A deep relearning method based on the recurrent neural network for land cover classification

Yunwei Tang†, Fang Qiu†, Bangjin Wang, Di Wu, Linhai Jing and Zhongchang Sun

†International Research Center of Big Data for Sustainable Development Goals, Beijing, China; ‡Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China; §Geospatial Information Sciences, The University of Texas at Dallas, Richardson, TX, USA; ¶Planning and Finance Department, Zhejiang Academy of Surveying and Mapping, Hangzhou, China; ¶Basic surveying and Mapping Center, Zhejiang Academy of Surveying and Mapping, Hangzhou, China

ABSTRACT
Recent developments in deep learning (DL) techniques have provided a series of new methods for land cover classification. However, most DL-based methods do not consider the rich spatial association of land cover classes embedded in remote sensing images. In this research, a deep relearning method based on the recurrent neural network (DRRNN) is proposed for land cover classification. The relearning approach has great potential to improve classification, which has never been used in DL-based land cover classification. To utilize the spatial association of the pixels’ information classes, a class correlated feature (CCF) is first extracted in a local window from an initial classification result. This feature can reflect both the spatial autocorrelation and spatial arrangement of land cover classes. Since the recurrent neural network (RNN) is designed to process sequential data, the CCF is formed as a feature sequence, allowing RNN to model the dependency between class labels. The relearning process is then applied to iteratively classify remote sensing images based on the CCF and spectral-spatial feature. At each relearning iteration, the CCF is learned from the previous classification result until a stopping condition is satisfied. This method was tested on five remote sensing images with different sensors and diverse environments. It was observed that noise in the classification result can be filtered by considering spatial autocorrelation, and misclassified areas can be corrected by incorporating spatial arrangement in the relearning process. The classification results indicate that compared to other state-of-the-art DL methods, the proposed method consistently achieves the highest accuracy.

1. Introduction
Image classification has been widely used for various applications in remote sensing. Numerous machine learning algorithms were developed for classification including maximum likelihood classifier (MLC), support vector machine (SVM), etc. To achieve a high classification accuracy, spatial features were introduced in classification such as a gray-level co-occurrence matrix (GLCM) (Haralick, Shanmugam, and Dinstein 1973), variogram features (Balaguer et al. 2010; Atkinson and Naser, 2010; Tang et al. 2019), etc. However, the extraction and selection of these representative handcrafted features require expert experience (Liang, Deng, and Zeng 2020), which would be a challenge with the increasing availability of imagery with various high spatial and/or spectral resolutions. To address this problem, the deep learning (DL) method was adopted by introducing artificial intelligence (AI). As a typical model in DL, the convolutional neural network (CNN) can automatically learn a variety of features from a large pool of image patches and has therefore been shown to be superior to traditional machine learning methods in feature selection.

Apart from CNN, another important branch of DL is the recurrent neural network (RNN). RNN is capable of dealing with sequential data by using previous outputs from a neural network as inputs, and thus has been widely used for time series analysis in remote sensing (lenco et al. 2017; Sun, Di, and Fang 2019; Feng et al. 2020; Khusni, Dewangkoro, and Arymurthy 2020; Zhu et al. 2021). A large number of spectral bands consist of a feature sequence; thus, the potential of RNN in hyperspectral image classification has also been explored (Mou, Ghamisi, and Zhu 2017; Wu and Prasad 2017; Hang et al. 2019; Ma et al. 2021). To introduce spatial information in hyperspectral images, CNN can be combined with RNN (Shi et al. 2015; Liu et al. 2017; Deng et al. 2018).
et al. 2017; Shi and Pun 2018). An alternative is to use the RNN itself to learn spatial information by taking a hyperspectral cube as a vector sequence and to model the dependence between these pixels (Liu et al. 2018a; Zhang et al. 2018b; Luo 2018).

Comparing the use of spatial features in machine learning and DL models, the former involves explicit handcrafted features (e.g., GLCM, variograms, etc.), while the latter automatically extracts spatial features by a series of convolutional operators. However, it is still unclear how the specifics of spatial features improve classification. In addition, different from many other applications in DL, classification images contain information classes with high spatial autocorrelation and typical spatial arrangement (e.g. self-similarity). This spatial association information is not considered in most DL methods. The class correlated feature (CCF) in the form of a class label sequence, which is extracted from a land cover classification, can reflect the local spatial association of pixels’ information classes. This feature can also be fed back to a classifier to improve its performance. The spatial correlation of land cover classes has been modeled from the CCF to enhance traditional classifiers such as k-nearest neighbors (k-NN) (Tang et al. 2016). To further explore its potential, a relearning method could be used to iteratively learn the CCF and recursively classify remotely sensed images. For example, a curve matching method was developed to improve classification by relearning the spatial association of class pairs (Tang et al. 2021). Previous studies also applied the SVM-based relearning by iteratively summarizing the histogram and primitive co-occurrence matrix (PCM) of the CCF (Huang et al. 2014). The relearning method has shown a great advantage in iteratively utilizing rich spatial association information embedded in the pixels’ information classes, but this method has never been exploited by DL.

In summary, DL methods (e.g. CNN and RNN) have been widely used for classification. However, problems still lie in several points. I) Most CNN methods extract spatial features based only on the convolutional operators. They were able to consider the spatial relationship of the original data values of the pixels (similar to texture), but unable to take advantage of the rich spatial association of resulted information classes of pixels. II) An RNN was often used to model temporal association but its potential in modeling the spatial association at the information class level has not yet been explored. III) The relearning approach has shown a great potential in modeling spatial associations using machine learning models (e.g. SVM), but it has never been tested in DL methods.

In our research, a deep relearning method is developed to incorporate the CCF for land cover classification based on an RNN model. The scientific contributions are as follows. I) A “deep relearning” model is developed to address the deficiency of current DL models by introducing a CCF-based spatial association and a relearning method based on RNN. II) An innovative relearning method designed to model spatial association of the information classes of the pixels is implemented using an RNN by taking advantage of its strength in modeling temporal association, greatly expanding its applicability. III) The introduction of CCF enables DL to model both spatial autocorrelation and spatial arrangement of classes for the first time, greatly improving the capability of DL models for image classification.

