



Cognitive digital modelling for hyperspectral image classification using transfer learning model

Mohammad SHABAZ^{1,*}, Mukesh SONI²

¹Model Institute of Engineering and Technology, Jammu, J&K, 181123, India

²Department of CSE, University Centre for Research & Development Chandigarh University, Mohali, Punjab-140413, India

Received: 02.06.2023

Accepted/Published Online: 17.06.2023

Final Version: 27.10.2023

Abstract: Deep convolutional neural networks can fully use the intrinsic relationship between features and improve the separability of hyperspectral images, which has received extensive in recent years. However, the need for a large number of labelled samples to train deep network models limits the application of such methods. The idea of transfer learning is introduced into remote sensing image classification to reduce the need for the number of labelled samples. In particular, the situation in which each class in the target picture only has one labelled sample is investigated. In the target domain, the number of training samples is enlarged by the homogenous region obtained by segmenting the target image. On this basis, the deep Siamese convolutional neural network is used to reduce the distribution difference between the source domain image and the target domain image to achieve the final result of the target hyperspectral image classification. The experimental results show that the combination of homogenous region and Siamese convolutional network can improve the classification effect of semisupervised transfer learning and better solve cross-regional hyperspectral image classification.

Key words: Deep learning, neural network, transfer learning, classification, hyperspectral image, image classification

1. Introduction

Hyperspectral imaging (HSI) is an important remote sensing technique that collects the electromagnetic spectrum from visible to near-infrared wavelengths. Due to the rich spectral information, hyperspectral remote sensing images can distinguish subtle spectral differences and have been widely used in many fields. The key to the successful application of hyperspectral imaging is accurate classification. Statistical results of published papers show that HSI classification (i.e., assigning each pixel to a specific class based on its spectral characteristics) is one of the most active areas of hyperspectral research. In HSI classification tasks, the limited available training samples are always a major challenge. It is quite challenging to label a lot of straining samples in time, and the basic idea of transfer learning becomes a potential hope to meet this challenge. As far as remote sensing image classification is concerned, the purpose of transfer learning is to use the abundant labeled samples in similar images to improve the classification accuracy of new images that lack labeled samples. The image with a large number of labeled samples is called the source domain, and the image to be classified with a lack of labeled samples is called the target domain. Strictly speaking, the task of using similar images to classify new images belongs to a branch of transfer learning - domain adaptation. The source domain image and the target domain image have certain similarities and differences at the same time, which is the reason for the existence

*Correspondence: bhatsab4@gmail.com

of domain adaptive methods.[1] suggested a deep network cascade for image SR that included a cooperative local autoencoder and nonlocal self-similarity search. When the domain information can be represented by a standard sparse coding model, a deep network paired with a sparse prior was described in [2] to increase efficiency. The model training speed is increased as a result of the sparse representation. Additionally, the deep spectral difference convolutional neural network (CNN) based SR model was created without distorting spectral information [3]. Additionally suggested SR techniques include those used by CNNs [4], generative adversarial networks [5], and linked deep autoencoders [6]. However, since there are so few training examples available in the hyperspectral picture domain, training these SR networks is highly challenging. The key to achieving the goal of transfer learning is to reduce the difference between the source and target domain images in the feature space. This task is closely related to the features used for classification. The earliest hyperspectral classifications were mainly limited to the use of spectral information. With the increasing awareness of the importance of spatial information, hyperspectral image classification begins to utilize both spectral and spatial information. In recent years, deep learning technology has developed rapidly, and the deep learning-based classification of hyperspectral images has produced encouraging results. Therefore, realizing the information transfer from the source domain to the target domain on the basis of deep learning has become the common goal of many researchers. In fact, with the introduction of convolutional neural network (CNN) [7], the corresponding transfer learning idea soon appeared: using a modest number of samples from the data set (target domain) that needs to be classified to fine-tune in huge data.

- This study combines Mean Shift and SEEDS, taking into account their individual benefits and drawbacks, to achieve the segmentation of the target image's various scales.
- It then removes up the segmentation output to provide high-quality homogenous sections.
- Use the mean shift technique to first segment the picture since it can determine the local gradient in pixel densities.
- Similar pixels will converge to the same maximum value if this estimate procedure is repeated, and it will also converge to the maximum value.

