Environmental Modelling & Software 91 (2017) 127–134

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Automatic land cover classification of geo-tagged field photos by deep learning

Guang Xu [a, *], Xuan Zhu [a], Dongjie Fu [b], Jinwei Dong [c, d], Xiangming Xiao [d]

[a] School of Earth, Atmosphere and Environment, Monash University, Clayton Campus, Clayton, VIC 3800, Australia
[b] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, China
[c] Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
[d] Department of Microbiology and Plant Biology, and Center for Spatial Analysis, University of Oklahoma, Norman, OK 73019, USA

ARTICLE INFO

Article history:
Received 4 July 2016
Received in revised form
19 January 2017
Accepted 2 February 2017

Keywords:
Deep learning
Convolutional neural network
Transfer learning
Multinomial logistic regression
Land cover
Crowdsourced photos

ABSTRACT

With more and more crowdsourcing geo-tagged field photos available online, they are becoming a potentially valuable source of information for environmental studies. However, the labelling and recognition of these photos are time-consuming. To utilise such information, a land cover type recognition model for field photos was proposed based on the deep learning technique. This model combines a pre-trained convolutional neural network (CNN) as the image feature extractor and the multinomial logistic regression model as the feature classifier. The pre-trained CNN model Inception-v3 was used in this study. The labelled field photos from the Global Geo-Referenced Field Photo Library (http://eomf.ou.edu/photos) were chosen for model training and validation. The results indicated that our recognition model achieved an acceptable accuracy (48.40% for top-1 prediction and 76.24% for top-3 prediction) of land cover classification. With accurate self-assessment of confidence, the model can be applied to classify numerous online geo-tagged field photos for environmental information extraction.

1. Introduction

Global land cover mapping is a fundamental method to monitor and evaluate global changes for environmental research and policy making. Remote sensing based classification is considered as the most efficient way for land cover mapping, but it always requires ground referencing data for training (calibration) and validation. Field survey is the general approach to acquiring the ground referencing data. During field surveys, photographs are often used to record detailed information of particular types of land cover at specific locations. Information provided by these photos can be used to help classify and validate land cover maps derived from analyses of aerial or satellite imagery. Lots of efforts have been made to archive these field photos. For example, from 1999 to 2011, the United States Geological Survey has conducted a project named "Land Cover Trends" (Gallant et al., 2004). During the project, 13,000 field photos were collected with ecoregion labels, as a nation-wide, geo-referenced dataset for land cover change mapping and as training or test site data for remote sensing image classification (Soulard and Sleeter, 2012).

However, field photo collecting by experts at a large scale is always labour-intensive and time-consuming. The crowd-sourced field photos have become a useful source employed by researchers. Since 2011, the University of Oklahoma has set up a Global Geo-Referenced Field Photo Library (Xiao et al., 2011), and also released the mobile app "Field Photo" (freely available in Google Play store and Apple Store for public use) to collect geo-referenced field photos from other researchers; and the library now contains more than 150,000 field photos (in public mode) with manually labelled land cover types. Furthermore, in 2013, the Geo-Wiki (Fritz et al., 2012) project released its mobile app "Geo-Wiki Pictures" which enables the public to share landscape photographs with detailed land cover types and other environmental information. This platform has accumulated more than 17,800 pictures so far.

These pictures can be used for validation of land cover maps at local to global scales (Fritz et al., 2012; Dong et al., 2013). However, with the aid of non-professional volunteers, crowdsourced field

* Corresponding author.
E-mail address: xg1990@gmail.com (G. Xu).

http://dx.doi.org/10.1016/j.envsoft.2017.02.004
1364-8152/

photos are sometimes misclassified with low accuracy. Average producer's accuracy of volunteers ranges from 52% to 62% (Foody et al., 2013). The experiment by Sparks et al., (2015) showed that the overall accuracy of volunteer-based Earth observation is around 70%, which is comparable to the result of GEO-Wiki. The accuracy of crowdsourced photo interpretation is becoming a bottleneck for its development.

Additionally, the volume of unlabelled online photos has been increasing at a rapid speed. In 2015, Yahoo released the YFCC (Yahoo Flickr Creative Commons) dataset (Thomee et al., 2015) containing 100 million online photos. Panoramio (http://www.panoramio.com/) from Google has also collected countless photos of the world, which remain to be utilised. However, the processing speed of public participated photo recognition is limited by the number of volunteers. Alternative efficient techniques should be developed.

With the fast development of deep learning technology, it becomes much more likely to make artificial intelligence to aid field work and help with the interpretation of ground referencing data for land cover types. In the image recognition field of deep learning, convolutional neural network (CNN) (Fukushima, 1980) is becoming the most promising algorithm, which incorporates convolutional and max-pooling layers into traditional neural networks for image feature extraction. It has already demonstrated satisfactory results for digit number recognition (LeCun et al., 1998), face detection (Garcia and Delakis, 2002; Osadchy et al., 2007; Strigl et al., 2010), pedestrian detection (Sermanet et al., 2013) and object detection (Long et al., 2015). However, these technologies have not been used specifically for the identification of land cover types. Therefore, the exploration of the state-of-the-art deep learning technology on photo recognition for land cover classification is needed to promote the automatic generation of the training and validation samples for large scale land cover mapping.

In this study, the classification model is built and tested for land cover type recognition by using the field photos from the global geo-referenced field photo library (http://eomf.ou.edu/photos), based on the CNN. Manually tagged pictures of land cover are used for model training and validation. The model performance and credibility are also assessed.

