coordinates as illustrated by Stockdale et al. (2015). Although classification is performed manually, their work demonstrates impressively how spatial information can be retrieved from landscape photographs.

4.3. Repeat photography in landscape monitoring

Repeat photography has shown to be a valuable tool for documenting and communicating landscape change. The remarkable richness of details in ground-based photographs is hardly to reach with other remote sensing imagery. Still, if photographs are to be used in systematic landscape monitoring, they need to fulfill some basic requirements:

- GPS coordinates and viewing direction must be recorded
- environmental conditions (lighting, weather, season) and photo location (perspective) should be as similar as possible to the initial photograph
- photo locations should be evenly distributed over the study area and photos need to be taken in several compass directions to ensure spatial representativeness

Particularly the last aspect is difficult to achieve with repeat photographs, which are based on a historical initial photograph. Many historical photographs have been taken rather unsystematically along roads and popular trails with the intention to capture special landscape characteristics. Thus, such photographs are often not spatially representative and typically cover a rather small portion of an area. These issues are also addressed by Kull (2005) and Pickard (2002). Nevertheless, for research questions related to tourism, these photographs might still be of high value. However, as a direct consequence of an uneven distribution, photographs can easily give a false impression of ongoing landscape changes. For example, it is a common problem that regrowth of shrubs and trees close to the camera location blocks the view partially or completely. Although single photographs from a certain region may show regrowth in the image foreground, we can hardly conclude that this is a general trend in the larger area. In this case, repeat photographs should be supplemented with additional information, for example from aerial photographs. In general, we think the linkage of ground-based photographs to other remote sensing data has great potential and allows us to make use of the advantages of seeing landscape changes from both perspectives.

To avoid the aforementioned issues, an even distribution of photographs over the study area is essential. To achieve this, initial photographs need to be taken with the intention of being used for landscape monitoring. Only then can it be ensured that photo locations are evenly distributed and that they give a representative picture of the study area. Predefined photo points and the taking of photographs in all four compass directions ensure uniform area coverage. However, experiences described by Puschmann et al. (2018) document that additional "free" images are recommended in order to capture elements not covered with fixed photo directions. In the future, drones could also be used for the automatic acquisition of repeat photographs and help to achieve an even distribution of photo points.

Beside the drawbacks described, repeat photographs offer also a range of advantages. The greatest strength of photographs is their high information content, especially for historical periods where other remote sensing imagery did not exist. By combining aerial and satellite imagery with ground-based photographs, the research period can be expanded into the past by almost one century. Moreover, the high image resolution and the oblique view of photographs allow capturing changes not detectable from the vertical view, for example, local changes in species composition or facade changes of buildings. While aerial photographs perform well in measuring horizontal changes, oblique photographs can tell us much more about vertical changes, for example in vegetation structure (stratification) and height.

With respect to landscape monitoring, photographs are thus capable of providing a much earlier indication of ongoing changes. Another advantage is that the acquisition of ground-based photographs is more flexible and cost-efficient compared to aerial photography (Kull, 2005). Finally, the ground-level perspective of photographs is closer to people's natural perception of landscapes. This is a significant advantage for communicating landscape changes in public and to policy makers. Even without experience in aerial photo interpretation, ground-based photographs are easy to understand and quite powerful in stimulating contemplation and debate (Kull, 2005).

Although the possibilities for quantitative analyses may still be limited, we should not forget the main purpose of repeat photographs: to serve as a qualitative supplement to quantitative monitoring results retrieved through remote sensing and field measurements. Moreover, repeat photography plays an important role in raising awareness of the dynamic character of landscapes among policy makers and public. Repeat photography can further provide valuable feedback for decision makers on the effectiveness of environmental subsidies and maintenance measures.

5. Conclusion

We demonstrated the application of a convolutional neural network for the automatic detection woody vegetation in repeat landscape photographs. Our results showed that the differentiation of woody regrowth vegetation from other vegetation types is a much more challenging task than the classification of vegetation as a single class. We found that image quality is an important factor influencing the performance of the automatic approach, while factors such as shadow cast and varying scale were additional key factors. Although the network did not provide perfect predictions, we do find our results promising. Regarding the measurement of changes between image pairs, the CNN-model was capable of recognizing rough trends of increasing and decreasing woody vegetation, which can be useful information in many research questions. In cases, where more accurate measurements are essential, the manual classification was superior. However, with further improvements as proposed, the automatic classification of woody vegetation might provide reasonable results, which would enable a more efficient analysis of large numbers of photographs more efficiently.

