

Article

The Verification of Land Cover Datasets with the Geo-Tagged Natural Scene Images

Liu Cui ¹, Hui Yang ^{1,2,3,*} , Liang Chu ⁴, Qingping He ¹, Fei Xu ⁵, Yina Qiao ¹, Zhaojin Yan ¹, Ran Wang ¹ and Hui Ci ¹

¹ School of Resources and Geosciences, China University of Mining and Technology, Xuzhou 221116, China
² Key Laboratory of Coalbed Methane Resources and Reservoir Formation Process of the Ministry of Education, China University of Mining and Technology, Xuzhou 221116, China
³ Artificial Intelligence Research Institute, China University of Mining and Technology, Xuzhou 221116, China
⁴ Xuzhou Municipal Bureau of Natural Resources and Planning, Xuzhou 221116, China
⁵ Xuzhou Registration Center of Natural Resources, Xuzhou 221116, China
* Correspondence: yanghui@cumt.edu.cn

Abstract: Land cover is important for global change studies, and its accuracy and reliability are usually verified by field sampling, which costs a lot. A method was proposed for the verification of land cover datasets with the geo-tagged natural scene images using a convolutional neural network. The nature scene images were firstly collected from the Land Use and Cover Area frame Survey (LUCAS) and global crowdsourcing images platform Flickr, then classified according to the Land Cover Classification System. The Nature Scene Image Classification (NSIC) model based on the GoogLeNet Inception network for recognition of natural scene images was then constructed. Finally, in the UK, as a verification area, the European Space Agency Climate Change Initiative Land Cover (ESA CCI-LC) datasets and the Global land-cover product with fine classification system (GLC-FCS) were verified using the NSIC-Inception model with the nature scene image set. The verification results showed that the overall accuracy verified by LUCAS was very close to the accuracy of the land cover product, which was 94.41% of CCI LC and 92.89% of GLC-FCS, demonstrating the feasibility of using geo-tagged images classified by the NSIC model. In addition, the VGG16 and ResNet50 were compared with GoogLeNet Inception. The differences in verification between LUCAS and Flickr images were discussed regarding the image's quantity, the spatial distribution, the representativeness, and so on. The uncertainties of verification arising from differences in the spatial resolution of the different datasets were explored by CCI LC and GCL-FCS. The application of the method has great potential to support and improve the efficiency of land cover verification.



Citation: Cui, L.; Yang, H.; Chu, L.; He, Q.; Xu, F.; Qiao, Y.; Yan, Z.; Wang, R.; Ci, H. The Verification of Land Cover Datasets with the Geo-Tagged Natural Scene Images. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 567. <https://doi.org/10.3390/ijgi11110567>

Academic Editor: Wolfgang Kainz

Received: 3 September 2022

Accepted: 4 November 2022

Published: 13 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: land cover verification; NSIC (Natural Scene Images Classification)-inception model; CCI LC (Climate Change Initiative Land Cover); geo-tagged natural scene images; GLC-FCS (global land-cover product with fine classification system)

1. Introduction

Land cover datasets on a global scale are essential for various geographic researches, including natural resource management [1], environmental monitoring [2], urban planning and sustainable development [3], Earth system modeling [4], etc. However, the accuracy of land cover and land use datasets differs widely due to the different methods used in their production. Therefore, the global and regional high-resolution classification accuracy evaluation of land cover products has become a hot research topic.

Initially, land cover data were validated through visual interpretation, a method that compared remote sensing images or field survey information with land cover maps to obtain verification accuracy [5]. In field survey data, more information could be conveyed in images than in text and labels, so geo-tagged field photographs have become highly utilized secondary verification data in land cover validation exercises.

Many institutions and organizations have invested in the collection of field photographs of regional land cover at a considerable human and economic cost to support the classification and verification of land cover maps, from two main sources. On the one hand, there was the collection of geographical field photographs organized by official survey agencies, such as the United States Geological Survey for the Land Cover Trends project [6], and the Land Use and Cover Area frame Survey conducted by European Space Agency in the European region [7,8]. On the other hand, there were crowdsourcing photos [8,9], including Flickr (<https://www.flickr.com/> (accessed on 28 May 2022)) and the Geo-Wiki project (<http://www.geo-wiki.org/> (accessed on 28 May 2022)), among others. These geo-tagged field images have been used to help classify and validate land cover maps derived from aerial or satellite image analysis [10,11].

Since crowdsourcing images offers easy collection, it is identified as having potential, especially for land cover mapping and validation. Many studies tried to improve the efficiency of image recognition and to accommodate the explosive growth of crowdsourcing image archives to be classified. Researchers have extracted image features using models, such as SVM models [12] and plain Bayes [13], to automatically identify image categories. In recent years, the convolutional neural network has become one of the most promising algorithms in the field of image recognition, with excellent performance in large-scale image processing tasks. Many convolutional neural network models with excellent performance emerged from the Large Scale Visual Recognition Challenge (ILSVRC), such as AlexNet [14], GoogleNet [15,16], VGG16 [17], ResNet [18], etc. Some research used a CNN model to extract features from images of ground natural scenes, resulting in improved accuracy and speed of image recognition [10,19].

Recently, some studies have been accumulated for the verification and production of land cover data using geo-tagged nature scene images [13,20,21]. Sitthi used a plain Bayesian classifier to classify a collection of Flickr images and produce a map of the land cover of Sapporo, Japan, with five categories: agricultural land, forest, water bodies, grassland, and urban areas [13]. Oba used CNN to extract high-level features from images to map the land use of the Stanford University campus, which was divided into eight categories, including research buildings, residences, hospitals, and parks [20]. Xing used VGG16 trained to learn Flickr image features to identify images in three categories, namely vegetation, artificial surfaces, and water, achieving 76.01% accuracy, and validated the Globeland30 2010 ground cover map [11].

The results obtained from these studies confirm that there is a correspondence between geo-tagged nature scene pictures as ground truth and land cover data. However, few studies used geo-tagged images as ground truth to verify the existing land cover products and explored the potential of crowdsourced images as reference data for land cover verification, especially the entire procedure of validation. At present, the possibility of images from different sources (such as Flickr and LUCAS) in land cover verification is not clear, and the suitable scale in the verification has not yet been explored.

Based on this, this study proposed the use of convolutional neural network to classify geo-tagged nature scene images for land cover data verification.

The specific objectives of this paper are:

- (1) Propose a land cover classification criterion based on the first stage of FAO/UNEP Land Cover Classification System (LCCS) which is flexible enough to accommodate different data sources, regions, and scales, that applies to both images and LULC maps, and correlate the natural scenes presented in the images with the land cover types.
- (2) Construct a nature scene images training set of land cover compatible with crowdsourced images and professional survey images and train an automatic recognition model of land cover images with strong generalization capability based on transfer learning to achieve the automated classification of natural scene images.

- (3) Compare the differences in the verification of land cover between Flickr and LUCAS crowdsourced images, and the differences in the verification using crowdsourced images between the CCI LC and GCL-FCS, in the UK as a verification area.

2. Materials and Methods

2.1. Overview

The verification of land cover datasets was conducted with the geo-tagged natural scene images using the model implemented in four main sections, as shown in Figure 1. Firstly, Flickr images and LUCAS images with EXIF information, such as geo-tagged and time-stamped, were collected to reflect the “ground truth” to construct natural scene image datasets. CCI LC and GCL-FCS land cover dataset were collected for the verification. Extent of the UK was collected for extract the land cover of verification area. Then, one part of natural scene images dataset distributed around the worldwide was manually annotated for training, validation, and to test the natural scene image classification model (NSIC). The NSIC model based on GoogLeNet Inception Network was constructed and evaluated for precision and recall. Thirdly, the other part of the images was filtered by time-stamped and geo-tagged for the verification of CCI-LC and GCL-FCS land cover data. The images in the UK were inputted in the NSIC model to obtain the category of images. The land cover type in the geo-tagged images was extracted from CCI LC and GCL-FCS. The confusion matrix was built according to the category of image and the type of land cover in the same geo-tagged to verify the land cover map. Finally, four metrics, including overall accuracy, product accuracy, user accuracy, and Kappa, were used to evaluate the land cover data. In order to further explore the universality and improve the reliability, the differences in spatial resolution land cover and different natural scene image source were further explored.