2. Methodology

2.1. Transfer learning

Training a complex neural network from scratch is always very slow on a large training set. Thus, transfer learning was proposed to apply a pre-trained neural network to another related problem (Caruana, 1995; Bengio et al., 2011; Bengio, 2012; Donahue et al., 2013). The idea of transfer learning is based on the fact that the knowledge learned from one task could be applied to solve other similar problems (Pan and Yang, 2010). In this way, the researchers could save much time for model training. A pre-trained neural network will include both its model structure and the network weights trained with large datasets. The pre-trained CNN models can always capture important features from common photos. Thus, they can be widely used for different applications.

There are mainly two kinds of strategies to take advantage of pre-trained models: feature extraction and fine-tuning. Fine-tuning means continuing training the pre-trained CNN model with another new dataset, according to the task of interest. This process will adjust the network weights of the pre-trained model to fit its outputs as close to new labels as possible, which has been proven to be effective by Yosinski et al., (2014). The benefit of fine-tuning is less time consuming because the training starts from pre-trained models. This technique has already been used for image style recognition (Karayev et al., 2013).

Unlike the fine-tune, feature extraction works by removing the last layer of a CNN model (output layer) and treating the output data of the second last layer as extracted features (also called CNN codes), which are always high dimensional vectors and implicitly represent characteristics of input images. The extracted features then can be analysed by other classifying models, such as logistic regression, multinomial logistic regression or support vector machine. In feature extraction, the pre-trained CNN model acts as the image feature extractor in the whole workflow. This framework has also shown competitive performance compared with other sophisticated models (Razavian et al., 2014).

Feature extraction is suitable when the research dataset is not similar to the original training dataset of the pre-trained CNN model in terms of sample size or sample content when it may take too much time to fine-tuning a CNN model. In this study, the Inception-v3 model was pre-trained by the ImageNet dataset (Russakovsky et al., 2015), which contains more than 10,000,000 labelled images depicting over 10,000 object categories. However, the Global Geo-referenced Field Photo Library has only nearly 30,000 training samples for landscape classification, which may cause overfitting if the CNN model is fine-tuned until the first layer. Thus feature extraction was chosen in this research.

The overall model framework for our study is shown in Fig. 1. The pre-trained CNN model Inception-v3 from Google (Szegedy et al., 2015) was chosen, because of its excellent performance on image recognition. The Inception-v3 model reached a top-5 error rate of 3.46%, which is even better than the error rate 5.1% of human (Karpathy, 2016) at the same image recognition challenge. By removing the last output layer of the pre-trained Inception-v3 model, the image feature extractor was then acquired with the output of 2048 CNN codes (image features). The CNN codes were then classified by a weighted multinomial logistic regression model for land cover type recognition.

2.2. Weighted multinomial logistic regression

In neural networks, multinomial logistic regression (Arbib, 2003) is the most widely used classification model as the last layer of a network, because it is straightforward and efficient. Multinomial logistic regression, also called softmax regression, is the generalised form of logistic regression, which can be used to model and predict probabilities that samples belong to more than two independent types. Its mathematic form is:

$$h_\theta(x) = \begin{bmatrix} P(y=1|x,\theta) \\ \cdots \\ P(y=k|x,\theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\left(\theta^T x\right)}} \begin{bmatrix} e^{\left(\theta^{(1)T} x\right)} \\ \vdots \\ e^{\left(\theta^{(k)T} x\right)} \end{bmatrix}, \quad (1)$$

where x represents an input variable, or a single sample, which is a $m \times 1$ dimensional vector, where m is the number of features of the input variable; k is the number of categories, into which the input variable to be classified; $h_\theta(x)$ represents the predicted probabilities that x belongs to each of k classes; θ is the multinomial logistic regression model parameter, an $m \times k$ matrix. In this research, the input variable x for each sample contains the CNN codes extracted by Inception-v3 model, which is a 2048 dimensional vector.

The process of model training is to find out the best model parameter θ that minimises the difference between the predicted and the actual probability that samples belong to each category. The gradient descent method is used to search for optimum parameter θ, which is an iterative algorithm that updates parameter θ step by step:

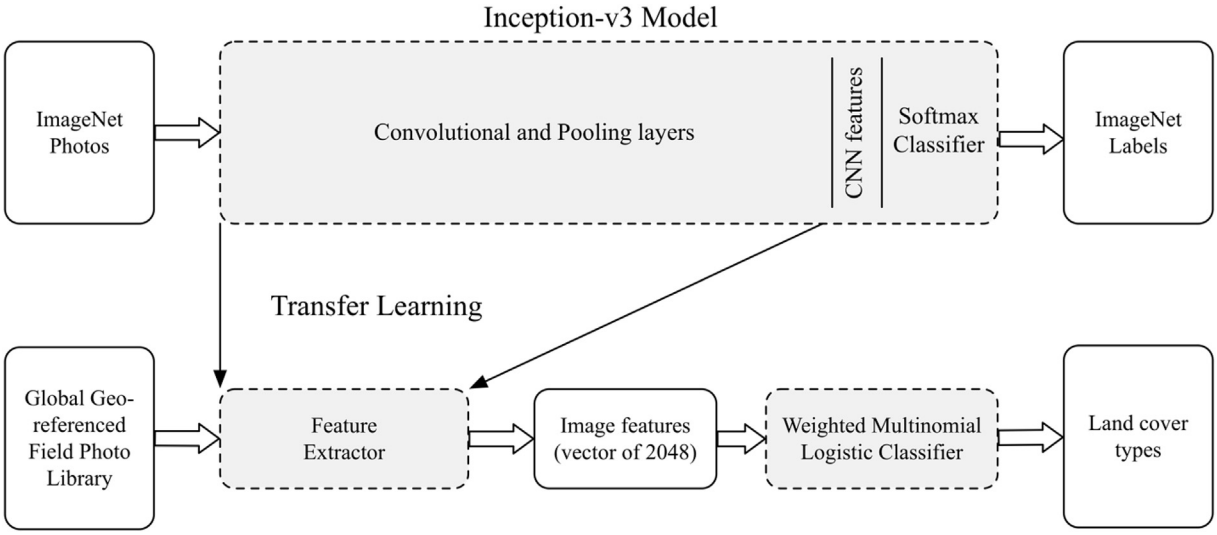

Fig. 1. The overall transfer learning framework of this study. All the convolutional and pooling layers except the last multinomial logistic classification layer of the Inception-v3 model were taken out as the feature extractor of this study. The extracted image features were then classified into different land cover types by weighted multinomial logistic model.

$$\theta := \theta - \gamma \left[X^T \left(h_\theta(X) - Y \right) + \lambda\theta \right], \tag{2}$$

where X is an $N \times m$ matrix and represents the CNN features of all the training samples, and N is the number of training samples; Y is an $N \times k$ matrix and represents the actual probabilities for each training sample belonging to each type; $h_\theta(X)$ is also an $N \times k$ matrix and represents the predicted probabilities for each sample; γ is the iteration step size. Because there are infinitely many solutions for any given training samples, $\lambda\theta$ is used in our iteration equation as a weight decay term to find out the unique solution. The L-BFGS (Limited-memory Broyden–Fletcher–Goldfarb–Shanno) algorithm was used to find optimal solutions. The learning rate is not fixed and varies as the training progresses and is decided by the solving algorithm.

Owing to the imbalanced numbers of training samples within different categories, the sample weighting should be taken into account when the multinomial logistic regression model is being trained. The land cover types with a small sample size should be treated with higher weights so that they will not be overwhelmed by those with a large sample size. Therefore, the sample weight w will be applied to the iteration equation of the ordinary multinomial logistic regression model as following:

$$\theta := \theta - \gamma \left[X^T \, diag(w) (h_\theta(x) - Y) + \lambda\theta \right], \tag{3}$$

where w is a vector with N elements and represents the training weighting of each sample; and $diag(w)$ is an operator that generates a diagonal matrix from a vector. The weight of each type of land cover is defined according to its sample size. All the samples within the same category are assigned with the same weighting. The weight for land cover type i is represented by w_i and calculated by the equation:

$$w_i = \frac{N}{n_i * k}$$

where N is the total number of training samples, n_i is the training samples within the i -th category, and k is the total number of types. This equation assures that the samples from small groups (with small n_i) will have large training weights, and yield the same effect on the training process.

3. Experiment design

3.1. Data sources

Field photos from Global Geo-Referenced Field Photo Library (http://www.eomf.ou.edu/photos/) are used for model training and validation. There are more than 150,000 field pictures available in this library, among which 35,887 pictures are labelled with land cover types. These labelled photos are selected for this research. This dataset has been accepted to have high quality and widely used for land cover mapping and validation (Dong et al., 2013; Leinenkugel et al., 2013; Dong et al., 2014; Tsarouchi et al., 2014; Qin et al., 2015; B. Chen et al., 2016; Y. Chen et al., 2016; Dong et al., 2016).

This dataset comprises 19 types of land cover as shown in Table 1. Due to the uneven distribution of samples within different land cover types, the land cover types were reclassified as nine new land cover types. Specifically, the new Forest type includes Deciduous Broadleaf Forest, Deciduous Needleleaf Forest, Evergreen Broadleaf Forest, Evergreen Needleleaf Forest and Mixed Forest; the new Shrublands type includes Open Shrublands and Closed Shrublands; the new Croplands type includes Croplands and Cropland/Natural Vegetation Mosaic; the new Savannas type includes Savannas and Woody Savannas. The Orchards type was discarded due to too few samples. (Table 1).

Spatial and temporal coverages of these photos are shown in Fig. 2 and Fig. 3. Most of the photos were taken after the year 2008. They were mainly taken in Northern America, India, part of Africa, East Asia and Australia, which are well distributed globally and include various landscapes all over the world.

3.2. Model training and validation

The repeated random sub-sampling validation (Dubitzky et al., 2007) has been performed in this study. Each time, the labelled samples are divided into two subsets: training and testing set. The model accuracy is evaluated by testing samples, which are

Table 1
Land cover types, sample sizes and training weights used in modelling.