Finally, we evaluated the usability of repeat photography in landscape monitoring. Results in this study illustrate that despite the large technological and methodical advances in image analysis, there are still difficulties in retrieving meaningful quantitative information from photographs. In order to examine further possibilities, our future research will focus on the geospatial analysis of photographs and their linkage to other remote sensing data. This might allow us to use the automatic classification from this study to retrieve real area coverage.

The results of our study can also be important for sharpening consciousness for how and under which conditions photographs for monitoring purposes should be taken in the future. Keeping in mind some of the mentioned issues such as illumination, shadow cast, spatial distribution and image resolution, the photographer is able to take photographs that are suitable for use in landscape monitoring.

Regardless of whether we can use photographs as quantitative data or not, based on our experiences, we consider repeat photography as a powerful tool in landscape research. Communicating landscape related issues to a broad audience and raising awareness of the dynamic character of landscapes are the main contributions of repeat photography. This awareness is essential to understand the consequences of decisions made by policy makers, management and farmers. Through this, repeat photography is no longer merely a glimpse into the past, it becomes even capable of forming the future.

Table A1
Classification results and accuracy measures for all images, sorted by overall accuracy (values above 90% are bold).

Image-Year	R	P	F1 (%)	MCC	OA (%) Woody vegetation	OA (%) All vegetation	Image quality
B01-2006	0.53	0.44	48.0	0.46	**96.4**	**97.6**	+
B09-2015	0.83	0.78	80.2	0.78	**95.5**	**96.9**	+
B11-2014	0.97	0.98	97.4	0.72	**95.3**	**99.5**	+
B03-2015	0.79	0.67	72.2	0.69	**94.4**	**94.3**	−
B04-2015[*]	0.17	0.01	2.7	0.03	**93.6**	**97.0**	+
B12-2015	0.91	0.97	94.1	0.87	**93.4**	**91.1**	−
B05-2018	0.95	0.90	92.7	0.86	**93.2**	**93.6**	+
B07-2006	0.84	0.84	84.0	0.80	**93.0**	**99.2**	−
B16-2011	0.98	0.93	95.4	0.78	**92.6**	**98.0**	+
B04-2009	0.85	0.87	85.9	0.81	**92.6**	**93.3**	+
B10-2009	0.90	0.95	92.7	0.85	**92.3**	**96.7**	+
B17-2010	0.93	0.86	89.4	0.83	**91.6**	**97.4**	+
B06-2017	0.89	0.96	92.2	0.82	**90.9**	**93.0**	+
B07-2017	0.73	0.90	80.7	0.75	**90.9**	**99.0**	+
B12-2009	0.86	0.86	85.9	0.78	**90.1**	**95.3**	+
B15-2006	0.68	0.73	70.6	0.64	89.8	**97.3**	+
B06-2002	0.82	0.87	84.5	0.76	88.8	**91.6**	−
B17-2000	0.89	0.94	91.6	0.74	88.6	**98.3**	−
B14-1994	0.96	0.83	89.4	0.76	87.6	**91.6**	+
B13-2005	0.90	0.89	89.6	0.74	87.4	**98.4**	+
B15-2018	0.98	0.86	91.6	0.67	87.1	**97.0**	+
B01-2018	0.94	0.77	84.4	0.73	86.2	**97.2**	+
B16-2016	0.87	0.92	89.1	0.70	86.1	88.4	+
B02-2018	0.82	0.31	44.7	0.44	84.0	**96.8**	−
B13-1999	0.95	0.73	82.3	0.70	83.9	**98.1**	−
B09-2010	0.87	0.89	87.7	0.64	83.8	**98.3**	+
B08-2002	0.76	0.99	85.9	0.66	81.7	**94.9**	−
B11-1994	0.79	0.64	71.1	0.58	81.3	**93.2**	−
B10-2001	0.69	0.79	73.5	0.58	80.2	87.7	−
B03-2010	0.54	0.93	68.4	0.59	80.0	**97.4**	−
B08-2017	0.78	0.94	85.6	0.54	79.5	86.7	+
B02-2002	0.81	0.23	36.3	0.36	78.0	88.1	−
B05-2002	0.94	0.41	57.4	0.52	77.8	**95.0**	−
B14-2004	0.71	1.00	82.9	0.43	73.5	**98.8**	+
Mean	**0.83**	**0.78**	**78.2**	**0.66**	**87.7**	**95.2**	