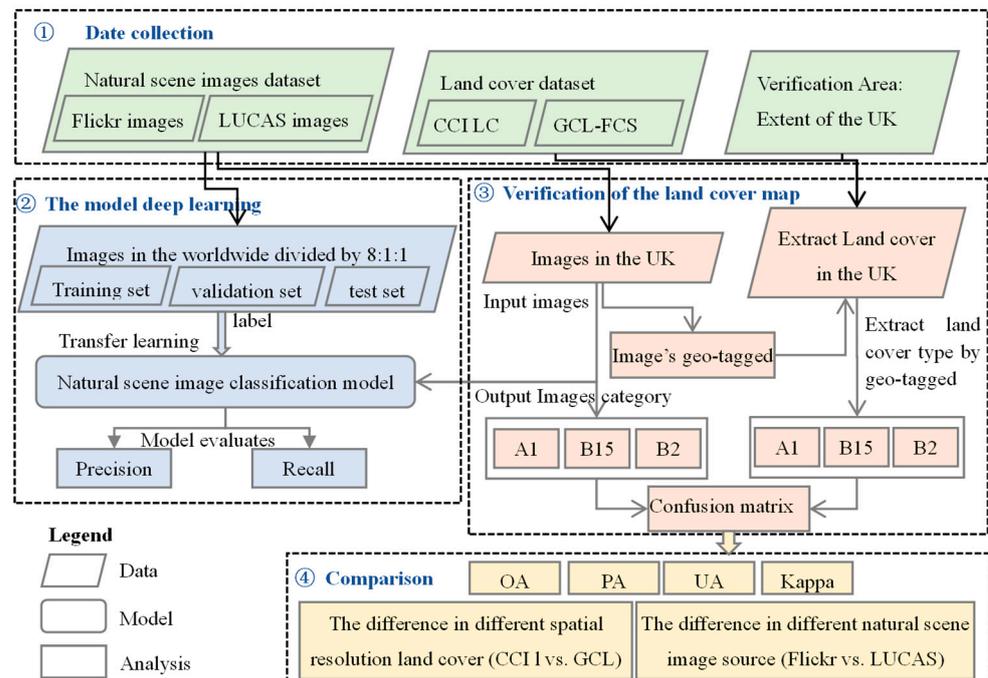


Figure 1. The workflow of the proposed method.

2.2. Datasets and Verification Area

Two datasets and one verification area were used to complete the development of the method and demonstrate its reliability.

The natural scene image datasets were distributed around the worldwide from 2012 to 2017, totaling 22,023 images, as shown in Figure 2a. The UK was selected as the verification area of the land cover according to the spatial distribution of the image collection, as shown

in Figure 2b. The land cover (LC) map of the European Space Agency Climate Change Initiative project (LC CCI) and the Global land-cover product with fine classification system (GLC-FCS) in the UK were verification with the geo-tagged natural scene images [22,23].

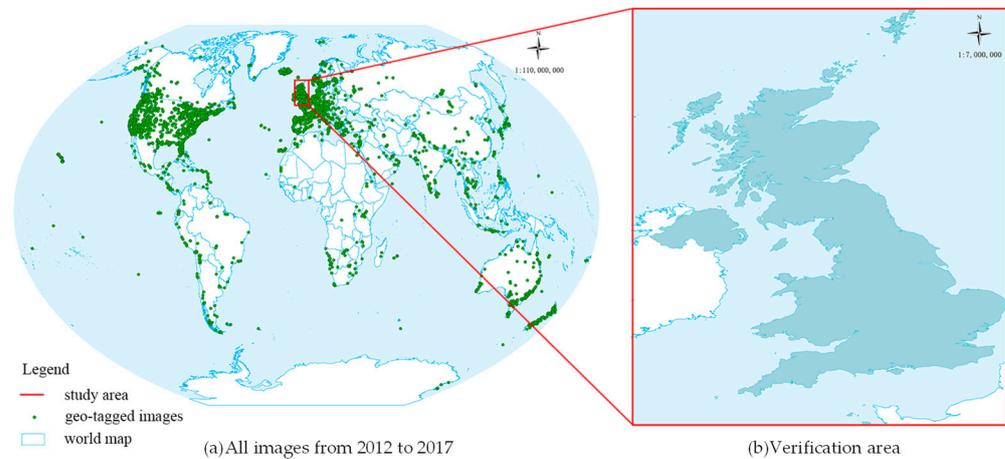


Figure 2. Spatial distribution of all geo-tagged natural scene images and verification area. (a) Spatial distribution of all geo-tagged natural scene images in dataset. (b) The verification area.

2.2.1. Natural Scene Image Datasets

The natural scene image datasets of this study have two sources, one is from Flickr (<https://www.flickr.com/> (accessed on 28 May 2019)) crowd-sourced images (Figure 3a) and the other is from the Land Use and Coverage Area frame Survey (LUCAS, <https://ec.europa.eu/eurostat/web/lucas/data> (accessed on 19 October 2021)) regularly carried out by EUROSTAT (Figure 3b). Both image sets contain a wealth of metadata information. Hence, parsing the image's EXIF information, information title, tags, the datetime of images acquired (time-stamped) and uploaded, GPS locations (geo-tagged), and so on could be obtained. However, the images from these two sources have different spatial distribution, camera angles, etc. Comparing the differences between these two sources in verifying the land cover can provide a greater evaluation of the images from the crowdsourcing in the land cover.



Figure 3. Part of natural scene image datasets. (a) Flickr images; (b) LUCAS images to the north, south, east, west and central.

The natural scene image dataset was cleaned by volunteers with geography background, and the images that reflect the “ground truth” were retained. There were 22,023 images in total.

Based on this, the image set was divided into two parts for the deep learning of the model and the verification of the land cover.

One part of the images was manually annotated except in the UK (land cover verification area) for training, validation, and testing of the natural scene image classification model. In all, 9144 images from all over the world were included, with 5933 Flickr and 3211 LUCAS compositions, ensuring that the model has a strong generalization ability to images from different sources. According to the same distribution of their categories, the images were divided into 8:1:1 as training, validation, and test datasets for the natural scene image classification model, respectively, as shown in Table 1.

Table 1. The quantity of images used for training, validation, and test.

	Training Dataset	Validation Dataset	Test Dataset	Total
A	4004	501	501	5006
B15	1644	205	205	2054
B16	515	65	65	645
B2	1151	144	144	1439
Total	7314	915	915	9144

The training set has 7314 samples, the validation set has 915 samples, and the test set has 915 samples. The quantity of A1, B15, B16, and B2 is 5006, 2054, 645, and 1439. There were the most A1 samples and the fewest B16 samples. The samples of each category are divided into training, validation, and test sets according to 8:1:1, as shown in Table 1.

The other part of the images was filtered according to time-stamped and geo-tagged, which were in 2015 and 2017 in the UK (Figure 4), totaling 12,879 images. The 12,879 images were annotated by the proposed model in this paper, then designated as the reference dataset of verification land cover data. Flickr and LUCAS images were compared to analyze the impact of different image sources on the verification of land cover in Section 3.3, and the spatial distribution of the two image sets is shown in Figure 4.