2. Methods

2.1 Recurrent neural network (RNN)

Since the developed method is based on an RNN, a brief review of RNNs is provided. First, time steps need to be defined in an RNN to record the elements in a feature sequence. The RNN is run to make a prediction at the current time step by considering the result derived from the previous one. Traditional RNNs often suffer from the problem of long-distance dependence (Bengio, Simard, and Frasconi 1994), which may cause gradient explosion and/or vanishing. Long short-term memory (LSTM) was proposed to avoid this problem by introducing a forget gate (Hochreiter and Schmidhuber 1997). LSTM can simulate the mechanisms of forgetting and memorizing as human brain, thus data can be disseminated in a certain distance (Luo 2018). A sequence of data \( (x_1, \ldots, x_n) \) is input into the RNN, where \( x_t \in (x_1, \ldots, x_n) \) represents the data (usually representing a feature sequence) at time step \( t \). The LSTM structure includes a memory unit \( c_t \), a hidden state \( h_t \), and three gates (input, forget and output
gates \(i_t, f_t, o_t\). A temporary cell state \(y_t\) is defined to resize the current input (Ndikumana et al. 2018). The memory unit \(c_t\) controls the kept information. The LSTM method is described as follows.

\[
\begin{align*}
    i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\
    f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\
    o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\
    y_t &= \tanh(W_y x_t + U_y h_{t-1} + b_y) \\
    c_t &= i_t \ast y_t + f_t \ast c_{t-1} \\
    h_t &= o_t \ast \tanh(c_t)
\end{align*}
\]

where \(W_i, W_f, W_o, W_y, U_i, U_f, U_o, U_y\) are weights, and \(b_i, b_f, b_o, b_y\) are biases; the asterisk (*) denotes the Hadamard product, and \(\sigma\) is a sigmoid function, which allows parts of the information to enter through the gates.

As a variant of LSTM, the gated recurrent unit (GRU) can achieve similar performance with a lower computational cost (Cho et al. 2014; Xiao et al. 2022). GRU includes an update gate \(z_t\) and a reset gate \(r_t\). The equations of GRU are as follows.

\[
\begin{align*}
    z_t &= \sigma(W_z x_t + U_z h_{t-1} + b_z) \\
    r_t &= \sigma(W_r x_t + U_r h_{t-1} + b_r) \\
    n_t &= \tanh(W_n x_t + U_n (r_t \ast h_{t-1}) + b_n) \\
    h_t &= z_t \ast h_{t-1} + (1 - z_t) \ast n_t
\end{align*}
\]

Similar to the LSTM method, \(W_z, W_r, W_n, U_z, U_r, U_n\) are weights, and \(b_z, b_r, b_n\) are biases, and \(n_t\) is defined as a candidate hidden state, storing the possible values to add to the current hidden state \(h_t\).

### 2.2 Class correlated feature (CCF)

In this study, the CCF is defined as a sequence of class labels that can reflect a local spatial association of the information classes. This feature is extracted from a prior land cover map, which can be obtained from a classification result. A template is used to extract the CCF from the prior land cover map in a local range. The class label is recorded in order along a certain direction (e.g. 0°, 90°, 180°, 270°) as a sequential feature, which is then directly fed into the RNN. Figure 1 demonstrates the process for extracting the CCF. Buildings and roads are easily misclassified, as the two classes have similar spectral features. In this example, the upper template is in a residential area, with a typical pattern of spatial arrangement in the form of “vegetation-shadow-building.” Dominated by this pattern, the pixels of the road class are unlikely to appear in such an area. If they do, they are likely to be reassigned to the correct class in a subsequent step. The lower template is located in a road area, where all the pixels are extracted as the road class, displaying a high spatial autocorrelation. The vehicles in the middle of the road, which are often misclassified as buildings based on the spectral feature only, are likely to be reclassified as the road class in the subsequent step. In summary, the CCF records the local spatial pattern; when fed into the RNN, it allows the spatial arrangement and spatial autocorrelation that exist in the remote sensing images to be fully utilized to improve the ability of the RNN classifier.

![Figure 1](image.png)

**Figure 1.** The process of extracting the CCF, along with the spatial arrangement and autocorrelation that are included in the CCF.
2.3 Deep relearning method based on the RNN (DRRNN)

A deep relearning method based on the RNN (DRRNN) is developed to recursively utilize the CCF to improve the classifier. We adapted the shortened spectral-spatial parallel RNN to derive the initial classification result. This method has been tested on the classification of hyperspectral images and has proven to be effective in achieving a satisfied result with a low computational cost (Luo 2018). The implementation of the DRRNN method is described below.

The training data are chosen randomly from the ground truth. Image patches are then extracted, centered on the training data, and used as the inputs in the network. Data augmentations, including reflections and rotations, are applied to enhance the training data. Three 3-dimensional (3D) convolutional layers are applied to capture the spatial features at multiple scale. The spectral-spatial feature is then fed into the parallel RNN, which consists of several RNN units to make the model more robust. The RNN unit with LSTM and GRU models are both tested. Following the RNN unit, a shortened model with a 1D convolutional layer is applied to boost the training speed. After the training process, the classification is performed on the entire image to derive an initial classification result. For each image patch centered on the training data, the predicted class labels are extracted from the image patch and shaped into a sequence as the CCF. Since image augmentation is applied, the CCF considers the order in different directions. Similarly, the shortened model with a 1D convolutional layer is applied to the CCF to reduce its dimension and thus can be combined with the spectral-spatial feature. These features are fed into the parallel RNN again to produce the next classification result. The above process is iteratively performed so that the CCF can be adjusted through deep relearning from the last predicted land cover map until a stopping condition is reached. In each relearning iteration, the validation accuracy is monitored (the validation sample is randomly selected from the training sample), and the relearning process stops if the increase in the validation accuracy is lower than a predefined threshold. The structure of the DRRNN method is displayed in Figure 2.