[*] Poor results due to generally low amount of positive class (woody vegetation) in this image.

Acknowledgments

This work was supported by the Norwegian Ministry of Agriculture and Food through the Research Council of Norway [nr. 194051]. We thank Geir-Harald Strand and Wenche E. Dramstad from NIBIO as well as Bernd Etzelmüller from the University of Oslo for their support during the project and their valuable comments on the final draft. We also appreciate the reviewers' suggestions and their great commitment to improve our work.

Disclosure statement

No potential conflict of interest was reported by the authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2019.01.012.

References

Amara, J., Bouaziz, B., Algergawy, A., 2017. A deep learning-based approach for banana leaf diseases classification. In: Mitschang, B., Nicklas, D., Leymann, F., Schöning, H., Herschel, M., Teubner, J., Härder, T., Kopp, O., Wieland, M. (Eds.), Datenbanksysteme für Business, Technologie und Web (BTW 2017) - Workshopband. Gesellschaft für Informatik e.V., Bonn (S. 79-88).

Amici, V., Maccherini, S., Santi, E., Torri, D., Vergari, F., Del Monte, M., 2017. Long-term patterns of change in a vanishing cultural landscape: a GIS-based assessment. Ecol. Informa. 37, 38–51. https://doi.org/10.1016/j.ecoinf.2016.11.008.

Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens. 65, 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004.

Bosch, A., Muñoz, X., Freixenet, J., 2007. Segmentation and description of natural outdoor scenes. Image Vis. Comput. 25, 727–740. https://doi.org/10.1016/j.imavis.2006.05.015.

Bozzini, C., Conedera, M., Krebs, P., 2012. A new monoplotting tool to extract georeferenced vector data and orthorectified raster data from oblique non-metric photographs. Int. J. Herit. Dig. Era 1, 499–518.

Buscombe, D., Ritchie, A., 2018. Landscape classification with deep neural networks. Geosciences 8, 244. doi.org/https://doi.org/10.3390/geosciences8070244

Byeon, W., Breuel, T.M., Raue, F., Liwicki, M., 2015. Scene labeling with LSTM recurrent neural networks, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Boston, MA, USA, pp. 3547–3555. doi.org/https://doi.org/10.1109/CVPR.2015.7298977

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2014. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. arXiv:1412.7062 [cs].

Chollet, F., Allaire, J.J., 2017. R interface to keras. GitHub. https://github.com/rstudio/keras (accessed 02.12.2018).

Clark, P.E., Hardegree, S.P., 2005. Quantifying vegetation change by point sampling landscape photography time series. Rangel. Ecol. Manag. 58, 588–597. https://doi.org/10.2458/azu_rangelands_v58i6_hardegree.

Clevert, D.-A., Unterthiner, T., Hochreiter, S., 2015. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). arXiv:1511.07289 [cs].

Conedera, M., Bozzini, C., Scapozza, C., Rè, L., Ryter, U., Krebs, P., 2013. Anwendungspotenzial des WSL-Monoplotting-Tools im Naturgefahrenmanagement. Schweiz. Z. Forstwes. 164, 173–180. https://doi.org/10.3188/szf.2013.0173.

Dramstad, W.E., Fjellstad, W.J., Strand, G.-H., Mathiesen, H.F., Engan, G., Stokland, J.N., 2002. Development and implementation of the Norwegian monitoring programme for agricultural landscapes. J. Environ. Manag. 64, 49–63. https://doi.org/10.1006/jema.2001.0503.

Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2010. The Pascal Visual Object Classes (VOC) Challenge. Int. J. Comput. Vis. 88, 303–338. https://doi.org/10.1007/s11263-009-0275-4.

Farabet, C., Couprie, C., Najman, L., LeCun, Y., 2013. Learning hierarchical features for scene labeling. IEEE Trans. Pattern Anal. Mach. Intell. 35, 1915–1929. https://doi.org/10.1109/tpami.2012.231.

Fjellstad, W.J., Dramstad, W.E., 1999. Patterns of change in two contrasting Norwegian agricultural landscapes. Landsc. Urban Plan. 45, 177–191. https://doi.org/10.1016/S0169-2046(99)00055-9.

Fortin, J.A., Fisher, J.T., Rhemtulla, J.M., Higgs, E.S., 2018. Estimates of landscape composition from terrestrial oblique photographs suggest homogenization of Rocky Mountain landscapes over the last century. Remote Sens. Ecol. Conserv. https://doi.org/10.1002/rse2.100.

Frankl, A., Nyssen, J., De Dapper, M., Haile, M., Billi, P., Munro, R.N., Deckers, J., Poesen, J., 2011. Linking long-term gully and river channel dynamics to environmental change using repeat photography (Northern Ethiopia). Geomorphology 129, 238–251. https://doi.org/10.1016/j.geomorph.2011.02.018.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., Chen, T., 2018. Recent advances in convolutional neural networks. Pattern Recogn. 77, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013.

Hall, F.C., 2001. Ground-based photographic monitoring. Gen. Tech. Rep. PNW-GTR-503. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR.

Han, D., Liu, Q., Fan, W., 2018. A new image classification method using CNN transfer learning and web data augmentation. Expert Syst. Appl. 95, 43–56. https://doi.org/10.1016/j.eswa.2017.11.028.

Harbaš, I., Prentašić, P., Subašić, M., 2018. Detection of roadside vegetation using Fully Convolutional Networks. Image Vis. Comput. 74, 1–9. https://doi.org/10.1016/j.imavis.2018.03.008.

Harper, K.A., Macdonald, S.E., Burton, P.J., Chen, J., Brosofske, K.D., Saunders, S.C., Euskirchen, E.S., Roberts, D., Jaiteh, M.S., Esseen, P.-A., 2005. edge influence on forest structure and composition in fragmented landscapes. Conserv. Biol. 19, 768–782. https://doi.org/10.1111/j.1523-1739.2005.00045.x.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, Springer Series in Statistics. Springer New York, New York, NY. https://doi.org/10.1007/978-0-387-84858-7.

Hastings, J.R., Turner, R.M., 1965. The Changing Mile. University of Arizona Press, Tucson.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition. arXiv:1512.03385 [cs].

Hendrick, L.E., Copenheaver, C.A., 2009. Using Repeat Landscape Photography to Assess Vegetation Changes in Rural Communities of the Southern Appalachian Mountains in Virginia. 29. Mountain Research and Development, USA, pp. 21–29. https://doi.org/10.1659/mrd.1028.

Herrero, H.V., Southworth, J., Bunting, E., Child, B., 2017. Using repeat photography to observe vegetation change over time in gorongosa national park. Afr. Stud. Quart. 17.

Hoffmann, T.M., Todd, S.W., 2010. Using fixed-point photography, field surveys, and GIS to monitor environmental change: An example from Riemvasmaak, South Africa. In: Webb, R.H., Boyer, D.E., Turner, R.M. (Eds.), Repeat Photography: Methods and Applications in the Natural Sciences. Island Press, Washington, pp. 46–56.

Julitta, T., Cremonese, E., Migliavacca, M., Colombo, R., Galvagno, M., Siniscalco, C., Rossini, M., Fava, F., Cogliati, S., Morra di Cella, U., Menzel, A., 2014. Using digital camera images to analyse snowmelt and phenology of a subalpine grassland. Agric. For. Meteorol. 198–199, 116–125. doi.org/https://doi.org/10.1016/j.agrformet.2014.08.007.

Kaim, D., 2017. Land cover changes in the polish Carpathians repeat photography. Carpath. J. Earth Environ. Sci. 12, 485–498.

Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: a survey. Comput. Electron. Agric. 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016.

Khan, S.F., Kamp, U., Owen, L.A., 2013. Documenting five years of landsliding after the 2005 Kashmir earthquake, using repeat photography. Geomorphology 197, 45–55. https://doi.org/10.1016/j.geomorph.2013.04.033.

Klett, M., 2010. Three methods of presenting repeat photographs. In: Webb, R.H., Boyer, D.E., Turner, R.M. (Eds.), Repeat Photography: Methods and Applications in the Natural Sciences. Island Press, Washington, pp. 32–45.

Klett, M., Manchester, E., Verburg, J., 1984. Second View: The Rephotographic Survey Project. Albuquerque University of New Mexico Press.

Kolecka, N., Kozak, J., Kaim, D., Dobosz, M., Ginzler, C., Psomas, A., 2015. Mapping secondary forest succession on abandoned agricultural land with LiDAR point clouds and terrestrial photography. Rem. Sens. 7, 8300–8322. https://doi.org/10.3390/rs70708300.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 84–90. https://doi.org/10.1145/3065386.

Kull, C.A., 2005. Historical landscape repeat photography as a tool for land use change research. Norsk Geografisk Tidsskrift – Norw. J. Geogr. 59, 253–268. https://doi.org/10.1080/00291950500375443.

Längkvist, M., Kiselev, A., Alirezaie, M., Loutfi, A., 2016. Classification and segmentation of satellite orthoimagery using convolutional neural networks. Remote Sens. 8, 329. doi.org/https://doi.org/10.3390/rs8040329

Lasanta, T., Nadal-Romero, E., Arnáez, J., 2015. Managing abandoned farmland to control the impact of re-vegetation on the environment. The state of the art in Europe. Environ. Sci. Pol. 52, 99–109. https://doi.org/10.1016/j.envsci.2015.05.012.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Lu, H., Fu, X., Liu, C., Li, L., He, Y., Li, N., 2017. Cultivated land information extraction in UAV imagery based on deep convolutional neural network and transfer learning. J. Mt. Sci. 14, 731–741. https://doi.org/10.1007/s11629-016-3950-2.

Luo, Y., El-Madany, T.S., Filippa, G., Ma, X., Ahrens, B., Carrara, A., Gonzalez-Cascon, R., Cremonese, E., Galvagno, M., Hammer, T.W., Pacheco-Labrador, J., Martín, M.P., Moreno, G., Perez-Priego, O., Reichstein, M., Richardson, A.D., Römermann, C., Migliavacca, M., 2018. Using near-infrared-enabled digital repeat photography to track structural and physiological phenology in Mediterranean tree–grass ecosystems. Remote Sens. 10, 1293. doi.org/https://doi.org/10.3390/rs10081293

Manier, D.J., Laven, R.D., 2002. Changes in landscape patterns associated with the persistence of aspen (Populus tremuloides Michx.) on the western slope of the Rocky Mountains, Colorado. For. Ecol. Manag. 167, 263–284. https://doi.org/10.1016/S0378-1127(01)00702-2.

Masubelele, M.L., Hoffman, M.T., Bond, W.J., 2015. A repeat photograph analysis of long-term vegetation change in semi-arid South Africa in response to land use and climate. J. Veg. Sci. 26, 1013–1023. https://doi.org/10.1111/jvs.12303.

McClaran, M.P., Browning, D.M., Huang, C., 2010. Temporal dynamics and spatial variability in desert grassland vegetation. In: Webb, R.H., Boyer, D.E., Turner, R.M. (Eds.), Repeat Photography: Methods and Applications in the Natural Sciences. Island Press, Washington, pp. 145–166.

Michel, P., Mathieu, R., Mark, A.F., 2010. Spatial analysis of oblique photo-point images for quantifying spatio-temporal changes in plant communities. Appl. Veg. Sci. 13, 173–182. https://doi.org/10.1111/j.1654-109X.2009.01059.x.