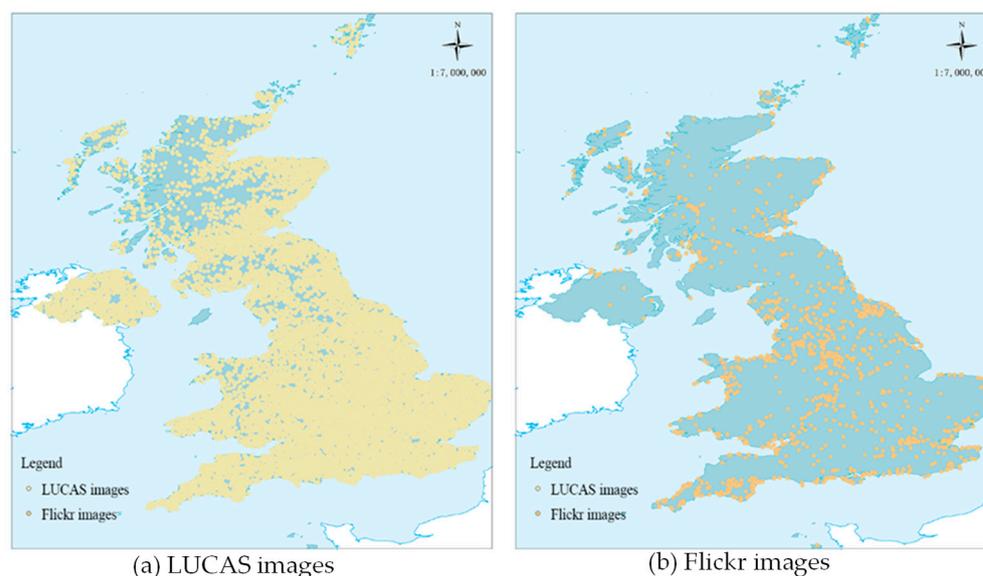


Figure 4. Spatial distribution of images from LUCAS and Flickr in UK (verification area). (a) LUCAS images, (b) Flickr images.

2.2.2. Land Cover Datasets

Two land cover datasets were used for verification with the geo-tagged natural scene image dataset. The product quality reports of land cover datasets were compared with the verification results using geo-tagged natural scene images in the UK to evaluate the reliability of this method.

The series data of Land Cover (LC) map data of the European Space Agency Climate Change Initiative project include consistent global land cover maps from 1992 to 2020 at the 300-m spatial resolution, which applies the FAO/UNEP Land Cover Classification System (LCCS) [24] to define land cover types. In terms of temporal availability, it was possible to coincide with the timing of the Flickr image set and the LUCAS image set used for land cover verification. Flickr images located in the UK for the full year of 2017 were used for the verification of the 2017 LC CCI data and the 2015 UK centrally oriented images were used for the verification of the 2015 LC CCI data. According to the product quality reports, the overall classification accuracy of the LC CCI for 2015 and 2017 was 93.39% and 95.27% [22].

Global land-cover product with fine classification system (GLC-FCS) at 30 m was produced by combining time series of Landsat imagery and high-quality training data from the GSPECLib (Global Spatial Temporal Spectra Library) on the Google Earth Engine computing platform, which provides more spatial details and a greater diversity of land-cover types than CCI_LC-2015. The verification results indicated the overall accuracy of 71.4% and kappa coefficient of 0.686 for the LCCS level-1 system (16 LCCS land-cover types) [23]. The spatial consistency of GLC_FCS30-2015 and CCI_LC-2015 was high because they shared the same classification system.

2.2.3. Verification Area

The UK was selected as the verification area for the land cover according to the spatial distribution of the image collection, as shown in Figure 2b.

There is low level of landscape diversity across many regions of the UK, especially in northern Scotland where mainly grassland or farmland is the main single land cover, while southern England is densely populated, with more artificial surface and a high level of landscape diversity [25]. There is a mass of geo-tagged natural scene images which contained mainly land cover types and reflect the “ground truth”, fulfilling the conditions for verification with geo-tagged natural scene images.

2.3. Photos Classification Using CNN Model

2.3.1. Category Definition of Land Cover and Natural Scene Images

The verification of land cover data using natural scene images needs to follow a uniform classification standard between land cover types and image classes. Yang reviewed and compared the major land cover classification methods at different scales and suggested that the Land Cover Classification System (LCCS) established by the Food and Agriculture Organization of the United Nations is the most promising land cover and land use classification standard with the potential to harmonize various classification methods [26]. FAO-LCCS is compatible with most land cover and land use classification methods and has a large impact on land cover classification worldwide. It is divided into two main stages. The first stage is a dichotomous stage, where eight main land cover types are defined (Table 2). The second stage is a modular classification stage, where further classification is performed by experimenting with pre-defined classification criteria (e.g., crop type, plant season, etc.) based on the first step.

Table 2. Basic Architecture of FAO/UNEP LCCS Phase I.

A. Vegetated	A1. Terrestrial	A11. Cultivated & Managed Areas
		A12. (SEMI) Natural Vegetation
	A2. Aquatic or regularly flooded	A23. Cultivated Aquatic Areas
		A24. (SEMI) Aquatic Vegetation
B. Non-Vegetated	B1. Terrestrial	B15. Artificial Surfaces
		B16. Bare Areas
	B2. Aquatic or regularly flooded	B27. Artificial Waterbodies, snow & ice
		B28. Natural Waterbodies, snow & ice

The legend of most of the land cover maps is based on the two-stage division of LCCS, and there are usually more than 20 types of land cover and land use. However, the Natural Scene Image Datasets cannot support the determination of vegetation attributes such as crop type now, so the classification and recognition of field pictures are achieved at the first stage of delineation. Combining the distribution of the quantity of categories of images which mainly include vegetation, urban landscape, desert, rivers, lakes, and seas, etc., four categories are made as the cases of LC CCI validation, including: terrestrial vegetated area (A1), artificial surface area (B15), bare area (B16), and aquatic or regularly flooded area (B2). The category of LCCS is supported to be the classification criterion for images and land cover.

2.3.2. Natural Scene Images Classification Model

Based on the category definition in Section 2.3.1, the automatic recognition model for natural scene images was constructed and trained. The model applied GoogLeNet Inception V3 [16] as the base model, namely the Natural Scene Images Classification Model (NSIC-Inception), removed the last fully connected layer of the base model, and added a maximum pooling layer and two fully connected layers. The last layer activation function was set to softmax. The overall structure of the model is shown in Figure 5. The NSIC-Inception model consists of convolutional layers, pooling layers, Inception modules, fully connected layers, and a softmax layer in turn.

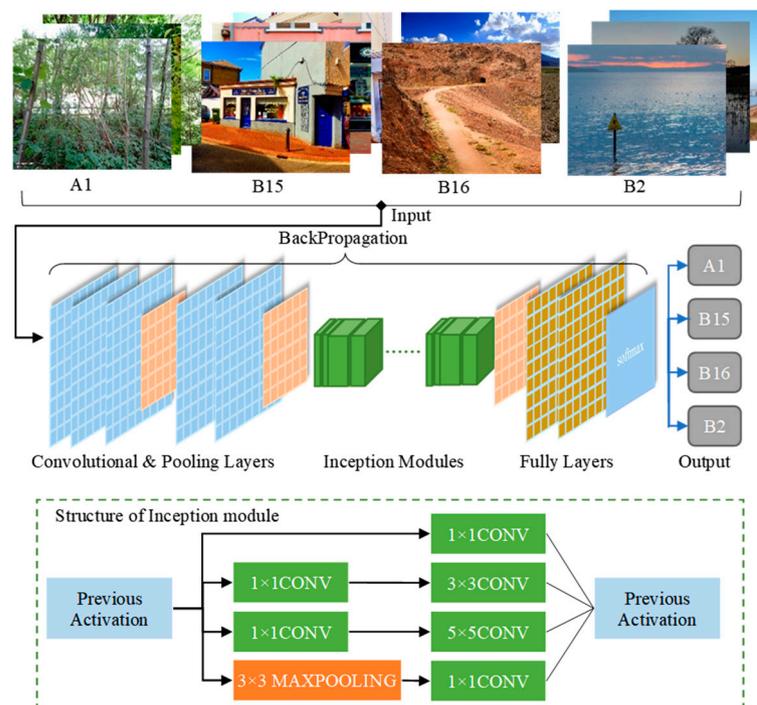


Figure 5. Structure of NSIC-Inception model based on transfer learning.

The Inception module is formed by arranging the convolutional and pooling layers according to a certain structure (Figure 4. Structure of inception module). It contains four branches: the first branch performs 1×1 convolution, which can organize information across channels and enhance the expressiveness of the model; the second branch performs feature extraction twice on the input features using two layers of convolution, 1×1 and 3×3 ; the third branch processes the input features twice using 1×1 and 5×5 convolution, and the fourth branch performs 1×1 convolution after 3×3 maximum pooling. These four branches can be processed in parallel. Finally, the input channels of each branch are merged. The inception module uses three different sizes of convolution kernels, 5×5 , 3×3 , 1×1 , and 3×3 maximum pooling kernels to design a sparse network structure that not

only improves processing efficiency, but also increases the model's processing of multi-scale features.