3. Results and analysis

3.1 Datasets and study areas

Throughout the study, five different remote sensing images containing diverse environments were tested. The first scene was acquired from the IKONOS satellite in May 2000, covering an urban area in Beijing, China. The spectral bands were enhanced to 1 m by applying image fusion. Six land cover classes were identified in this image. The second scene is a well-known hyperspectral image from the ROSIS sensor acquired in July 2002. This image had 103 bands with a 1.3 m resolution. Nine classes of interest were defined in this image. The third image is a Landsat 8 OLI subscene of Zhongwei City, China, acquired in April 2017. Similar to the IKONOS image, seven multispectral bands were increased to the 15 m resolution. Seven classes were

Figure 2. Flowchart of the DRRNN method.
used to analyze the Landsat image. The fourth image is a subscene of the Gaofen-2 (GF-2) image that was obtained from the Gaofen Image Dataset (GID) (Tong et al. 2020). In total, there were 15 annotated categories for GID, and in the selected subscene, there were 10 categories with four multispectral bands. The fifth image was derived from a Kaggle competition released by the Defense Science and Technology Laboratory (Dstl). Three-band images with RGB natural color were collected from the WorldView-3 (WV-3) sensor. A pseudo geo-coordinate was manually created by Dstl to obscure the location. Ten classes were defined for the WV-3 dataset, and six appeared in the selected subscene. The study areas have a distinctive range of environments (urban and suburban) with varied land cover classes. The five images include multispectral and hyperspectral bands (all the bands were included for classification), with resolutions varying from 15 m to 1 m, thus providing a robust evaluation of the developed method.

Figure 3 shows these five images and the associated ground truth data used. Ten percent of the ground truth was randomly selected as the training data using a stratified sampling scheme, and the remaining data were for the testing purpose. Image patches centered at each sample data point were extracted, which provided sufficient data for training (Luo 2018; Guo et al. 2020; Ding et al. 2021). The training sample information is displayed in Figure 4.

3.2 Experimental settings

To testify the effectiveness of the DRRNN method, eight advanced CNN methods were used for comparison. LeNet is representative of early CNNs, including two convolutional and two pooling layers, followed by three fully connected layers (LeCun et al. 1998). ResNet is a popular architecture (He et al. 2016) that introduces a shortcut connection that skips one or more layers to prevent the gradients to be vanished.

![Figure 3. The images of the study areas.](image-url)
or exploded. ResNet-18 includes the first convolutional layer, four modules consisting of four convolutional layers, and a fully connected layer, thus containing 18 layers in total. LeNet and ResNet-18 predict one label for each image patch and are therefore called patch-based CNN methods. The fully convolutional network (FCN) has been reported to be more efficient in utilizing the training information than patch-based CNN methods (Liu et al. 2018c). Therefore, we used the FCN with a self-defined structure and the well-known SegNet (Badrinarayanan, Kendall, and Cipolla 2017) as the pixel-based CNN for comparison. Both structures include an encoder and a decoder for a pixel-wise classification, while the FCN has one up-sampling layer, and the SegNet has three up-sampling layers for the decoder. FCN and SegNet predict all the pixels for an image patch; thus they are called pixel-based CNN methods. Because research indicates that multiscale CNN architecture is effective at capturing the semantics of an image (Gao et al. 2021; Yeung et al. 2022), two multiscale CNNs were used as benchmark methods. The multiscale network structure (MSNN) allows the network to accept the patches of varying sizes. The multiscale features are combined after two convolutional and pooling layers, attached with four fully connected layers. A wide contextual residual network (WCRN) includes a convolutional layer and a residual unit after multiscale features are combined (Liu et al. 2018b). The three input sizes of image patches were used in the MSNN and WCRN methods. Finally, LSTM and GRU (two typical RNN methods with shortened spectral-spatial parallel RNN units) were applied, but without the CCF and relearning process. For the proposed DRRNN method, the number of filters was set to 16 for both the 3D and 1D convolutional layers, as suggested by Luo (2018). The neuron numbers $H$ and the reduced time steps $T$ in

Figure 4. The training sample information of the study areas.
the shortened model were set to 128 and 3, respectively. Three parallel RNN units based on the LSTM and GRU models were used. The threshold of the stopping condition was set to 0.01%, which meant that the iteration would stop if the difference in the validation accuracy between the latest two runs was less than 0.01%. Overall, eight CNN methods were used as benchmarks, including two patch-based CNNs (LeNet and ResNet), two pixel-based CNNs (FCN and SegNet), two multiscale CNNs (MSNN and WCRN), and two RNNs (LSTM and GRU).

For all the CNN methods, the training data were augmented five times by flipping the images in two directions (horizontally and vertically), and rotating them in three directions (90°, 180°, 270°). The size of image patches $W$ was set to 17 based on the trial-and-error method (tested in Subsection 3.5), and the three input sizes were set to $W$, $W-2$, $W-4$ (i.e. 17, 15 and 13) for the MSNN, WCRN and two DRRNN methods to capture multiscale feature. The input batch included 128 images in the training stage. The maximum number of epochs equaled to $10^6$. Adam optimization was used to update the learning rate, which was set to $10^{-3}$. Half of the training data were used as validation data to monitor the accuracy and adjust the parameters of the classification methods. All the methods were run independently ten times for a robustness test, based on which the average overall accuracy (OA) was assessed. The confusion matrix was computed for each classification method, and the per-class producer’s accuracy (PA), user’s accuracy (UA) and F1-score were reported. The McNemar’s test was applied to determine whether the accuracy of the DRRNN method was significantly improved compared with the best benchmark method at the 95% significance level.