Original Type Name	Sample Count	New Type Name	Sample Count	Training Samples	validation Samples	Training weight
Deciduous Broadleaf Forest	1670	Forest	3975	3875	100	0.80
Deciduous Needleleaf Forest	38					
Evergreen Broadleaf Forest	349					
Evergreen Needleleaf Forest	1595					
Mixed Forest	323					
Open Shrublands	475	Shrublands	1007	907	100	3.42
Closed Shrublands	532					
Grasslands	1627	Grasslands	1627	1527	100	2.03
Barren Or Sparsely Vegetated	1085	Barren	1085	985	100	3.15
Croplands	5794	Croplands	5984	5884	100	0.53
Cropland/Natural Vegetation Mosaic	190					
Plantations	341	Plantations	341	241	100	12.88
Permanent Snow And Ice	673	Snow and Ice	673	573	100	5.42
Permanent Wetlands	13,369	Wetlands	13,369	13,269	100	0.23
Savannas	1361	Savannas	1627	1527	100	2.03
Woody Savannas	266					
Urban And Built-Up	2919	Urban	2919	2819	100	1.10
Water Bodies	2635	Water	2635	2535	100	1.22
Orchards	29	–	–	–	–	–
Total Count	35,271		35,242	34,142	1100	

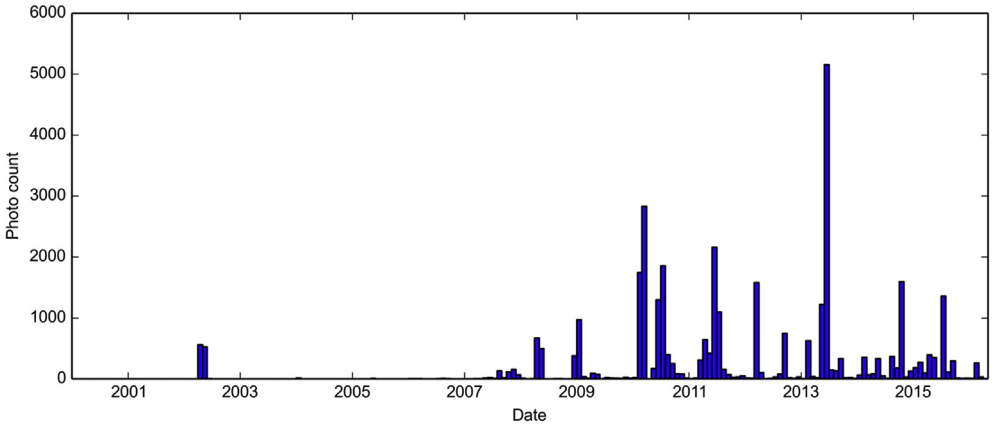

Fig. 2. Temporal coverage of photo samples from Global Geo-Referenced Field Photo Library.

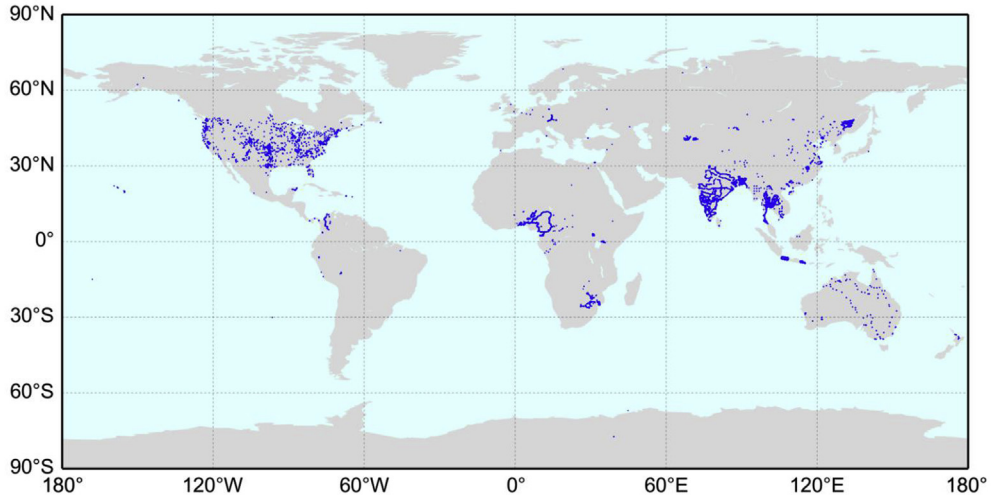

Fig. 3. Spatial coverage of photo samples (blue points) from Global Geo-Referenced Field Photo Library. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

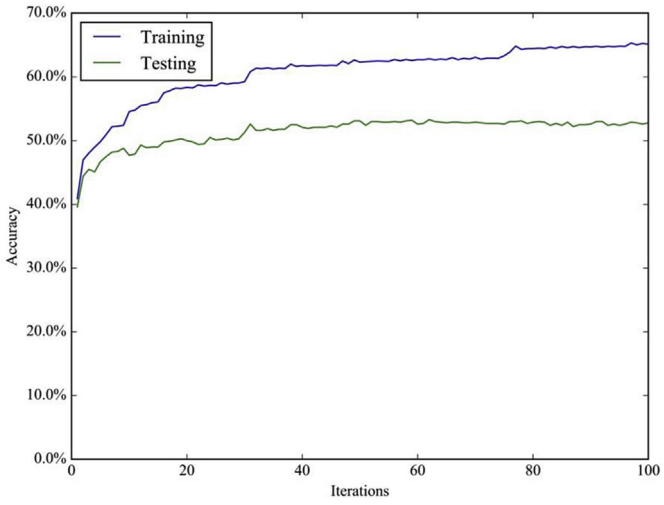

Fig. 4. Model fitting history of training and testing accuracies.

independent of training samples. However, due to imbalanced samples sizes within different land cover types, 100 testing samples are drawn from each land cover types in order to ensure the validation result is unbiased, as shown in Table 1. The training and testing process are repeated by 100 times, to avoid the effects of random factors and provide a comprehensive model evaluation. Their average accuracies are used for the following discussion in our paper. Different sample weights were applied to each training sample according to the categories they belong to, which are also listed in Table 1.

Fig. 4 shows how the model training and testing accuracies change with fitting iterations. Both training and testing accuracies monotonically increased with fitting iterations. No sign of overfitting could be identified.