In the training of the model, the inputted natural scene images are processed by several convolution layers and pooling layers as well as multi-scale feature extraction by Inception modules to obtain the feature map. They are then passed into the fully connected layer for feature mapping, and finally the feature vector is passed into softmax to calculate the classification probability. The transfer learning based on ImageNet weights are used to migrate the features and parameters to achieve a better model fit in a smaller natural scene images dataset. The error between the true label and the predicted label is calculated using cross-entropy loss, and the error is returned layer by layer to calculate the corresponding error for each layer, and the network weights are adjusted in a continuous cycle. The specific optimization objective function can be expressed as follows.

$$L_c = \frac{1}{m \sum_{i=1}^n \sum_{k=1}^K y_{ik} \log \tilde{y}_i} \quad (1)$$

where m represents the number of samples participating in training, and y and \tilde{y} are the true and predicted labels of the one-hot vector, respectively.

A total of 9144 images from Flickr and LUCAS were used to build the model dataset, and the training set, validation set, and test set were divided according to the ratio of 8:1:1. The method of manual tuning was selected to determine the hyperparameters of CNN. The ImageNet trained weights as the initial weights for the CNN network, the batch size was set to 32, and the learning rate was set to 0.01. The images were resized to 299×299 to fit the input size of the InceptionV3 model, and the backpropagation algorithm was used in the process of weight optimization using batch gradient descent. An early end function was set to end the training when the accuracy could not be improved in two epochs. Cross-entropy loss and accuracy were used to monitor the model training process, and precision and recall were used to quantify the performance of the model in the test set.

To retain reliable images for verification of land cover datasets, images were filtered by the probability of the image's classification before being used to verify the land cover data. The probability threshold for model classification was set to 0.9. A higher threshold would make the results more reliable, but this reduces the quantity of valid results. Therefore, a threshold of 0.9 was taken [11].

2.4. Verification Land Cover Maps with Natural Scene Images

In this paper, the land cover type of LC-CCI data was first extracted using a pre-sampling method based on the geo-tags of natural scene photos, and then verified by the natural scene photo categories using reference data as "ground truth", establishing a confusion matrix [27,28].

The pre-sampling method based on image location is the most convenient sampling method to validate land cover maps using geo-tagged images and can be used to validate land cover data directly by a confusion matrix. The confusion matrix is a statistical matrix which counts and summarizes the number of consistent and inconsistent land cover and "ground truth" (geo-tagged natural scene images) data, and then subdivide them by category (as shown in Table 2). Each row of the confusion matrix indicates the number of samples in land cover map type, and each column indicates the number of samples in that category of "ground truth". The metrics from confusion matrix were used to express the verification results including overall accuracy, user accuracy, product accuracy, and Kappa coefficient [27,29]. Overall accuracy indicates the probability that the classification result for each random land cover sample agrees with the field image type. The product accuracy of each category is the degree of consistency of the land cover category with the ground truth. User accuracy is the degree of consistency of the field image type with the land cover type. The Kappa coefficient is a quantitative description of the agreement between the land

cover data to be evaluated and the reference images. The confusion matrix is as Table 3 and formula are as follows.

$$OA = \frac{\sum_{i=1}^n x_{ii}}{N} \quad (2)$$

$$PA = \frac{x_{ii}}{x_{+i}} \quad (3)$$

$$UA = \frac{x_{ii}}{x_{i+}} \quad (4)$$

$$Kappa = \frac{N \sum_{i=1}^n x_{ii} - \sum_{i=1}^n (x_{i+} x_{+i})}{N^2 - \sum_{i=1}^n (x_{i+} x_{+i})} \quad (5)$$

where n represents the total number of feature classes of land cover data involved in the assessment; x_{ii} refers to the values on the main diagonal of the confusion matrix; x_{i+} and x_{+i} are the total number of samples in row i and column i , respectively; N is the overall number of samples in the accuracy assessment.

Table 3. Confusion matrix.

		Land Cover Map					
		x_{11}	x_{21}	...	x_{i1}	...	x_{n1}
Ground truth		x_{12}	x_{22}	...	x_{i2}	...	x_{n2}
	
		x_{1i}	x_{2i}	...	x_{ii}	...	x_{ni}
	
		x_{1n}	x_{2n}	...	x_{in}	...	x_{nn}

Kappa coefficient takes values in the range of $[-1, 1]$ and is generally classified as $[0.4, 0.6]$ for moderate consistency, $[0.6, 0.8]$ is highly consistent, $[0.8, 1]$ is almost completely consistent, and a value closer to 1 means better consistency.

3. Results and Discussion

3.1. LUCAS and Flickr Images Classification

The automatic classification model of natural scene images is the key under this framework, which will directly determine the reliability of the land cover data verification results. The NSIC-Inception model was used for the photo classification accuracy of the test set, and the results showed that the top1 accuracy of the test set was 95.48%, and the top-3 accuracy was 99.47%. The precision and recall rates of each category are shown in Table 4.

Table 4. Precision and recall of test set classification.

Category	Precision	Recall
A1	99.80%	93.69%
B15	95.07%	98.97%
B16	60.32%	90.47%
B2	94.52%	97.18%
OA	95.21%	

The accuracy of A1, B15, and B2 were all above 90%, with A1 having the highest accuracy rate of 99.80%. However, B16's precision was lower, only 60.32%. The recall rate of all four categories was above 90%, the recall rate of B16 was lower compared to the other three, and the recall rate of A1, B2, and B15 increased in order.

The NSIC-Inception model, resnet50 model and the VGG16 model were used for the same transfer learning, and the model accuracy and land cover verification accuracy of

both (Table 5) were compared. There is a clear difference in the structure of the three models. The NSIC-Inception model sets multiple convolutional kernels of different sizes in parallel to extract image features, while VGG16 is a typical deep network structure consisting of five groups of convolutions, two fully connected, and one classification layer. All convolutional layers use 3×3 convolutional kernels and take a vertical expansion and deepening approach to build the network. The Resnet50 model is designed to overcome the problem of inefficient learning and ineffective accuracy due to the deepening of the network by using a shortcut connection to introduce the data output from one of the previous layers directly into the input part of the later layers by skipping several layers.

Table 5. Comparison of verification results of different models.

Model (Pre-Training)	Accuracy	OA with Flickr	OA with LUCAS
NSID-Inception	94.38%	61.75%	94.41%
Resnet50	93.59%	56.63%	92.49%
VGG16	88.98%	43.08%	86.20%

Both the accuracy of the model and the agreement of the image classification results with CCI-LC with NSIC-Inception pre-training were slightly higher than those of the VGG16 pre-trained model and Resnet50 model. Moreover, the agreement of the LUCAS images was much higher than those of the Flickr images. This showed that although the natural scene images classification model has an impact on the verification results of the land cover data, the impact of different images sources on the verification results is more pronounced than that of the image classification model.

In summary, the NSIC-Inception model proposed in this study had better performance and could classify the natural scene images of categories A1, B15, and B2 with high accuracy to satisfy the verification of the land cover map. Therefore, the 12,879 natural scene images used for LC-CCI verification were classified by NSIC-Inception model, and only those with classification confidence greater than 0.9 were retained. The results are shown in Table 6.

Table 6. Classification results of the natural scene image datasets in the UK.

Category	Flickr Images in 2017	LUCAS Images in 2015
A1	539	7719
B15	623	487
B2	521	108
Sum	1683	8314

As can be seen from Table 6, a total of 9997 geo-tagged natural scene images were classified with a confidence level of more than 0.9. Category A1 accounts for about 92.84% of 2015 LUCAS image datasets, 7719 images. While the quantity of B2 images in the set was only 108, accounting for only 1.30%. Since the LUCAS survey was a spatially homogeneous sampling, both showed the uneven feature that the quantity of images in category A1 was much higher than the quantity of images in other categories. In contrast, the quantity of images distributed across categories on Flickr in 2017 was more even.