### 3.3 Classification results

The average OA over ten runs and the result of the McNemar test are shown in Table 1. The OA indicates that the DRRNN method constantly achieved higher accuracy for all five images. In most cases, the DRRNN was able to improve the accuracy of each class based on its corresponding RNN method. Compared with traditional CNN methods, the test demonstrated that eight out of the ten significant improvements over the best benchmark method (i.e. ResNet) could be achieved for the IKONOS image and all ten runs for the other four images, indicating a rather stable performance of the developed method.

The classification results from one run (selected randomly) are shown in the following figures along with the confusion matrices of each classification method and the PA, UA, and F1-score of each class (Figures 5–14). The confusion matrices are presented in the form of heat maps, and the color of the dialog elements was removed to highlight the confusion between classes.

As can be seen, the building and road classes in the IKONOS classification results were likely to be mixed with each other. The grass class was misclassified as trees in the LeNet, FCN and MSNN results, leading to a rather low UA for these three methods. In comparison, the ResNet, SegNet and WCRN methods have higher F1-scores on the road, grass, and tree classes (Figure 6). For the two RNN methods, LSTM and GRU do not show much difference. By introducing the spatial autocorrelation in the CCF, the DRRNN method acts as a filter to correct the misclassified pixels in the corresponding RNN results. In some cases, the DRRNN method is not affected by the majority class in a neighborhood. For example, nearly all the pixels in the two buildings of the GRU result (circles in

| Method      | IKONOS | ROSIS | Landsat | GF-2 | WV-3 |
|-------------|--------|-------|---------|------|------|
| LeNet       | 88.98  | 89.81 | 91.68   | 82.86| 79.78|
| ResNet      | 96.00* | 97.35*| 97.09   | 84.47| 83.91|
| FCN         | 91.52  | 91.95 | 92.43   | 82.88| 81.23|
| SegNet      | 95.25  | 97.18 | 94.27   | 84.31| 83.28|
| MSNN        | 89.25  | 88.70 | 93.47   | 84.60| 84.89*|
| WCRN        | 95.71  | 94.38 | 93.92   | 84.72*| 84.39|
| RNN with LSTM | 94.90 | 96.04 | 97.74*  | 83.02| 81.93|
| RNN with GRU | 93.02 | 94.29 | 97.74   | 83.18| 80.31|
| DRRNN with LSTM | 96.49 (8) | 98.67 (10) | 98.63 (10) | 89.62 (10) | 87.29 (10) |
| DRRNN with GRU | 97.18 (10) | 98.40 (10) | 98.69 (10) | 89.88 (10) | 87.00 (10) |
Figure 5i) are misclassified as roads, and most of these pixels are correctly classified by relearning the spatial arrangement in the CCF with the DRRNN (circles in Figure 5j). The DRNN method resulted in the highest F1-score across all classes, and the accuracy of the grass class improved the most (Figure 6).

Similarly, in the ROSIS image, the LeNet, FCN and MSNN methods tended to mix some spectrally similar classes, such as bare soil and meadows, asphalt and bitumen, and gravel and self-blocking bricks. The ResNet, SegNet and WCRN methods had higher F1-scores for these classes (Figure 8). For the RNN methods, the LSTM and GRU correctly classified most of the bare soil and meadow classes. However, the RNN with the GRU method resulted in a complete misclassification of the bitumen class (circle in Figure 7i). On the other hand, the DRRNN with the GRU method (circle in Figure 7j) was able to rectify most of the misclassified asphalt and gravel classes to the bitumen class. The DRRNN method improved the F1-score of the bitumen class to 0.9, indicating a great advantage in utilizing the CCF. The gravel class and self-blocking bricks also showed great improvements using the DRRNN methods (Figure 8).

In the Landsat image, vegetation and farmland tended to be mixed with one other. Among the eight benchmark methods, the LeNet, FCN, SegNet, MSNN and WCRN methods misclassified vegetation as
farmland due to a small sample size. Only the ResNet and two RNN methods demonstrated a clearer distinction between these two classes (middle two circles in Figure 9). The two DRRNN methods further decreased the noise of farmland within the vegetation class by introducing the spatial autocorrelation in the CCF. A similar result was observed with the road class (lower circle in Figure 9). The DRRNN with the GRU method correctly classified the road class, while all other methods misclassified roads as buildings. The F1-score of the RNN with the LSTM and GRU methods was above 0.85 for the vegetation class, and the corresponding DRRNN methods improved to 0.95 (Figure 10).

Urban and rural residential areas in the GF-2 classification results were most likely to be confused with one other. The irrigated land was also confused with rural residential areas and paddy fields. A severe misclassification appears for all the traditional CNN methods on the left side of the image, where rural residential area is interspersed among urban residential area which is bisected by a traffic road. Only two DRRNN methods show less misclassification in this area (the circle in Figures 11h and 11j) by considering spatial association of the information classes. The noise in the paddy field and irrigated land classes are decreased in the DRRNN methods. The average OA does not show much difference between the
traditional CNN methods, ranging from 82% to 84%. The OA reaches 89% with the two DRRNN methods, with great improvements in the F1-score for nearly all classes (Figure 12).

In the WV-3 image, the confusion matrix in Figure 13 shows a misclassification of all classes. The LeNet, FCN and two RNN results show the roads were misclassified as buildings. The DRRNN methods corrected most of the misclassification along the edges of the road, leading to an increase in PA. Miscellaneous manmade structures appear in thin linear shapes, which are easy to interrupt, causing the lowest accuracy in this class. Compared with the RNN methods, the DRRNN method filtered more noise (upper circle in Figure 13) and grouped separated patches (lower circle in Figure 13) of miscellaneous manmade structures by addressing the spatial association, resulting in an increase in both the UA and PA. Crops were likely to be
confused with trees since there was only a small patch of crops in the image. The UA of crops showed significant improvement after the two DRRNN methods corrected the misclassified trees (Figure 14).