Milioto, A., Lottes, P., Stachniss, C., 2017. Real-time blob-wise sugar beets vs weeds classification for monitoring fields using convolutional neural networks. In: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, IV-2/W3, pp. 41–48. https://doi.org/10.5194/isprs-annals-IV-2-W3-41-2017.

Molnia, B.F., 2010. Repeat photography of Alaskan glaciers and landscapes from ground-based photo stations and airborne platforms. In: Webb, R.H., Boyer, D.E., Turner, R.M. (Eds.), Repeat Photography: Methods and Applications in the Natural Sciences. Island Press, Washington, pp. 59–76.

Moore, C.E., Brown, T., Keenan, T.F., Duursma, R.A., van Dijk, A.I.J.M., Beringer, J., Culvenor, D., Evans, B., Huete, A., Hutley, L.B., Maier, S., Restrepo-Coupe, N., Sonnentag, O., Specht, A., Taylor, J.R., van Gorsel, E., Liddell, M.J., 2016. Reviews and syntheses: Australian vegetation phenology: new insights from satellite remote sensing and digital repeat photography. Biogeosciences 13, 5085–5102.

NIBIO, 2018. Tilbakeblikk: Norwegian landscapes in retrospect. Available at http://www.tilbakeblikk.no/ (accessed 10.12.2018).

Nüsser, M., 2001. Understanding cultural landscape transformation: a re-photographic survey in Chitral, eastern Hindukush, Pakistan. Landsc. Urban Plan. 57, 241–255. https://doi.org/10.1016/S0169-2046(01)00207-9.

Pickard, J., 2002. Assessing vegetation change over a century using repeat photography. Aust. J. Bot. 50, 409–414. https://doi.org/10.1071/BT01053.

Puschmann, O., Dramstad, W.E., 2002. Documenting landscape change through fixed angle photography. In: NIJOS/OECD Workshop on Agricultural Landscapes Indicators, Norway, Oslo.

Puschmann, O., Dramstad, W.E., Hoel, R., 2006. Tilbakeblikk: norske landskap i endring [Flashback: Norwegian landscapes in retrospect]. Tun Forlag, Oslo.

Puschmann O., Eiter, S., Fjellstad, W.J., Krøgli, S.O., 2018. Preparing future flashbacks – repeat photography as a method in landscape monitoring. NIBIO POP. 2018, 4 (24).

Reimers, B., Griffiths, C., Hoffman, M., 2014. Repeat photography as a tool for detecting and monitoring historical changes in South African coastal habitats. Afr. J. Mar. Sci. 36, 387–398. https://doi.org/10.2989/1814232X.2014.954618.

Rhemtulla, J.M., Hall, R.J., Higgs, E.S., Macdonald, S.E., 2002. Eighty years of change: vegetation in the montane ecoregion of Jasper National Park, Alberta, Canada. Can. J. For. Res. 32, 2010–2021. https://doi.org/10.1139/x02-112.

Roush, W., Munroe, J.S., Fagre, D.B., 2007. Development of a spatial analysis method using ground-based repeat photography to detect changes in the alpine treeline ecotone, Glacier National Park, Montana, U. S. A. Arctic Antarct. Alp. Res. 39 (2), 297–308.

Sang, N., Dramstad, W.E., Bryn, A., 2014. Regionality in Norwegian farmland abandonment: Inferences from production data. Appl. Geogr. 55, 238–247. https://doi.org/10.1016/j.apgeog.2014.09.015.

Sanseverino, M.E., Whitney, M.J., Higgs, E.S., 2016. Exploring landscape change in mountain environments with the mountain legacy online image analysis toolkit. Mt. Res. Dev. 36, 407–416. https://doi.org/10.1659/mrd-journal-d-16-00038.1.

Shelhamer, E., Long, J., Darrell, T., 2016. Fully Convolutional Networks for Semantic Segmentation. arXiv:1605.06211 [cs].

Shuai, B., Zuo, Z., Wang, B., Wang, G., 2016. DAG-Recurrent Neural Networks for Scene Labeling. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Las Vegas, NV, USA, pp. 3620–3629. https://doi.org/10.1109/CVPR.2016.394.

Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs].