The classification model trained on the set (source domain) is used to predict the categories of other samples in the target domain through the fine-tuned classification model. This scheme is simple and effective, many times, it is used by researchers to classify hyperspectral data [8], [9], [10]. Data scientists handle machine learning model training by using hyperparameters, which are external configuration variables. Before training a model, the hyperparameters also known as model hyperparameters are manually set. They differ from parameters, which are intrinsic parameters that organically emerge throughout the learning process and are not determined by data scientists. A decision tree's number of branches and the quantity of nodes and layers in a neural network are two examples of hyperparameters. Hyperparameters control important characteristics like model complexity, learning rate, and model architecture. There is also a view that this simple fine-tuning cannot reduce the large data shift (Data Shift) between the source and the target domain. Therefore, the maximum mean difference (MMD) technique in domain adaptation is introduced to solve the problem [11], [12]. MMD is based on reducing the difference in feature distribution between unlabeled samples. The selection and training samples require a certain amount of experience, and a large number of target domain samples are needed to ensure the results are valid. Reference [13] utilizes a few target domain labeled samples to overcome the

shortcomings of unsupervised domain adaptation. This method uses the classic convolutional Siamese network to cleverly design a deep domain adaptive network structure, and uses the source domain samples and limited number of labeled samples in the target domain to train it, reducing the source domain. The distance between similar samples in the target domain and the distance between different samples are increased at the same time to achieve semantic alignment of features between the two domains. The loss function used in this method is CCSA (Classification and Contrastive Semantic Alignment).

One-shot, the situation when there is just one labelled sample for each class in the target domain, is an extreme case of supervised transfer and has received much attention recently [14],[15],[16]. Experiments have proved that CCSA can also achieve good results under one-shot conditions, and the classification accuracy will be significantly improved for each additional labeled sample in the target domain. Moreover, it is inferred that if one actual labeled sample for each class in the target domain is effectively augmented, then the effect of transfer learning can be improved. Reduce the disparity in the feature space between the pictures from the source and target domains in order to successfully implement transfer learning. The characteristics that are employed for categorization have a direct bearing on this task. Hyperspectral image classification starts to use both spectral and spatial information as people become more aware of the value of spatial information. When applied to remote sensing photos, many transfer learning techniques now in use frequently have little to no noticeable impact, however the Siamese network method may somewhat enhance the classification effect of nearly all experimental images. In light of this network topology, increasing the number of target domain training samples and altering the training sample pairing metrics would increase the classification impact of transfer learning.

This paper focuses on sample augmentation for the one-shot case to explore the highest classification accuracy that can be obtained with 1 true labeled sample for each class in the target domain. Sample augmentation is a concept that appears when training a CNN scene classification model. It increases the number of training samples by transforming existing labeled images in the form of rotation, deformation, and translation. The traditional method cannot be used to process the spectral feature data of pixel classification. Therefore, an amplification method based on homogeneous regions is proposed in this paper. The sample amplification effect of the new method is closely related to the image segmentation method for obtaining homogeneous regions. In the field of image processing, image segmentation is an important classic topic, and several image segmentation algorithms have been presented [17],[18],[19]. In this paper, the process of generating the homogeneous area involves image segmentation on two areas with different characteristics, one is the original remote sensing image itself; the other is the irregular area near the edge of the object. The first type of area is suitable for using ordinary image segmentation algorithms; the second type of area is suitable for using super pixel segmentation method that can form small patches. Because there have only been a few literature reviews, and none of them have provided any crucial insights on this emerging domain, the application of cognitive digital modelling for hyperspectral image classification through transfer learning model in healthcare is still in its preliminary stages. The reason for this is because cognitive digital modelling is used for hyperspectral image classification through transfer learning model. The rationale for this is because there are not that many literature reviews since, stated previously, there have not been that many literature reviews written. Because of this, obtaining an understanding of the possible benefits that cognitive computing could bring to boosting the quality of hyperspectral pictures and making use of those benefits can be difficult and puzzling. Furthermore, researchers can find it challenging to monitor and employ, particularly in terms of the capabilities it possesses or the effect it has on healthcare. As a result, we conducted research to capture relevant literature from a variety of sources with

the following objectives in mind: to reveal the recent research scientific work on cognitive computing in terms of the process, methodologies, apps, and outcomes used in the medical industry for modelling and classifying the hyperspectral images; to investigate the burgeoning aspects of cognitive computing in healthcare; to reveal the new vision of cognitive computing with respect to the healthcare industry.