### 3.4 Relearning process

It is relevant to check how the relearning process affects the classification results. The GRU result and the intermediate results using the DRRNN with the GRU method are shown in Figure 15 for the IKONOS, ROSIS and Landsat images. The number of relearning iterations using the DRRNN method was nine for the IKONOS image and four for the ROSIS and Landsat images. Three intermediate results are displayed.

Generally, more iterations lead to a smoother result due to the impact of the spatial autocorrelation in the CCF. The noise of the road class diminishes, and the pattern of the building class tends to become increasingly connected with more iterations in the IKONOS image. In the ROSIS image, the bitumen class is completely misclassified in the RNN with the GRU result, yet it is classified correctly in some areas after one iteration by introducing the CCF. Subsequently, the CCF plays a key role in continuously correcting this class. The gravel, self-blocking bricks and meadows display a smoothing effect with more iterations. As the majority land cover type in the Landsat image, the farmland class also tends to be connected in a continuous pattern with more iterations. The

---

**Figure 8.** Confusion matrix and per-class accuracy of the ROSIS classification results.
curvilinear pattern of the road and the river classes does not appear to be over-smoothed after four iterations. This test demonstrates that the relearning process plays an important role in decreasing noise, while some curvilinear details can be kept in the relearning process.

In the DRRNN method, the validation accuracy was used to control the number of relearning iterations. Changes in the testing accuracy against the number of iterations when this limitation was lifted were also explored (Figure 16). Forty iterations were tested for the DRRNN with the LSTM and GRU methods for the
IKONOS image. Both methods presented an upward trend of accuracy with more iterations and a decline after the peak. The accuracy tends to be stable after 30 iterations, indicating that with enough iterations, the method is able to converge.

### 3.5 Image size

The image size was set to 17 in the experiment. To analyze the impacts of different image sizes on the final result, a sensitivity analysis was carried out on the IKONOS image. The image size varied from 5 to 33 to calculate the overall accuracy for all classification methods (Figure 17). The accuracy of the LeNet, FCN and MSNN methods fluctuated with changes in image size, with sizes of 17 to 21 achieving the highest scores. The accuracy of the LSTM and GRU methods decreased when the image size was larger than 17, while the accuracy of the corresponding DRRNN methods was not greatly affected, proving the robustness of the DRRNN methods. The remaining methods did not show an obvious preference regarding image size. To obtain a fair comparison for most of the methods, we set the image size to 17. The large-sized sample images included complicated environments labeled by a single class and caused a decline

![Figure 10. Confusion matrix and per-class accuracy of the Landsat classification results.](image-url)
in accuracy. The results of the tests with the other four images showed similar trends and thus are not displayed.

### 3.6 Feature separability

To demonstrate the advantage of the developed DRRNN method, the feature separability was analyzed. For all methods, high-dimensional features were reduced by applying the $t$-distributed stochastic neighbor embedding ($t$-SNE) tool (van der Maaten and Hinton 2008) for optimal visualization. We excluded methods with poor performance, retaining only ResNet, SegNet, WCRN, two RNN and two DRRNN methods (for the first iteration). The $t$-SNE plots of these methods for three images (IKONOS, ROSIS and Landsat) are shown in Figure 18.
For all three image scenes, the two DRRNN methods have better feature separability than the other methods. Grass and trees in the IKONOS image, as well as farmland and vegetation in the Landsat image, are mixed with all other methods; only the DRRNN methods indicate a clear distinction. The DRRNN methods also have greater separability in buildings and roads for these two scenes. In the ROSIS image, the two DRRNN methods are able to separate bare soil and meadows, while all the other methods fail to do so. The asphalt and bitumen classes are also severely mixed with other classes in all benchmark methods. During the first iteration using the DRRNN methods, the asphalt and bitumen classes were separated from the other classes, and these two classes also tended to move away from each other. Therefore, this study proves that the feature separability is enhanced in the DRRNN methods by introducing the CCF.
4. Discussion

The proposed DRRNN method is designed to make full use of spatial features. Numerous methods have been proposed based on machine learning to incorporate spatial features, such as a filtering-based method (Shang and Zhang 2014; Cao et al. 2017), Markov random field (MRF) (Tarabalka et al. 2010; Lu et al. 2016), geostatistics (Atkinson and Lewis 2000; Atkinson and Naser, 2010), multiple-point statistics (MPS) (Ge and Bai 2011), etc. Most of these methods involve complicated spatial modeling processes. The advancement of DL seems to allow spatial features to be automatically extracted, making the spatial feature extraction strategies used in machine learning classification methods unnecessary. However, this statement is not entirely accurate for the following two

Figure 13. Classification results based on the WV-3 image.
First, because these DL techniques perform similarly to a data-driven method, their generalization ability is weaker than the model-based methods if the samples are not enough or not representative. Second, the CNN and related methods are more suited for scene understanding (Zhang et al. 2018a; Khan et al. 2019) and object recognition (Ren, Zhu, and Xiao 2018; Zhang et al. 2021), as the underlying features and spatial contextual information are more obvious for a typical scene or object. For land cover classification, however, some atypical categories, such as barren land and paddy fields, do not have discriminative features that can be captured with the convolutional operators alone. Therefore, extra spatial features (i.e. CCF) were introduced to guide the CNN method in a more manageable way, instead of feeding the algorithm numerous samples without considering which features the CNN can learn.

**Figure 14.** Confusion matrix and per-class accuracy of the WV-3 classification results.
Some researchers realized that introducing extra features to the CNN can lead to higher accuracy (Li, Stein, and de Beurs 2020). A challenge with this is that the input image patches used in the CNN are not technologically ready to be combined with traditional features. Most studies combined different features at a concatenate layer (Zhang et al. 2020) or used a joint probability for different predictions (Zhang et al. 2018a; Li et al. 2019). With this technique, the role of the CNN is undermined because the introduced features are not actually learned by the CNN. Therefore, we introduce the CCF with a size that equals an image patch and fed it together with the image patch into the networks. Superior to previous studies, the DRRNN method does not need to assign weights to different features, and the CCF does not need to be manually modeled using summarized statistics such as histograms or PCMs (Huang et al. 2014) but can be directly utilized by the RNN. Finally, the performance of CNN is sometimes unstable, especially for minority classes. For example, the bitumen class was completely misclassified using the RNN with the GRU method in the ROSIS image. Therefore, a relearning mechanism such as DRRNN can be employed to enhance the

Figure 15. The RNN and DRRNN methods based on the GRU model.
learning ability of the CNN method and achieve a stable result.