In this research, top-1 accuracy and top-3 accuracy were used to adequately reveal the model performance. Top-1 accuracy is the percentage of testing samples whose most possible land cover types match their actual types. Top-3 accuracy is defined as the percentage of testing samples whose actual types are among the most three possible land cover types predicted by the multinomial logistic regression model.

In addition, to avoid the effect of possible random factors, the training and testing processes were repeated ten times, with different training and validation samples each time. The overall results of the model validation were then analysed.

4. Experiment result

4.1. Overall accuracy

Examples of model prediction are shown in Fig. 5, in which the probabilities that a photo belongs to every land cover types are given by the multinomial logistic regression. The type with the highest probability will be taken as the predicted type within the photo. And the corresponding probability is referred as “predicted probability”. This probability represents how confident the model is about its prediction. So it can also be regarded as prediction confidence, which will be discussed in the next subsection.

Based on repeated cross-validation, top-1 and top-3 accuracies of the model were calculated from testing samples, which are listed in Table 2. To evaluate the model performance, random guessing accuracies are also calculated as benchmarks. Compared with the random guessing accuracies, this model produced a very promising prediction accuracy.

4.2. Predicted probability

As illustrated in Fig. 5, land cover type predictions are accompanied with probabilities, which can be used as indicators of the confidences of model predictions. It is possible that the predictions with high probabilities will probably be correct. In order to prove this assumption, further understanding how predicted probabilities are related to the model accuracy is important.

As shown by the histogram in Fig. 6, each bar represents the number of testing samples whose probabilities of top predictions are within the given interval. The red and blue bars represent incorrect and correct predictions respectively. And the correctness of any prediction depends on whether its top prediction is identical to the label of the sample. Very few samples have predicted

Open Water:	94.74%	Croplands:	95.39%	Urban:	95.49%
Urban:	1.51%	Barren:	2.10%	Croplands:	2.55%
Grasslands:	1.06%	Wetlands:	1.58%	Barren:	1.17%
Croplands:	0.73%	Grasslands:	0.38%	Shrublands:	0.44%
Forest:	0.67%	Savannas:	0.20%	Wetlands:	0.11%
Barren:	0.66%	Snow and Ice:	0.12%	Open Water:	0.11%
Shrublands:	0.17%	Urban:	0.10%	Snow and Ice:	0.08%
Snow and Ice:	0.16%	Shrublands:	0.09%	Forest:	0.02%
Wetlands:	0.16%	Open Water:	0.03%	Savannas:	0.01%
Savannas:	0.12%	Forest:	0.00%	Grasslands:	0.01%
Plantations:	0.02%	Plantations:	0.00%	Plantations:	0.01%

Fig. 5. Example outputs of the field photo classification model (The source of the photos is Global Geo-Referenced Field Photo Library. The usernames of the uploaders of these three photos are cocorahs, subbuteo and xiao 2007).

Table 2
Classification accuracies for model training, model validation, and random guessing.

Accuracy	Training	Validation	Random guessing
Top 1 accuracy	63.54%	48.40%	9.09%
Top 3 accuracy	87.32%	76.24%	27.27%

probabilities less than 20%. Obviously, the distributions of the predicted probabilities of most of the incorrect samples are bell-shaped and right skewed. Most of them tend to be around 40%. And very few are higher than 80%.

For the correct predictions, the distribution of predicted probabilities showed a monotone distribution. More samples have higher predicted probabilities. It can be inferred that if a prediction has a probability higher than 60%, it is more likely to be a correct prediction. In order to further illustrate the likelihood of correct predictions with their corresponding predicted probabilities, a new measurement named "posterior probability" will be defined.

In this study, the posterior probability is defined as the likelihood that a prediction is correct according to its predicted probability given by the multinomial regression model. This posterior probability can be derived by calculating the ratio of correct prediction frequency against total frequency at each predicted probability level in Fig. 6.

Fig. 7 shows the relationship between our model predicted probability and the posterior probability based on all the testing samples in this study. Each bar represents the percentage of correct predictions within given intervals. Obviously, they have a nearly linear relationship when predicted probability is greater than 20%.

This linear relationship implies that a prediction is more likely to be correct whenever the model gives a higher predicted probability. For example, a prediction is probably correct with a nearly 80% likelihood when the model gives a predicted probability of 90%. Such a linear relationship makes it possible to use predicted probability as an efficient indicator to distinguish potential correct and incorrect predictions. A certain threshold can be applied to ignore uncertain predictions and increase the reliability of the classification model.

4.3. User and producer accuracy

As mentioned above, model performance could be improved by applying a threshold of minimum predicted probability. To further elaborate how predicted probabilities can be used to find out better predictions, the user and producer accuracies with different probability thresholds were calculated and shown in Table 3. If the predicted probability of a sample is lower than a certain threshold (0%, 50%, and 75% are applied in Table 3), this sample will not be counted in the corresponding accuracy measurement. For most of the land cover types, both user accuracy and producer accuracy are improved when applying higher thresholds. Moreover, the overall accuracy reaches more than 73% when the threshold of 75% applied.

Thus, it is important to know how many samples are left after filtering out uncertain predictions. Fig. 8 shows the percentage of remaining samples when applying different minimum predicted probability thresholds. It represents the percentages of samples whose predicted probabilities are higher than any given filtering thresholds. It is obvious that 100% of the samples have more than 0% prediction confidence; and 0% of the samples have more than 100% prediction confidence. If the predictions with a probability of less than 60% are dropped out, there will be nearly 40% samples left. It means this model can not only produce a reliable prediction but also be aware of which samples are predicted reliably according to its predicted probabilities.