3.2. LC CCI Verification and Analysis in the UK

The verification of the land cover map requires the spatial locations of samples first. Then, it is possible to verify the land cover types with “ground truth” of the sampled points, and to calculate the quantity and proportion of samples with their classification. The spatial pre-sampling method based on the image spatial distribution was used, and the samples were laid out with the same spatial distribution as the geo-tagged images. The geo-tagged 2017 Flickr images were used to sample the LC CCI in 2017, and the geo-tagged 2015 LUCAS images were used to sample the LC CCI in 2015. The distribution of the quantity of land cover categories at the sampling points is shown in Table 7.

Table 7. Distribution of pre-sampling categories of LC CCI maps.

LC CCI Category	Sample Size in 2017	Sample Size in 2015
A1	886	7759
B15	466	448
B2	331	107
Sum	1683	8314

The very small percentage (0.24%) of the area covered by class B16 in the UK resulted in even fewer samples for B16 in the pre-sampling. It has been shown that the small sample size has an impact on the reliability of the verification results [30,31]. Consequently, type B16 was removed from the image datasets verification and A1, B15, and B2 were reserved for the agreement analysis in this study. As shown in Table 6, there were a total of 1683 LC CCI samples in 2017 and 8314 LC CCI samples in 2015.

A confusion matrix was built using pivot tables (Tables 6 and 7), shown in Figure 6. Natural scene images were used as a ground reference to judge whether the land cover data of the sample are correct.

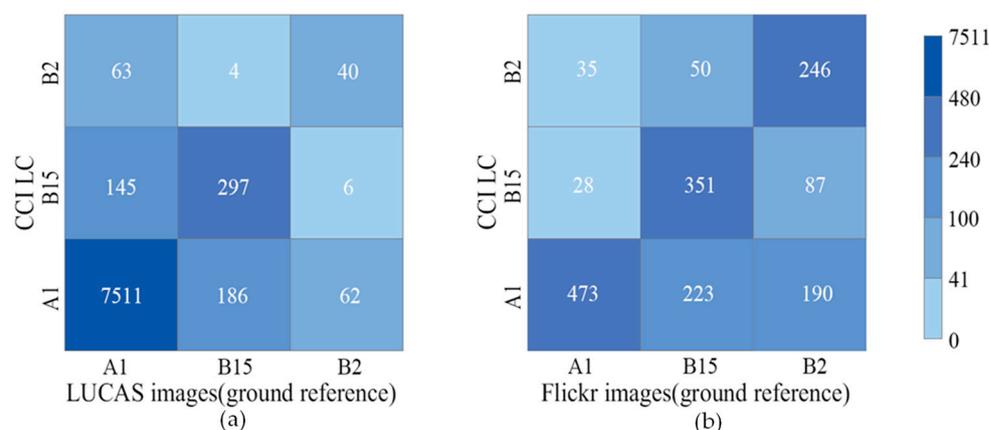


Figure 6. Confusion matrix between CCI LC and ground reference. (a) Confusion matrix between CCI LC and LUCAS images, (b) Confusion matrix between CCI LC and Flickr images.

Overall accuracy, product accuracy, user accuracy, and Kappa coefficient of map verification for LC CCI are shown in Table 8, respectively. The classifications of geo-tagged images were both based on the NSIC-Inception model. However, the verification results varied widely between LUCAS and Flickr.

Table 8. The verification results with LUCAS and Flickr images.

	OA (%)	A1	PA (%) B15	B2	A1	UA (%) B15	B2	Kappa
LUCAS	94.41	97.31	60.99	37.04	96.80	66.29	37.38	0.57
Flickr	65.58	88.25	56.25	47.04	53.39	75.32	74.32	0.46

The kappa coefficients, with the verification result of CCI LC with LUCAS and Flickr being 0.57, 0.46, respectively, showed medium consistency, but the Kappa coefficient of LUCAS was 0.11 higher than that of Flickr.

In terms of the overall accuracy of the land cover maps, the overall accuracy of the verification results for the LUCAS images was 30.83% higher than that for the Flickr images. However, the product accuracy difference of verification for each category in the LUCAS and Flickr image sets was at least 20.83% narrower than the difference in overall accuracy, and the consistency verification for each category behaved differently in different datasets,

which may be related to factors, such as the size of images sample, spatial distribution, and spatial heterogeneity of land cover.

3.3. Comparison and Analysis of LUCAS vs. Flickr

The differences between LUCAS and Flickr were compared and analyzed regarding the quantity of images, spatial distribution, representativeness of images, and the camera angle.

In order to quantitatively measure the four impacts, the verification area was divided by grids of 3000×3000 m. The quantities of LUCAS and Flickr image and the number of land cover types were counted within the grid. The number of land cover types was between 1 and 4, which meant 1–4 land cover types (A1, B15, B16, B2) within each grid, which was used to express the spatial heterogeneity of land cover. The average accuracy of verification in each grid and in each spatial heterogeneity of land cover were calculated and expressed as the ratio of the number of verified consensuses to the total number of images.

(1) The quantity of the images

Flickr had a much smaller total number of images than LUCAS. For category A1 images in particular, Flickr had 7180 fewer images than LUCAS, as shown in Table 5. The PA of LUCAS was higher than Flickr, at 9.06%. This huge volume gap made the PA and OA of Flickr lower than that of LUCAS. Similar to category A1, Flickr had 224 more images than LUCAS in category B2, and the PA of Flickr was higher than that of LUCAS than 10.0%. Thus, quantity of images plays an important role in the verification results.

(2) The spatial distribution of images

The spatial distribution of sample is completely determined by the distribution of images using images for verification. The spatial distribution of LUCAS and Flickr images collections was compared. LUCAS takes photos according to certain sampling rules, while Flickr photo spatial distribution is random. Hence, this could have a certain influence on the verification of the land cover maps [32].

As shown in Figure 7, LUCAS mainly distributed areas within one or two land cover types, accounting for 57% and 37% respectively. Flickr's photos accounted for 25% and 45% of the one and two land cover types. There were more Flickr images than LUCAS images in areas with higher spatial heterogeneity of land cover, which may result in verification inconsistency more often.

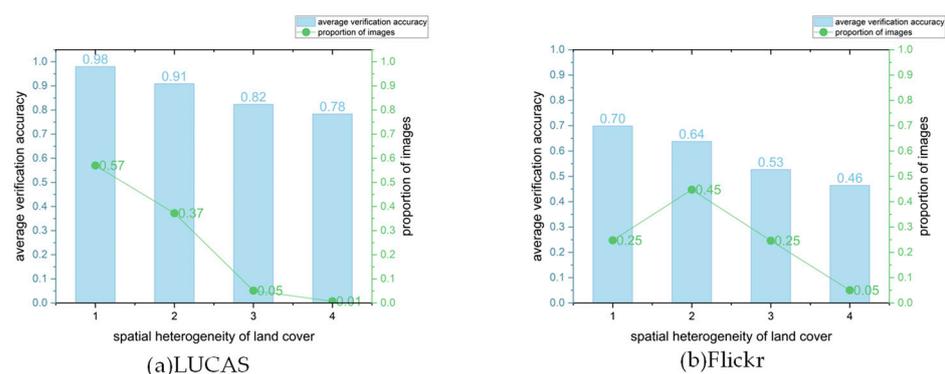


Figure 7. The proportion of image and average verification accuracy in different spatial heterogeneity of land cover. (The average verification accuracy was represented by blue bar graphs, and the proportion of images was represented by green folded line chart) (a) The LUCAS images, (b) The Flickr images.

In addition, the average verification accuracy in the low spatial heterogeneity grid is higher than that in the high spatial heterogeneity grid, for both LUCAS and Flickr. As the spatial heterogeneity increases from 1 to 4, the LUCAS's average verification accuracy

decreases from 0.98 to 0.78, and Flickr’s average verification accuracy decreases from 0.70 to 0.46.

It could be assumed that the consistency of the verification results was related to the diversity of the land cover at the location of the samples, and the lower spatially heterogeneous the images are distributed over the area of land cover, the more reliable the validation of this method is. Maybe in areas with high spatial heterogeneity of land cover, the verification of land cover requires higher representativeness of images.