Overall, the results from five experiments indicate that the use of the CCF can improve classification based on the CNN model alone, as predicted. Several issues can be further explored to improve the developed method. First, the number of nodes that consisted of the CCF is required to be the same with the number of pixels included in the image patch and may lead to an over-fitting problem if the image size is too large. The analysis of the size of the image patch in Figure 17 reflects that the accuracy decreases when the size is larger than 25 for the DRRNN methods, which is partly caused by the over-fitting problem of the CCF. Therefore, a potential improvement in the DRRNN method is to define a different effective range of the CCF without relying on the input image size. Second, the relearning process is currently controlled by the increase in validation accuracy. For each iteration, the input features and parameters of the method remain the same. By introducing an intelligent updating mechanism, the input features and parameters can be adjusted automatically based on the output result of each iteration. Third, the RNN with the LSTM and GRU methods do not have an advantage over the other methods (Table 2). Therefore, the CCF does not need to be learned from an RNN classification result, but instead can be transferred to other types of CNN results. Finally, since each relearning iteration included classification errors, an alternative is to use class probabilities to represent spatial features.
5. Conclusion

A deep relearning method based on an RNN model (DRRNN) was developed for land cover classification. The class correlated feature (CCF) containing the spatial association of the pixels’ information classes was derived from the classification result and combined with the CNN-derived spectral-spatial features to iteratively improve classification. The proposed method was tested on five different types of remote sensing images with distinct environments, and the results indicated that the classification capability was improved by introducing the CCF. Compared with...
other advanced CNN models, the DRRNN method consistently achieved the best performance. The isolated noise was filtered by considering spatial auto-correlation, and the curvilinear details were well retained in the relearning iteration. Additionally, certain misclassified patches were corrected by addressing spatial arrangements that were included in the CCF. The DRRNN method demonstrated that significant improvements can be achieved for easily confused classes such as industrial land, urban residential and rural residential areas. The interpretability of DL was enhanced by using extra spatial features to guide the learning process.

Acknowledgements

The authors thank anonymous reviewers for providing helpful comments, and Avery Zhao of Plano West and Matthew Drouillard of UT Dallas for their language editorial work.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Key R&D Program of China [2021YFB3900503], the National Science Foundation (NSF) [BCS-1826839], and the National Natural Science Foundation of China [41972308, 42071312, 42171291].

ORCID

Yunwei Tang  http://orcid.org/0000-0002-2257-8749

Data availability statement

The data that supports the findings of this study are available from the corresponding author [F. Qiu], upon reasonable request. ROSIS data an be accessed from https://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes. GID data can be accessed from https://xk-ytong.github.io/project/GID.html. Dstl an be accessed from https://www.kaggle.com/competitions/dstl-satellite-imagery-feature-detection/data. The other data are not publicly available due to their containing information that could compromise the privacy of research participants.

References

Atkinson, P. M., and P. Lewis. 2000. “Geostatistical Classification for Remote Sensing: An Introduction.” Computers & Geosciences 26 (4): 361–371. doi:10.1016/S0098-3004(99)00117-X.

Atkinson, P. M., and D. K. Naser. 2010. “A Geostatistically Weighted K -NN Classifier for Remotely Sensed Imagery.” Geographical Analysis 42 (2): 204–225. doi:10.1111/j.1538-4632.2010.00790.x.

Badrinarayanan, V., A. Kendall, and R. Cipolla. 2017. “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation.” IEEE Transactions on Geoscience and Remote Sensing 39 (12): 2481–2495.

Balagué, A., L. A. Ruiz, T. Hermosilla, and J. A. Recio. 2010. “Definition of A Comprehensive Set of Texture Semivariogram Features and Their Evaluation for Object-Oriented Image Classification.” Computers & Geosciences 36 (2): 231–240. doi:10.1016/j.cageo.2009.05.003.

Bengio, Y., P. Simard, and P. Frasconi. 1994. “Learning Long-Term Dependencies with Gradient Descent Is Difficult.” IEEE Transactions on Neural Networks 5 (2): 157–166. doi:10.1109/72.2279181.

Cao, X., B. Ji, Y. Ji, L. Wang, and L. Jiao. 2017. “Hyperspectral Image Classification Based on Filtering: A Comparative Study.” Journal of Applied Remote Sensing 11 (3): 035007. doi:10.1117/1.JRS.11.035007.

Cho, K., B. van Merrienboer, Ç. Gülçehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. 2014. “Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation.” Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 1724–1734.

Ding, Y., Y. Chong, S. Pan, Y. Wang, and C. Nie. 2021. “Spatial-Spectral Unified Adaptive Probability Graph Convolutional Networks for Hyperspectral Image Classification.” IEEE Transactions on Neural Networks and Learning Systems, 1-15. doi:10.1109/TNNLS.2021.3112268.

Feng, Q., J. Yang, Y. Liu, C. Ou, D. Zhu, B. Niu, J. Liu, and B. Li. 2020. “Multi-Temporal Unmanned Aerial Vehicle Remote Sensing for Vegetable Mapping Using an Attention-Based Recurrent Convolutional Neural Network.” Remote Sensing 12 (10): 1668. doi:10.3390/rs12101668.

Gao, H., J. Guo, P. Guo, and X. Chen. 2021. “Classification of Very-High-Spatial-Resolution Aerial Images Based on Multiscale Features with Limited Semantic Information.” Remote Sensing 13 (3): 364. doi:10.3390/rs13030364.