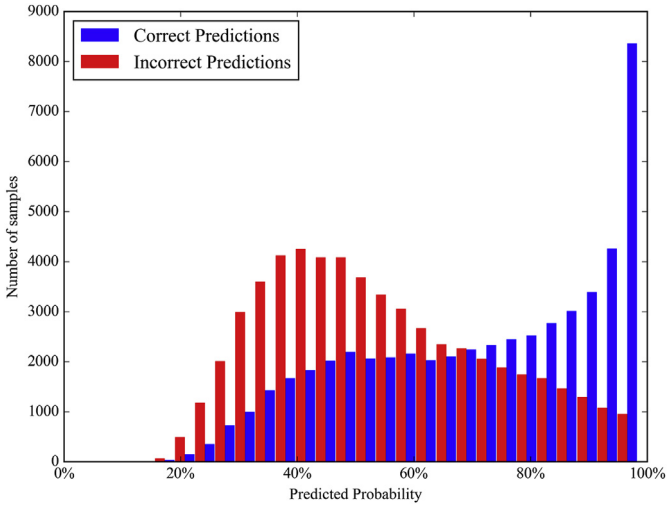

Fig. 6. Density distribution of right and wrong predicted probabilities of 1000 training samples with 100 repeats.

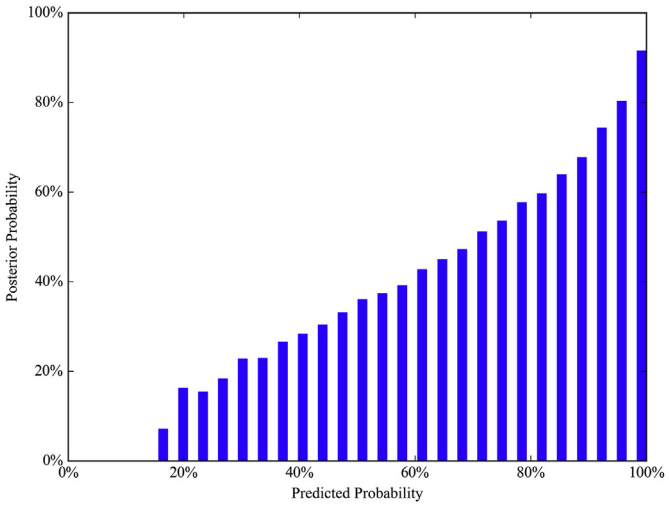

Fig. 7. Posterior probability of our model predictions.

5. Discussion

5.1. Model performance

In this study, the land cover classification model for field photos showed the top-1 accuracy of 48.40% and top-3 accuracy of 76.24%. Because no comparable researches were found for land cover classification of field photos, it is difficult to assess the model accuracy by comparison with the results of others. However, the model accuracy is much better than the accuracy of random guessing. If taking similar transfer learning researches into consideration, such as the image style recognition with the accuracy of 36.8% (Karayev et al., 2013) and image scene classification with the accuracy of 40.94% (Donahue et al., 2013), the model accuracy of 48.40% is relatively good for its research task.

What is more important, this model could provide the self-assessment of prediction confidence, which has been proved to be helpful to increase the prediction accuracy. After filtering out the predictions with predicted probabilities of less than 75%, the overall accuracy increased to 73.61%, which implies that the model is fully

Table 3
User and Producer Accuracies with probability filtering thresholds of 0%, 50%, and 75%.

Type	No filtering		Predicted probability higher than 50%		Predicted probability higher than 75%	
	User accuracy	Producer accuracy	User accuracy	Producer accuracy	User accuracy	Producer accuracy
Forest	48.97%	59.34%	58.56%	70.40%	71.23%	82.45%
Shrublands	39.78%	40.90%	47.79%	50.21%	57.02%	60.39%
Savannas	39.71%	39.96%	48.32%	48.66%	62.76%	61.23%
Cropland	51.44%	56.26%	62.64%	66.67%	74.97%	78.73%
Plantations	63.01%	39.26%	69.63%	51.59%	77.41%	69.85%
Grassland	48.71%	47.60%	60.37%	58.34%	75.82%	69.72%
Wetlands	58.23%	65.43%	70.95%	76.73%	85.44%	88.63%
Urban	34.62%	36.65%	43.63%	40.70%	55.24%	45.87%
Barren	41.81%	37.63%	51.29%	44.48%	62.28%	52.44%
Open Water	45.53%	48.63%	54.59%	58.07%	65.88%	69.71%
Snow and Ice	67.21%	61.12%	76.33%	75.05%	86.90%	88.72%
Overall Accuracy	48.40%		59.24%		73.61%	

aware of its prediction confidence and can precisely assess the degree of confidence by predicted probability. Therefore, this model can be applied to classify field photos and extract useful environmental information with high confidence.

5.2. Sample quality

The overall accuracy of this model is still far from perfect. One of the factors that affect the model performance is the quality of training samples. Uneven sample sizes for each land cover type limited the model's ability to understand the feature of different land cover types. For example, as shown in Table 1, there are only 27 samples for Deciduous Needleleaf Forest, which is why we merged all the forest types into a single type. Besides that, some definitions of land cover types could be ambiguous or overlapped with each other. For example, the Plantation type could be regarded as Forest or Croplands; the Wetlands could be labelled as the Water Bodies. What is more, as the training samples were classified manually, it is inevitable that some photos could be tagged with wrong types of land cover.