(3) The representativeness of images

Theoretically, the layout of sampling points should be combined with the spatial heterogeneity of land cover. The locations with high spatial heterogeneity need more samples, while the locations with low spatial heterogeneity need fewer samples. In other words, the sample points should be as representative as possible. There was a gap in terms of representativeness between Flickr images and LUCAS images, which can be confirmed from Figures 8 and 9.

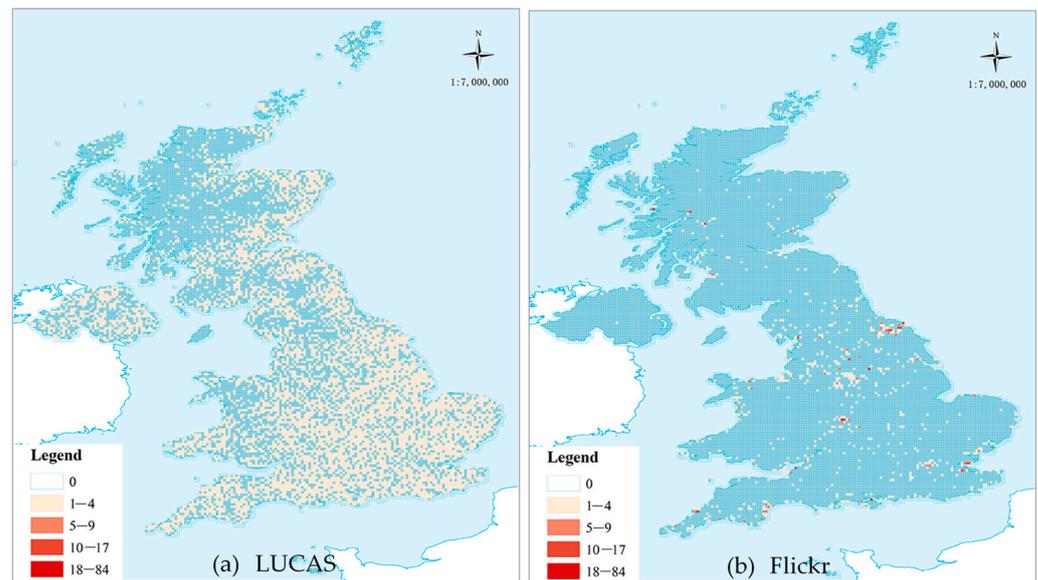


Figure 8. Spatial distribution of the image quantity in grid of 3000×3000 m. (a) The spatial distribution of LUCAS images, (b) The spatial distribution of Flickr images.

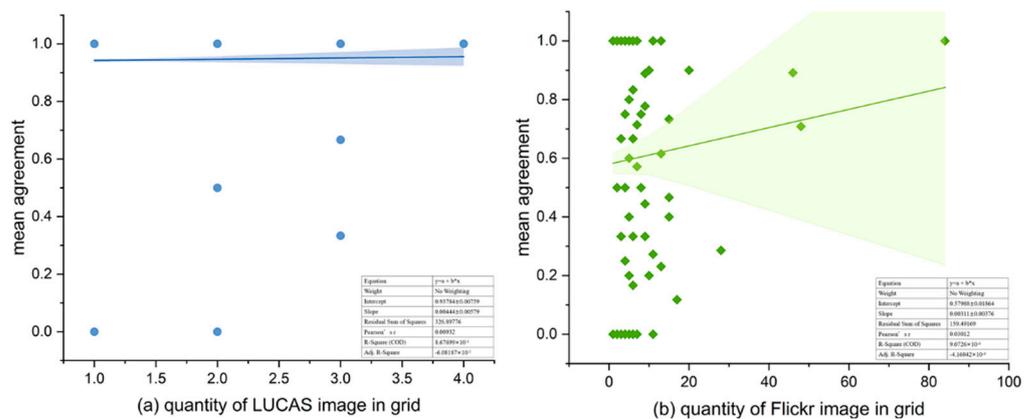


Figure 9. Scatter plot of quantity of images versus average agreement within the grid of 3000×3000 m. (The lines indicate linear fits to the data, the band around the line is the 0.95 confidence interval) (a) The quantity of LUCAS images versus average agreement, (b) The quantity of Flickr images versus average agreement.

The quantity of LUCAS and Flickr image in the grid of 3000×3000 were counted and then plotted based on the UK map, as shown in Figure 8. A scatter plot of the quantity of images and the average accuracy in different grids was plotted, and a linear fit was analyzed for it within the 0.95 confidence intervals, as shown in Figure 9.

From Figure 8a, LUCAS images were distributed in almost the whole UK, and there were only 1–4 images in each grid. However, the average consistency within the grid is as high as 93.78%, as shown Figure 9a. This indicates that the LUCAS image, as a record of land cover type, is highly representative.

From Figure 9b, Flickr images had a limited distribution range, mostly gathered in London, Manchester, York, and other large cities in England. The quantity of images in the grid ranged from 0 to 84, mostly concentrated in 0–20 images, but the corresponding average consistency also fluctuated in a wide range, and the overall starting value was only 57.99%, 35.79% lower than LUCAS. This suggests that as a record of land cover type, Flickr images are still not spatially representative enough. The coefficient of the fitted line is positive, and the degree of consistency increases with the increase of the quantity of images. That is to say, although Flickr images do not have very typical spatial representation of land cover type, the increase in the quantity of images in a unit grid could increase the verification result and thus be more reliable.

(4) Camera Angle

The images of LUCAS and Flickr demonstrated different performances in the same land cover verification case due to different shooting purposes, different photographers and different sharing methods. In addition to the above reasons obtained from analysis, there were also some subjective behaviors.

For example, LUCAS shooting group had a clear requirement to find a fixed shooting angle before shooting. In this study, images from a central perspective were used. Such a shooting angle represented the main content of the pictures mostly because given the surface or close natural scenes, it can reflect the “ground truth”. However, the images shared by the public on Flickr may be tourist records or special scene commemorations, such as buildings clearly standing in the grass and beautiful flower pools in the city. The NSIC-Inception model can accurately identify buildings as B15 land and flower pools as A1 land cover. However, it turns out that grassland and artificial land cover are the types of land cover in the 300-m resolution map. There are probably a lot of these images in Flickr. Such a minimum mapping unit for photo content records and a minimum unit for land cover records may be more common in Flickr collections than in LUCAS.

Therefore, the differences caused by these subjective factors also affect the verification results with land cover dataset.

3.4. Verification CCI LC and GLC-FCS

The verification of GLC-FCS was compared with that of CCI LC using LUCAS which had a more reliable result than Flickr. The spatial resolution of GLC-FCS dataset is different from CCI LC and can be explored with respect to the impact of spatial heterogeneity on the method proposed.

The results of verification CCI LC and GLC-FCS using LUCAS images were compared. The confusion matrix of GLC-FCS and the difference between the confusion matrix of GLC FCS and that of CCI LC were drawn, as shown in Figure 10. When the spatial resolution of land cover was improved, the land cover sample points of B15 category were increased, and compared with CCI LC, there were 71 new verification samples whose photo category was consistent with the land cover category. However, the sample points of B2 and A1 were reduced, and the verification sample points of photo category consistent with land cover category were also reduced by 23 and 170, respectively.

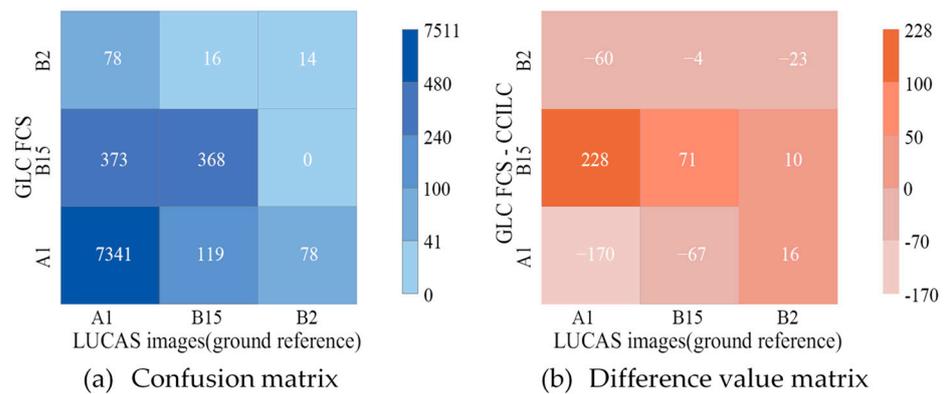


Figure 10. (a) Confusion matrix between GLC-FCS and LUCAS, (b) Difference value matrix GLC-FCS minus CCI LC.