Ge, Y., and H. Bai. 2011. “Multiple-Point Simulation-Based Method for Extraction of Objects with Spatial Structure from Remotely Sensed Imagery.” International Journal of Remote Sensing 32 (8): 2311–2335. doi:10.1080/01431160103698278.

Guo, H., J. Liu, Z. Xiao, and L. Xiao. 2020. “Deep CNN-Based Hyperspectral Image Classification Using Discriminative Multiple Spatial-Spectral Feature Fusion.” Remote Sensing Letters 11 (9): 827–836. doi:10.1080/2150704X.2020.1779374.

Hang, R., Q. Liu, D. Hong, and P. Ghamisi. 2019. “Cascaded Recurrent Neural Networks for Hyperspectral Image
Classification.” IEEE Transactions on Geoscience and Remote Sensing 57 (8): 5384–5394. doi:10.1109/TGRS.2019.2899129.
Haralick, R. M., K. Shanmugam, and I. Dinstein. 1973. “Textural Features for Image Classification.” IEEE Transactions on Systems, Man, and Cybernetics SMC-3 (6): 610–621. doi:10.1109/TSMC.1973.4309314.
He, K., X. Zhang, S. Ren, and J. Sun. 2016. “Deep Residual Learning for Image Recognition.” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 770–778. doi:10.1109/CVPR.2016.90.
Hochreiter, S., and J. Schmidhuber. 1997. “Long Short-Term Memory.” Neural Computation 9 (8): 1735–1780. doi:10.1162/neco.1997.9.8.1735.
Huang, X., Q. Lu, L. Zhang, and A. Plaza. 2014. “New Postprocessing Methods for Remote Sensing Image Classification: A Systematic Study.” IEEE Transactions on Geoscience and Remote Sensing 52 (11): 7140–7159. doi:10.1109/TGRS.2014.2308192.
Ienco, D., R. Gaetano, C. Dupaquier, and P. Maurel. 2017. “Land Cover Classification via Multitemporal Spatial Data by Deep Recurrent Neural Networks.” IEEE Geoscience and Remote Sensing Letters 14 (10): 1685–1689. doi:10.1109/LGRS.2017.2728698.
Khan, N., U. Chaudhuri, B. Banerjee, and S. Chaudhuri. 2019. “Graph Convolutional Network for Multi-Label VHR Remote Sensing Scene Recognition.” Neurocomputing 357: 36–46. doi:10.1016/j.neucom.2019.05.024.
Khusni, U., H. I. Dewangkoro, and A. M. Arymurthy. 2020. Urban Area Change Detection with Combining CNN and RNN from Sentinel-2 Multispectral Remote Sensing Data. 2020 3rd International Conference on Computer and Informatics Engineering (IC2IE), Yogyakarta, Indonesia, 171–175. doi: 10.1109/IC2IE50715.2020.9274617
LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner. 1998. “Gradient-Based Learning Applied to Document Recognition.” Proceedings of the IEEE 86 (11): 2278–2324. doi:10.1109/5.726791.
Li, M., A. Stein, and K. M. de Beurs. 2020. “A Bayesian Characterization of Urban Land Use Configurations from VHR Remote Sensing Images.” International Journal of Applied Earth Observation and Geoinformation 92: 102175. doi:10.1016/j.jag.2020.102175.
Li, H., S. Zhang, S. Zhang, and P. M. Atkinson. 2019. “A Hybrid OSVM-OCNN Method for Crop Classification from Fine Spatial Resolution Remotely Sensed Imagery.” Remote Sensing 11 (20): 2370. doi:10.3390/rs11202370.
Liang, J., Y. Deng, and D. Zeng. 2020. “A Deep Neural Network Combined CNN and GCN for Remote Sensing Scene Classification.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13: 4325–4338. doi:10.1109/JSTARS.2020.3011333.
Liu, T., A. Abd-Elrahman, J. Morton, and V. L. Wihelm. 2018c. “Comparing Fully Convolutional Networks, Random Forest, Support Vector Machine, and Patch-Based Deep Convolutional Neural Networks for Object-Based Wetland Mapping Using Images from Small Unmanned Aircraft System.” GIScience & Remote Sensing 55 (2): 243–264. doi:10.1080/15481603.2018.1426091.
Liu, S., H. Luo, Y. Tu, Z. He, and J. Li. 2018b. “Wide Contextual Residual Network with Active Learning for Remote Sensing Image Classification.” IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 7145–7148. doi:10.1109/IGARSS.2018.8517855.
Liu, B., X. Yu, A. Yu, P. Zhang, and G. Wan. 2018a. “Spectral-Spatial Classification of Hyperspectral Imagery Based on Recurrent Neural Networks.” Remote Sensing Letters 9 (12): 1118–1127. doi:10.1080/2150704X.2018.1511933.
Liu, Q., F. Zhou, R. Hang, and X. Yuan. 2017. “Bidirectional-Convolutional LSTM Based Spectral-Spatial Feature Learning for Hyperspectral Image Classification.” Remote Sensing 9 (12): 1330. doi:10.3390/rs9121330.
Lu, Q., J. Huang, L. Zhang, and L. Zhang. 2016. “A Novel MRF-Based Multi-feature Fusion for Classification of Remote Sensing Images.” IEEE Geoscience and Remote Sensing Letters 13 (4): 515–519. doi:10.1109/LGRS.2016.2521418.
Luo, H. 2018. “Shorten Spatial-Spectral RNN with Parallel-GRU for Hyperspectral Image Classification.” ArXiv. https://arxiv.org/abs/1810.12563.
Ma, A., A. M. Fillippi, Z. Wang, Z. Yin, D. Huo, X. Li, and B. Güneral. 2021. “Fast Sequential Feature Extraction for Recurrent Neural Network-Based Hyperspectral Image Classification.” IEEE Transactions on Geoscience and Remote Sensing 59 (7): 5920–5937. doi:10.1109/TGRS.2020.3018449.
Mou, L., P. Ghamisi, and X. Z. Zhu. 2017. “Deep Recurrent Neural Networks for Hyperspectral Image Classification.” IEEE Transactions on Geoscience and Remote Sensing 55 (7): 3639–3655. doi:10.1109/TGRS.2016.2636241.
Ndikumana, E., D. Ho Tong Minh, N. Baghdadi, D. Courault, and L. Hossard. 2018. “Deep Recurrent Neural Network for Agricultural Classification Using Multitemporal SAR Sentinel-1 for Camargue, France.” Remote Sensing 10 (8): 1217. doi:10.3390/rs10081217.
Ren, Y., C. Zhu, and S. Xiao. 2018. “Small Object Detection in Optical Remote Sensing Images via Modified Faster R-CNN.” Applied Sciences 8 (5): 813. doi:10.3390/app8050813.
Shang, K., and X. Zhang. 2014. “A Novel Spectral Spatial Filtering Approach for Hyperspectral Image Classification.” 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Lausanne, Switzerland, 1–4.
Shi, X., Z. Chen, H. Wang, D. –, Y. Yeung, W. –, K. Wong, and W. –. C. Woo 2015. “Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting.” Proceedings of the 28th International Conference on Neural Information Processing Systems, Montreal, Canada, 1: 802–810.
Shi, C., and C.-M. Pun. 2018. “Multi-Scale Hierarchical Recurrent Neural Networks for Hyperspectral Image Classification.” Neurocomputing 294: 82–93. doi:10.1016/j.neucom.2018.03.012.
Sun, Z., L. Di, and H. Fang. 2019. “Using Long Short-Term Memory Recurrent Neural Network in Land Cover Classification on Landsat and Cropland Data Layer Time Series.” International Journal of Remote Sensing 40 (2): 593–614. doi:10.1080/01431161.2018.1516313.