Moreover, not all the field photos have well-defined land cover types. There are always uncertainties when trying to classify photos. It is possible that some photographs contain more than one land cover type. But all the photo samples are supposed to represent only one type of land cover and are labelled with only one category in the database, which makes it difficult to validate the existence of multiple types within one photo. Besides, by using the multinomial logistic regression model, any samples are assumed to belong to only one category. And the classifier is supposed to be able to identify the most significant type within the image. Any ambiguous predictions are supposed to be the result of model bias.

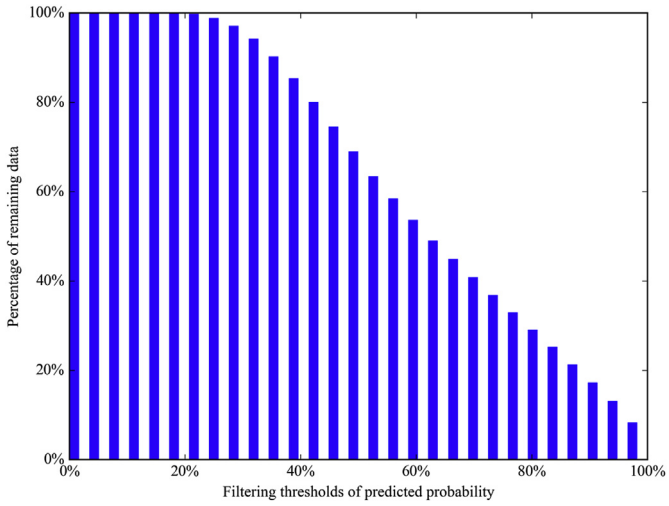

Fig. 8. Data percentile and corresponding minimum predicted probability.

It is also important to note that the photos in the field photo library were taken with various qualities, dependent upon the data providers, ranging from scientific researchers who follow precisely the field photo acquisition protocols, to amateur citizens who have only limited experience with the protocols. In this study, all training samples are assumed to be correctly labelled for model testing purposes. And a higher accuracy could be anticipated when better training samples were used for model training. Therefore, it is important to develop and deliver more training activities and materials that provide effective communications with citizen scientists, which will improve their skills in taking photos in the fields.

6. Conclusions

This research demonstrated that the field photo recognition model based on Inception-v3 CNN and weighted multinomial logistic regression produced the top-1 accuracy of 48.40% and the top-3 accuracy of 76.24% when applied with the Global Geo-Referenced Field Photo Library for land cover classification. The model is able to describe its prediction confidence and enables users to distinguish reliable and unreliable predictions.

The contribution of this research is not about a new image recognition algorithm. Both the Inception-v3 model and transfer learning are well-established techniques. This research is focused on proposing a new approach to applying the deep learning for geographic and environmental studies. The main contribution is to prove the possibility that artificial intelligence can help with land cover classification and to evaluate how well the model can perform. It provides a new research direction in citizen science. And hopefully, our research would be a benchmark for future studies.

Concerning future research, more photos should be taken for those land cover types with very few samples in the existing library in order to improve the training sample quality. A new land cover classification scheme will also need to be developed to avoid class overlapping. Besides the improvement of labelled data from data providers, the transfer learning framework could also be further improved. A CNN based classification model could be trained completely from scratch. But this may require extensive computing time and a large sample size. Besides, other more advanced pre-trained CNN models in the future could be used as the feature extractor. Such photo classification models will be applied to countless online geo-tagged field photos for land use/cover research and will form an important information source for environmental observation.

Acknowledgements

We would like to thank Google for making the TensorFlow library and the Inception-v3 model available. We also would like to thank all the contributors of the Global Geo-Referenced Field Photo Library who make this study possible. This study was supported in part by the NASA Land Use and Land Cover Change program (NNX14AD78G), and the Key Research Program of Frontier Sciences, the Chinese Academy of Sciences (QYZDB-SSW-DQC005), the "Thousand Youth Talents Plan". This research was also supported by the Youth Science Funds of State Key Laboratory of Resources and Environmental Information System (O8R8A080YA), Chinese Academy of Sciences, the research grants (41501473) funded by National Science Foundation of China and the research grants (Y6V60206YZ) funded by Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences.

References

Arbib, M.A., 2003. The Handbook of Brain Theory and Neural Networks. MIT press.

Bengio, Y., 2012. Deep learning of representations for unsupervised and transfer learning. Unsupervised Transf. Learn. Challenges Mach. Learn. 7, 19.

Bengio, Y., Bastien, F., Bergeron, A., Boulanger-Lewandowski, N., Breuel, T.M., Chherawala, Y., Cisse, M., Côté, M., Erhan, D., Eustache, J., 2011. Deep Learners Benefit More from Out-of-distribution Examples (International Conference on Artificial Intelligence and Statistics).

Caruana, R., 1995. Learning many related tasks at the same time with backpropagation. Adv. neural Inf. Process. Syst. 657–664.

Chen, B., Li, X., Xiao, X., Zhao, B., Dong, J., Kou, W., Qin, Y., Yang, C., Wu, Z., Sun, R., 2016. Mapping tropical forests and deciduous rubber plantations in Hainan Island, China by integrating PALSAR 25-m and multi-temporal Landsat images. Int. J. Appl. Earth Observation Geoinformation 50, 117–130. http://www.sciencedirect.com/science/article/pii/S0303243416300423.