Skovsen, S., Dyrmann, M., Mortensen, A., Steen, K., Green, O., Eriksen, J., Gislum, R., Jørgensen, R., Karstoft, H., 2017. Estimation of the botanical composition of clover-grass leys from RGB Images using data simulation and fully convolutional neural networks. Sensors 17, 2930. doi.org/https://doi.org/10.3390/s17122930

Snyder, K., Wehan, B., Filippa, G., Huntington, J., Stringham, T., Snyder, D., 2016. Extracting plant phenology metrics in a great basin watershed: Methods and considerations for quantifying phenophases in a cold desert. Sensors 16, 1948. doi.org/https://doi.org/10.3390/s16111948

Stockdale, C.A., Bozzini, C., Macdonald, S.E., Higgs, E., 2015. Extracting ecological information from oblique angle terrestrial landscape photographs: Performance evaluation of the WSL Monoplotting Tool. Appl. Geogr. 63, 315–325. https://doi.org/10.1016/j.apgeog.2015.07.012.

Tang, K., Paluri, M., Fei-Fei, L., Fergus, R., Bourdev, L., 2015. Improving image classification with location context, in: 2015 IEEE International Conference on Computer Vision (ICCV). Presented at the 2015 IEEE International Conference on Computer Vision (ICCV), IEEE, Santiago, Chile, pp. 1008–1016. doi.org/https://doi.org/10.1109/iccv.2015.121.

Tewkesbury, A.P., Comber, A.J., Tate, N.J., Lamb, A., Fisher, P.F., 2015. A critical synthesis of remotely sensed optical image change detection techniques. Remote Sens. Environ. 160, 1–14.

Tompson, J., Goroshin, R., Jain, A., LeCun, Y., Bregler, C., 2014. Efficient Object Localization Using Convolutional Networks. arXiv:1411.4280 [cs].

Traore, B.B., Kamsu-Foguem, B., Tangara, F., 2018. Deep convolution neural network for image recognition. Ecol. Informa. 48, 257–268.

Van Bogaert, R., Haneca, K., Hoogesteger, J., Jonasson, C., De Dapper, M., Callaghan, T.V., 2011. A century of tree line changes in sub-Arctic Sweden shows local and regional variability and only a minor influence of 20th century climate warming: Twentieth century tree line changes in the Swedish sub-Arctic. J. Biogeogr. 38, 907–921. https://doi.org/10.1111/j.1365-2699.2010.02453.x.

Vihinen, M., 2012. How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis. BMC Genomics 13, S2. https://doi.org/10.1186/1471-2164-13-S4-S2.

Voulodimos, A., Doulamis, N., Doulamis, A., Protopapadakis, E., 2018. Deep learning for computer vision: a brief review. Comput. Intell. Neurosci. 1–13. https://doi.org/10.1155/2018/7068349.

Wäldchen, J., Rzanny, M., Seeland, M., Mäder, P., 2018. Automated plant species identification—Trends and future directions. PLoS Comput. Biol. 14, e1005993. https://doi.org/10.1371/journal.pcbi.1005993.

Webb, R.H., Boyer, D.E., Turner, R.M. (Eds.), 2010. Repeat Photography: Methods and Applications in the Natural Sciences. Island Press, Washington.

Weiss, K., Khoshgoftaar, T.M., Wang, D., 2016. A survey of transfer learning. J. Big Data 3. https://doi.org/10.1186/s40537-016-0043-6.

Wiesmann, S., Steiner, L., Pozzi, M., Bozzini, C., Bauder, A., Hurni, L., 2012.

Reconstructing historic glacier states based on terrestrial oblique photographs. In: Proceedings of the AutoCarto International Symposium on Automated Cartography, Columbus, OH, USA.
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. https://doi.org/10.1016/j.envsoft.2017.02.004.
Zhang, L., Verma, B., Stockwell, D., 2016. Spatial contextual superpixel model for natural roadside vegetation classification. Pattern Recogn. 60, 444–457. doi.org/https://doi.org/10.1016/j.patcog.2016.05.013
Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., Atkinson, P.M., 2018. An object-based convolutional neural network (OCNN) for urban land use classification. Remote Sens. Environ. 216, 57–70. https://doi.org/10.1016/j.rse.2018.06.034.