Tang, Y., L. Jing, H. Li, and P. M. Atkinson. 2016. “A Multiple-Point Spatially Weighted K-NN Method for Object-Based Classification.” International Journal of Applied Earth Observation and Geoinformation 52: 263–274. doi:10.1016/j.ijag.2016.06.017.

Tang, Y., L. Jing, F. Shi, X. Li, and F. Qiu. 2019. “A Hybrid Model Integrating Spatial Pattern, Spatial Correlation, and Edge Information for Image Classification.” Remote Sensing 11 (13): 1599. doi:10.3390/rs11131599.

Tang, Y., F. Qiu, L. Jing, F. Shi, and X. Li. 2021. “A Recurrent Curve Matching Classification Method Integrating Within-Object Spectral Variability and Between-Object Spatial Association.” International Journal of Applied Earth Observation and Geoinformation 101: 102367. doi:10.1016/j.ijag.2021.102367.

Tarabalka, Y., M. Fauvel, J. Chanussot, and J. A. Benediktsson. 2010. “SVM- and MRF-Based Method for Accurate Classification of Hyperspectral Images.” IEEE Geoscience and Remote Sensing Letters 7 (4): 736–740. doi:10.1109/LGRS.2010.2047711.

Tong, X.-Y., G.-S. Xia, Q. Lu, H. Shen, S. Li, S. You, and L. Zhang. 2020. “Land-cover Classification with high-resolution Remote Sensing Images Using Transferable Deep Models.” Remote Sensing of Environment 237: 111322. doi:10.1016/j.rse.2020.111322.

van der Maaten, L. J. P., and G. E. Hinton. 2008. “Visualizing High-Dimensional Data Using T-SNE.” Journal of Machine Learning Research 9: 2579–2605.

Wu, H., and S. Prasad. 2017. “Convolutional Recurrent Neural Networks for Hyperspectral Data Classification.” Remote Sensing 9 (3): 298. doi:10.3390/rs9030298.

Xiao, B., J. Liu, J. Jiao, Y. Li, X. Liu, and W. Zhu. 2022. “Modeling Dynamic Land Use Changes in the Eastern Portion of the Hexi Corridor, China by CNN-GRU Hybrid Model.” GiScience & Remote Sensing 9 (1): 501–519. doi:10.1080/15481603.2022.2037888.

Yeung, H. W. F., M. Zhou, Y. Y. Chung, G. Moule, W. Thompson, W. Ouyang, W. Cai, and M. Bennamoun. 2022. “Deep-learning-based Solution for Data Deficient Satellite Image Segmentation.” Expert Systems with Applications 191: 116210. doi:10.1016/j.eswa.2021.116210.

Zhang, C., X. Pan, H. Li, A. Gardiner, I. Sargent, J. Hare, and P. M. Atkinson. 2018a. “A Hybrid MLP-CNN Classifier for Very Fine Resolution Remotely Sensed Image Classification.” ISPRS Journal of Photogrammetry and Remote Sensing 140: 133–144. doi:10.1016/j.isprsjprs.2017.07.014.

Zhang, X., Y. Sun, K. Jiang, C. Li, L. Jiao, and H. Zhou. 2018b. “Spatial Sequential Recurrent Neural Network for Hyperspectral Image Classification.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11 (11): 4141–4155. doi:10.1109/JSTARS.2018.2844873.

Zhang, N., X. Wei, H. Chen, and W. Liu. 2021. “FPGA Implementation for CNN-Based Optical Remote Sensing Object Detection.” Electronics 10 (3): 282. doi:10.3390/electronics10030282.

Zhang, C., P. Yue, D. Tapete, B. Shangguan, M. Wang, and Z. Wu. 2020. “A Multi-Level Context-Guided Classification Method with Object-Based Convolutional Neural Network for Land Cover Classification Using Very High Resolution Remote Sensing Images.” International Journal of Applied Earth Observation and Geoinformation 88: 102086. doi:10.1016/j.ijag.2020.102086.

Zhu, Y., C. Geiß, E. So, and Y. Ji. 2021. “Multitemporal Relearning with Convolutional LSTM Models for Land Use Classification.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14: 3251–3265. doi:10.1109/JSTARS.2021.3055784.