In this paper, the Mean shift algorithm [17] is used in the first type of region, because of its good stability and robustness, simple principle and less parameters to be set. Amplification of samples requires high purity of the homogenous region plaques. It is generally considered that the patches generated by the Mean shift algorithm are not compact enough, contain less semantic information, and contain shadow areas that are prone to under-segmentation. However, the compactness and semantic information do not affect the amplification of the sample, and the separation of the shadow area from the illumination can avoid the over-segmentation phenomenon that hinders the sample amplification. A comparative study [20] shows that the SEEDS algorithm can maintain the main edges in the image, with high operating efficiency, and the irregularity of its patches does not affect the expansion of training samples. Therefore, this paper chooses the SEEDS algorithm to perform super pixel segmentation on the second type of regions. After extensive analysis and comparison, it is found that CCSA has the characteristics of simple principle, clear network structure, easy training and good effect. In this study, a sample amplification scheme based on a homogenous region was proposed to improve CCSA, in order to improve the classification accuracy when the target domain only contains one real marker point per class as shown in Table 1.

Table 1. Classification accuracies obtained by different methods.

Class	2-D CNN [21]	3-D CNN [22][23]	SUNN [24]	SSRN [25]	S3Net [26]
Alfalfa	100 \pm 0.0	98.55 \pm 2.37	100 \pm 0.0	89.36 \pm 9.96	100 \pm 0.0
Corn	89.08 \pm 8.44	93.86 \pm 4.44	96.74 \pm 2.72	79.39 \pm 11.93	99.49 \pm 1.08
Grass/Tree	96.62 \pm 3.55	93.87 \pm 4.14	90.15 \pm 5.21	96.92 \pm 2.18	96.63 \pm 2.69
Oat	100 \pm 0.0	98.75 \pm 4.80	100 \pm 0.0	43.56 \pm 19.22	100 \pm 0.0
Wheet	99.57 \pm 0.99	98.20 \pm 3.02	99.08 \pm 1.35	90.60 \pm 6.11	99.14 \pm 1.61

The goal of transfer learning is to make better use of the information gained from the images of the source domain by applying it to the images of the target domain. If we were to start from zero while training the model, it would be at a disadvantage compared to the pretrained model, which already has knowledge of the facets that make up the domain. The presence or absence of labelled data in the source domain determines which of two further subcategories are included in this further division. These include learning new skills while juggling multiple responsibilities at once and teaching oneself new skills. The transductive transfer learning approach is used in situations when the domains of the source image and the target image are not exactly the same but are connected to one another. It is possible to identify parallels between the source task and the target task. In most of these situations, the source domain contains a substantial amount of labelled data, but the target domain is composed entirely of unlabeled information.

The hybridization of the homogeneous region and the Siamese convolutional neural network helps in solving the cross-region hyperspectral image classification by making use of the homogeneous region to amplify the one sample that is taken from the target domain and using the Siamese convolutional neural network to reduce the gap in time and space that exists between the data coming from the source domain and the data

coming from the target domain. This helps in helping in helping in helping in helping in helping. Because of this, the strategy is able to achieve the highest possible levels of success across all three categorization assessment indicators. The overall accuracy and the average accuracy both come in at 82.35%, which is a difference of approximately 2% to 3% in comparison to the CCSA method. In addition to that, there has been an increase of 0.04% in the Kappa coefficient.

This study investigates sample augmentation for the one-shot case in order to investigate the best classification accuracy that can be achieved with a single true labelled sample for each class in the study domain. The purpose of this investigation is to investigate the best classification accuracy that can be achieved with a single sample. The concept of sample augmentation arises at some point during the training process of a CNN scene categorization model. It does this by applying transformations, such as rotation, deformation, and translation, on images that have already been labelled. This increases the total number of training instances. Using the traditional method is not an option for processing the spectral feature data that is associated with pixel classification. An approach to amplification that is predicated on the existence of homogenous zones is proposed in this research. The sample amplification effect of the cutting-edge method is intricately connected to the method of image segmentation that is applied in order to generate uniform zones. Plaques with a homogenous area and high purity are required for the sample amplification process. In general, it is thought that the patches produced by the Meanshift method are overly big, lack semantic detail, and feature shadow areas that are prone to under-segmentation. The problem of over-segmentation, which prevents the sample from being amplified, can be avoided, however, by physically separating the area of shade from the area of illumination. The compactness of the sample does not alter the sample amplification, nor does semantic information. When there is only one authentic marker point available for each class in the target domain, this research suggests using a sample amplification strategy that is based on a homogenous region to improve CCSA and increase classification accuracy. Because the majority of points are labelled in the same category as their true category, all of the points in the homogenous region will be given the main category in accordance with the simple amplification scheme. These points will be considered amplified sample points with pseudo-labels because the main category will be given to all of the points in the region. These enlarged sample points are useful for supervised training and do not pose any safety concerns. This study is brought closer to the scenario of the classic technique as a result of the homogenous region-based amplification procedures, which result in a bigger number of target domain labelled samples.