As shown in Table 9, The overall accuracy of GLC-FCS was slightly lower than that of CCI LC in the verification. It should be noticed that the PA of B2 was higher in GLC-FCS than that of CCI LC. Both PA and UA of A1 are slightly increased or decreased, which may be caused by the error of the product itself. The sample points of B2 were too few, so we will not further explain it here.

Table 9. Comparison of verification results between CCI LC and GLC-FCS.

	OA (%)	PA (%)			UA (%)			Kappa
		A1	B15	B2	A1	B15	B2	
CCI LC	94.41	97.31	60.99	37.04	96.80	66.29	37.38	0.57
GLC-FCS	92.89	95.31	75.56	12.96	97.39	48.61	82.35	0.52

We presumed that the change in the verification results of B15 was due to the higher spatial heterogeneity of artificial surfaces land cover types, such as cities, parks, neighborhoods, etc. Land cover maps with higher resolution could make the images close to the mapping units of land cover. Thus, PA verification results were improved with regard to land cover with 300 m spatial resolution, as shown in Figure 11.

In CCI LC land cover, A1 land cover was found near the sampling points, while A1 and B15 land cover were found near the sampling points in GLC-FCS land cover. Comparing the land cover at high and low resolution, B15 shows a stronger spatial heterogeneity than A1, which is expected. The minimum mapping unit of land cover and that of images do not match. When the spatial resolution is increased from 300×300 to 30×30 , many verification results of the sample points will be positive, but not all. The key to this method was the matching of the smallest unit of mapping and that of the image.

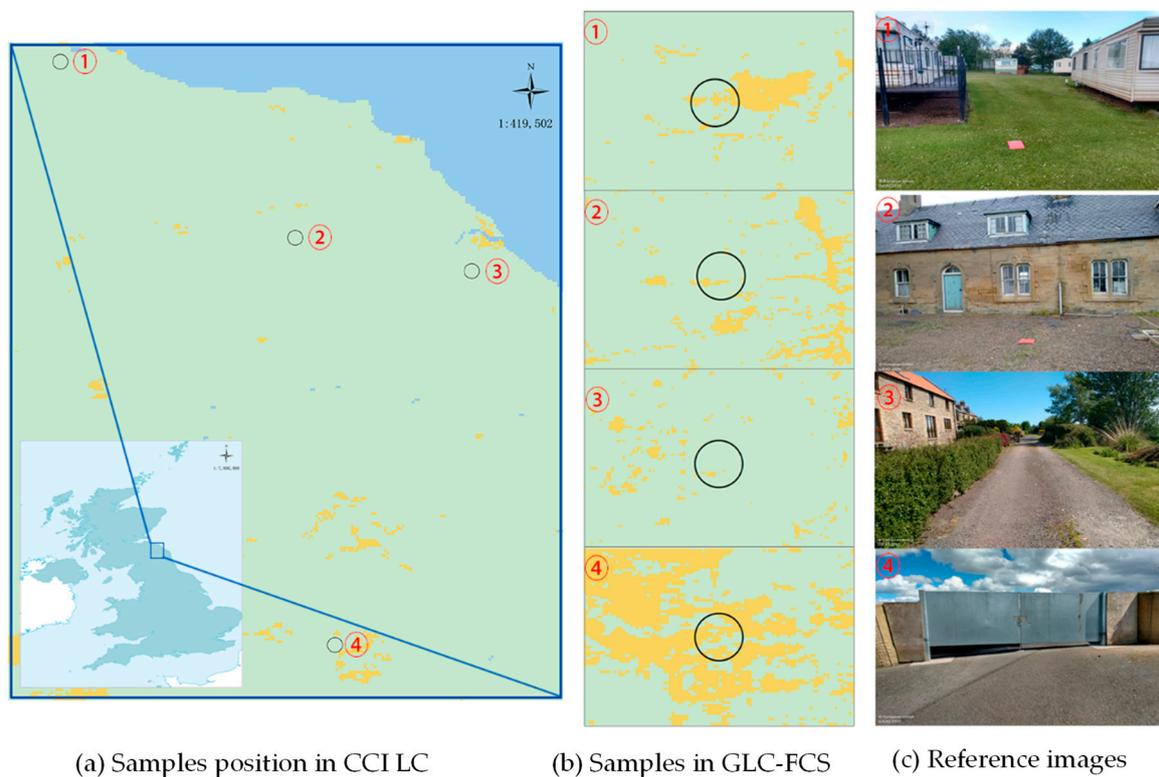


Figure 11. The difference verification result between CCI LC and GLC-FCS. (a) Samples position in CCI LC (samples were around by only A1 land cover), (b) Samples in GLC-FCS (samples were distributed in A1 and B15 land cover), (c) The natural scene images in samples position as reference.

4. Conclusions

Land cover maps represent the basis for various studies and applications, and ensuring the quality of the maps through verification is an essential step. However, this verification process requires a lot of manpower and money, and it is especially difficult to validate maps on a global scale. In this study, a framework was developed to improve the efficiency of land cover map verification by using the NSIC-Inception model to recognize images and directly evaluate the accuracy of land cover maps using a pre-sampling method based on image location.

The results show that LUCAS image performs much better than Flickr in the land cover verification. There are great differences between them in terms of quantity, spatial distribution, and representation in land cover. LUCAS images have a large number, strong representation in land cover, and uniform distribution. Compared with Flickr images, LUCAS images have less distribution in areas with high land cover heterogeneity. It is very suitable for land cover verification as a reference dataset based on NSIC-Inception model recognition, which is simpler than Flickr. The spatial distribution of Flickr was relatively concentrated, especially some special landscapes in the surface environment which were often recorded and shared, so the minimum mapping unit of the image has obvious conflicts with the land cover map. Moreover, the verification result will be less accurate than that of the land cover product if it is directly used for the verification of land cover. In other words, the accuracy of land cover products will be underestimated. However, the analysis shows that when the quantity of Flickr images increases, its verification result will be improved, making the results more reliable.

The comparison between different spatial resolution of land cover data showed that the minimum mapping unit of land cover and that of images do not match. When the spatial resolution is increased from 300 to 30, many of the verification results of sample points will be positive in category B15, but the opposite is true for A1 and B2. The key to this method is the matching of smallest unit of mapping and that of the image.

In this paper, only three land cover types were used to validate and discuss the land cover data. In the future, the training of models in more detailed categories will be needed to adapt the verification of land cover maps in more fields, and more detailed classification methods can still refer to the LCCS classification standard, which can promote the construction and use of the global land cover images reference datasets.

Author Contributions: Conceptualization, Hui Yang; Data curation, Liang Chu; Formal analysis, Fei Xu and Zhaojin Yan; Investigation, Liang Chu and Ran Wang; Methodology, Liu Cui and Hui Yang; Software, Liu Cui; Supervision, Hui Ci; Validation, Qingping He and Yina Qiao; Writing—view original draft, Hui Yang; Writing—review & editing, Liu Cui. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Third Comprehensive Scientific Investigation Project of Xinjiang (2022xjkk1006), the National Natural Science Foundation of China (grant number 41971335 and 51978144), the Xinjiang Uygur Autonomous Region Key Research and Development Program (2022B01012-1), the Graduate Innovation Program of China University of Mining and Technology (2022WLKXJ036) and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX22_2603).

Data Availability Statement: Natural Scene Image Data in this experiment can be downloaded at the platform Flickr (<https://www.flickr.com/> (accessed on 28 May 2022)) and the Land Use and Cover Area frame Survey (<https://ec.europa.eu/eurostat/web/lucas/data> (accessed on 19 October 2021)); Land Cover datasets can be found at European Space Agency Climate Change Initiative project (<http://maps.elie.ucl.ac.be/CCI/viewer/download.php> (accessed on 2 February 2022)); Global land-cover product with fine classification system (GLC-FCS) is available at <https://zenodo.org/record/3986872> (accessed on 10 October 2022) All data and materials are available for publication.