Chen, Y., Dong, J., Xiao, X., Zhang, M., Tian, B., Zhou, Y., Li, B., Ma, Z., 2016. Land Claim and Loss of Tidal Flats in the Yangtze Estuary. Scientific Reports 6. http://www.nature.com/articles/srep24018.

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T., 2013. Decaf: a Deep Convolutional Activation Feature for Generic Visual Recognition. arXiv preprint arXiv:1310.1531.

Dong, J., Xiao, X., Chen, B., Torbick, N., Jin, C., Zhang, G., Biradar, C., 2013. Mapping deciduous rubber plantations through integration of PALSAR and multi-temporal Landsat imagery. Remote Sens. Environ. 134, 392–402.

Dong, J., Xiao, X., Menarguez, M.A., Zhang, G., Qin, Y., Thau, D., Biradar, C., Moore, B., 2016. Mapping paddy rice planting area in northeastern Asia with landsat 8 images, phenology-based algorithm and google earth engine. Remote Sens. Environ. 185, 142–154. http://www.sciencedirect.com/science/article/pii/S003442571630044X.

Dong, J., Xiao, X., Sheldon, S., Biradar, C., Zhang, G., Duong, N.D., Hazarika, M., Wikantika, K., Takeuhci, W., Moore III, B., 2014. A 50-m forest cover map in Southeast Asia from ALOS/PALSAR and its application on forest fragmentation assessment. PloS one 9 (1).

Dubitzky, W., Granzow, M., Berrar, D.P., 2007. Fundamentals of Data Mining in Genomics and Proteomics. Springer Science & Business Media.

Foody, G.M., See, L., Fritz, S., Van der Velde, M., Perger, C., Schill, C., Boyd, D.S., 2013. Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project. Trans. GIS 17 (6), 847–860.

Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., Van der Velde, M., Kraxner, F., Obersteiner, M., 2012. Geo-Wiki: an online platform for improving global land cover. Environ. Model. Softw. 31, 110–123.

Fukushima, K., 1980. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybern. 36 (4), 193–202.

Gallant, A.L., Loveland, T.R., Sohl, T.L., Napton, D.E., 2004. Using an ecoregion framework to analyze land-cover and land-use dynamics. Environ. Manag. 34 (1), S89–S110.

Garcia, C., Delakis, M., 2002. A Neural Architecture for Fast and Robust Face Detection. Pattern Recognition, 2002. Proceedings. 16th International Conference on, 2, pp. 44–47.

Karayev, S., Trentacoste, M., Han, H., Agarwala, A., Darrell, T., Hertzmann, A., Winnemoeller, H., 2013. Recognizing Image Style. arXiv preprint arXiv:1311.3715.

Karpathy, A., 2016. What I Learned from Competing against a ConvNet on ImageNet from. http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324.

Leinenkugel, P., Kuenzer, C., Oppelt, N., Dech, S., 2013. Characterisation of land surface phenology and land cover based on moderate resolution satellite data in cloud prone areas—a novel product for the Mekong basin. Remote Sens. Environ. 136, 180–198.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation. CVPR (to appear).

Osadchy, M., Cun, Y.L., Miller, M.L., 2007. Synergistic face detection and pose estimation with energy-based models. J. Mach. Learn. Res. 8, 1197–1215.

Pan, S.J., Yang, Q., 2010. A survey on transfer learning. Knowl. Data Eng. IEEE Trans. 22 (10), 1345–1359.

Qin, Y., Xiao, X., Dong, J., Zhou, Y., Zhu, Z., Zhang, G., Du, G., Jin, C., Kou, W., Wang, J., 2015. Mapping paddy rice planting area in cold temperate climate region through analysis of time series Landsat 8 (OLI), Landsat 7 (ETM+) and MODIS imagery. ISPRS J. Photogrammetry Remote Sens. 105, 220–233.

Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S., 2014. CNN Features Off-the-shelf: an Astounding Baseline for Recognition (Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops).

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

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y., 2013. Overfeat: Integrated Recognition, Localization and Detection Using Convolutional Networks. arXiv preprint arXiv:1312.6229.

Soulard, C.E., Sleeter, B.M., 2012. Late twentieth century land-cover change in the basin and range ecoregions of the United States. Reg. Environ. Change 12 (4), 813–823.

Sparks, K., Klippel, A., Wallgrün, J.O., Mark, D., 2015. Citizen Science Land Cover Classification Based on Ground and Aerial Imagery. Spatial Information Theory. Springer, pp. 289–305.

Strigl, D., Kofler, K., Podlipnig, S., 2010. Performance and Scalability of GPU-based Convolutional Neural Networks. Parallel, Distributed and Network-Based Processing (PDP), 2010 18th Euromicro International Conference on, pp. 317–324.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2015. Rethinking the Inception Architecture for Computer Vision. arXiv:1512.00567.

Thomee, B., Shamma, D.A., Friedland, G., Elizalde, B., Ni, K., Poland, D., Borth, D., Li, L.-J., 2015. The New Data and New Challenges in Multimedia Research. arXiv preprint arXiv:1503.01817.

Tsarouchi, G., Mijic, A., Moulds, S., Buytaert, W., 2014. Historical and future land-cover changes in the upper Ganges basin of India. Int. J. Remote Sens. 35 (9), 3150–3176.

Xiao, X., Dorovskoy, P., Biradar, C., Bridge, E., 2011. A library of georeferenced photos from the field. EOS, Trans. Am. Geophys. Union 92 (49), 453–454.

Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? Adv. Neural Inf. Process. Syst. 3320–3328.