Acknowledgments: Thanks to the European Space Agency Climate Change Initiative Land Cover project for the Land Use and Land Cover dataset.

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

References

1. Aneseyee, A.B.; Soromessa, T.; Elias, E.; Noszczyk, T.; Feyisa, G.L. Evaluation of Water Provision Ecosystem Services Associated with Land Use/Cover and Climate Variability in the Winike Watershed, Omo Gibe Basin of Ethiopia. *Environ. Manag.* **2022**, *69*, 367–383. [[CrossRef](#)] [[PubMed](#)]
2. Qiao, R.; Dong, C.; Ji, S.; Chang, X. Spatial Scale Effects of the Relationship between Fractional Vegetation Coverage and Land Surface Temperature in Horqin Sandy Land, North China. *Sensors* **2021**, *21*, 6914. [[CrossRef](#)] [[PubMed](#)]
3. Szarek-Iwaniuk, P.; Dawidowicz, A.; Senetra, A. Methodology for Precision Land Use Mapping towards Sustainable Urbanized Land Development. *Int. J. Environ. Res. Public Health* **2022**, *19*, 3633. [[CrossRef](#)] [[PubMed](#)]
4. Ma, T.; Li, X.; Bai, J.; Ding, S.; Zhou, F.; Cui, B. Four decades' dynamics of coastal blue carbon storage driven by land use/land cover transformation under natural and anthropogenic processes in the Yellow River Delta, China. *Sci. Total Environ.* **2019**, *655*, 741–750. [[CrossRef](#)] [[PubMed](#)]
5. Tarko, A.; Tsendbazar, N.; de Bruin, S.; Bregt, A. Influence of image availability and change processes on consistency of land transformation interpretations. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *86*, 102005. [[CrossRef](#)]
6. Pengra, B.; Gallant, A.L.; Zhu, Z.; Dahal, D. Evaluation of the Initial Thematic Output from a Continuous Change-Detection Algorithm for Use in Automated Operational Land-Change Mapping by the U.S. Geological Survey. *Remote Sens.* **2016**, *8*, 811. [[CrossRef](#)]
7. D'Andrimont, R.; Yordanov, M.; Lemoine, G.; Yoong, J.; Nikel, K.; van der Velde, M. Crowdsourced Street-Level Imagery as a Potential Source of In-Situ Data for Crop Monitoring. *Land* **2018**, *7*, 127. [[CrossRef](#)]
8. Bayas, J.C.L.; See, L.; Bartl, H.; Sturn, T.; Karner, M.; Fraisl, D.; Moorthy, I.; Busch, M.; Van Der Velde, M.; Fritz, S. Crowdsourcing LUCAS: Citizens Generating Reference Land Cover and Land Use Data with a Mobile App. *Land* **2020**, *9*, 446. [[CrossRef](#)]
9. Fritz, S.; See, L.; McCallum, I.; You, L.; Bun, A.; Moltchanova, E.; Duerauer, M.; Albrecht, F.; Schill, C.; Perger, C.; et al. Mapping global cropland and field size. *Glob. Chang. Biol.* **2015**, *21*, 1980–1992. [[CrossRef](#)]
10. Xu, G.; Zhu, X.; Fu, D.; Dong, J.; Xiao, X. Automatic land cover classification of geo-tagged field photos by deep learning. *Environ. Model. Softw.* **2017**, *91*, 127–134. [[CrossRef](#)]
11. Xing, H.; Meng, Y.; Wang, Z.; Fan, K.; Hou, D. Exploring geo-tagged photos for land cover validation with deep learning. *ISPRS J. Photogramm. Remote Sens.* **2018**, *141*, 237–251. [[CrossRef](#)]
12. Pal, M.; Foody, G.M. Feature Selection for Classification of Hyperspectral Data by SVM. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2297–2307. [[CrossRef](#)]

13. Sitthi, A.; Nagai, M.; Dailey, M.; Ninsawat, S. Exploring Land Use and Land Cover of Geotagged Social-Sensing Images Using Naive Bayes Classifier. *Sustainability* **2016**, *8*, 921. [[CrossRef](#)]
14. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *NIPS* **2017**, *60*, 84–90. [[CrossRef](#)]
15. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; Volume 37, pp. 448–456.
16. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
17. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2015**, arXiv:1409.1556.
18. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
19. Xu, S.; Zhang, S.; Zeng, J.; Li, T.; Guo, Q.; Jin, S. A Framework for Land Use Scenes Classification Based on Landscape Photos. *Proc. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 6124–6141. [[CrossRef](#)]
20. Zhu, Y.; Newsam, S. Land Use Classification using Convolutional Neural Networks Applied to Ground-Level Images. In Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, Seattle, WA, USA, 3–6 November 2015; pp. 1–4.
21. Oba, H.; Hirota, M.; Chbeir, R.; Ishikawa, H.; Yokoyama, S. Towards Better Land Cover Classification Using Geo-tagged Photographs. In Proceedings of the 2014 IEEE International Symposium on Multimedia, Washington, DC, USA, 10–12 December 2014; pp. 320–327. [[CrossRef](#)]
22. Defourny, P.; Kirches, G.; Brockmann, C.; Boettcher, M.; Peters, M.; Bontemps, S.; Lamarche, C.; Schlerf, M.; Santoro, M. Land Cover CCI: Product User Guide Version. 2016. Available online: http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf (accessed on 5 March 2016).
23. Zhang, X.; Liu, L.; Chen, X.; Gao, Y.; Xie, S.; Mi, J. GLC_FCS30: Global land-cover product with fine classification system at 30 m using time-series Landsat imagery. *Earth Syst. Sci. Data* **2021**, *13*, 2753–2776. [[CrossRef](#)]
24. Ahlqvist, O. In Search of Classification that Supports the Dynamics of Science: The FAO Land Cover Classification System and Proposed Modifications. *Environ. Plan. B Plan. Des.* **2008**, *35*, 169–186. [[CrossRef](#)]
25. Hill, R.A.; Smith, G.M. Land cover heterogeneity in Great Britain as identified in Land Cover Map. *Int. J. Remote Sens.* **2005**, *26*, 5467–5473. [[CrossRef](#)]
26. Yang, H.; Li, S.; Chen, J.; Zhang, X.; Xu, S. The Standardization and Harmonization of Land Cover Classification Systems towards Harmonized Datasets: A Review. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 154. [[CrossRef](#)]
27. Yang, Y.; Xiao, P.; Feng, X.; Li, H. Accuracy assessment of seven global land cover datasets over China. *ISPRS J. Photogramm. Remote Sens.* **2017**, *125*, 156–173. [[CrossRef](#)]
28. Qian, T.; Kinoshita, T.; Fujii, M.; Bao, Y. Analyzing the Uncertainty of Degree Confluence Project for Validating Global Land-Cover Maps Using Reference Data-Based Classification Schemes. *Remote Sens.* **2020**, *12*, 2589. [[CrossRef](#)]
29. Saha, A.; Pal, S.; Arabameri, A.; Blaschke, T.; Panahi, S.; Chowdhuri, I.; Chakraborty, R.; Costache, R.; Arora, A. Flood Susceptibility Assessment Using Novel Ensemble of Hyperpipes and Support Vector Regression Algorithms. *Water* **2021**, *13*, 241. [[CrossRef](#)]
30. Nguyen, L.H.; Joshi, D.R.; Clay, D.E.; Henebry, G.M. Characterizing land cover/land use from multiple years of Landsat and MODIS time series: A novel approach using land surface phenology modeling and random forest classifier. *Remote Sens. Environ.* **2020**, *238*, 111017. [[CrossRef](#)]
31. Stehman, S.V.; Wickham, J.D.; Fattorini, L.; Wade, T.D.; Baffetta, F.; Smith, J.H. Estimating accuracy of land-cover composition from two-stage cluster sampling. *Remote Sens. Environ.* **2009**, *113*, 1236–1249. [[CrossRef](#)]
32. Ishii, Y.; Iwao, K.; Kinoshita, T. Global Land Cover Assessment Using Spatial Uniformity Validation Dataset. *Remote Sens.* **2021**, *13*, 2950. [[CrossRef](#)]