2. Homogeneous region acquisition and training sample amplification

2.1. Homogeneous region generation based on Mean Shift and SEEDS superpixel segmentation

According to the respective advantages and disadvantages of Mean Shift and SEEDS, this paper combines these two segmentation methods to achieve the segmentation of different scales of the target image, and purifies the segmentation results, so as to finally obtain high-quality homogeneous regions. The specific steps of the segmentation operation are described in detail below.

1. The hyperspectral remote sensing images have several bands, and it is not suitable to directly apply the segmentation algorithm. Firstly, the unsupervised band selection method based on the band similarity measure [27] was used to select the target image, and the three most informative but least similar bands of the target image were selected as new color features vector and normalize it.

2. Use the mean shift algorithm to initially segment the image: the mean shift technology can estimate the local density gradient of similar pixels. This estimation process is repeatedly performed, and similar pixels will converge to the same maximum value and will converge to the same maximum value. The pixels that meet the adjacent conditions are combined into the same super pixel to complete the entire segmentation process.
3. Purify the initial segmentation results: The area of the initial segmentation patches obtained by the mean shift algorithm is large, and the phenomenon of under-segmentation often occurs. At the same time, the homogeneous area will contain impurity pixels. Homogeneous patches located at the edge of the feature area span multiple categories. In this paper, a plaque purification scheme is designed with the help of the idea of mean shift, which can filter out the initial segmentation patches. The relatively pure area in the original segmentation block is removed, and the inaccurate part is removed to obtain a high-purity homogeneous area. The following homogeneous region purification algorithm gives the specific implementation steps. The number of iterations T is in the range of $[3, 7]$, and the value of q is in the range of $[0.5, 0.8]$. In this experiment, T is 5; q is 0.6.

Algorithm 1: Homogeneous region purification algorithm

Input: All pixels in the initial segmented patch of the homogenous region $P = p_{i=1}^n$

Output: P_1

Feature vector X for all pixels $X = x_{i=1}^n$ The number of iterations T , the image is retained after purification the ratio of prime elements q .

Step 1: Initialize the mean value of high-purity point set features: $\tilde{x}_c = \frac{1}{n} \sum_{i=1}^n x_i$

Step 2: $d_i = x_i - \tilde{x}_c, (i = 1 : n)$

Step 3: $\lceil n * q \rceil$ points with smaller distances form a high-purity point set P_1

Step 4: Recalculate the feature mean \tilde{x}_c of the high-purity point set

Step 5: Repeat steps 2 to 4 iteratively calculate T times

4. Perform small-scale super pixel segmentation on the remaining image parts: the above-mentioned mean-shift segmentation and purification scheme can effectively obtain high-purity homogeneous patches in hyperspectral images, but at the same time, it will produce nonhigh-purity homogeneous patches. These remaining parts are mostly near the edge of the feature or in the mixed area of multiple types of features. This type of area is complex and changeable. In this paper, the SEEDS super pixel segmentation algorithm is used to segment the purified remaining part. The area of super pixel patches is small, and noise points are rarely mixed, so each super pixel can be regarded as a purified patch. After this step, each plaque is a high-purity homogeneous region. The overall flow chart of the homogenous region acquisition in this paper is shown in Figure 1.

The segmentation technique cannot be used directly for the multiband hyperspectral remote sensing images. First, the target image is chosen using the unsupervised band selection method based on the band similarity measure. The three most insightful but least similar bands of the target image were then chosen as new colour feature vectors and normalised. The mean shift algorithm is used to first segment the image because it can determine the local gradient in pixel densities. Similar pixels will converge to the same maximum value if this estimation process is repeated, and it will also converge to the maximum value. To finish the segmentation procedure, the pixels that satisfy the nearby conditions are joined into a single superpixel. After this initial segmentation results are purified. The mean shift technique produces initial segmentation patches with a wide

surface area, and under-segmentation is a common occurrence. At the same time, impurity pixels will be present in the homogenous area. At the margin of the feature region, homogeneous patches are present that cross many categories. The concept of mean shift is used in this research to construct a plaque purification system that can screen out patches associated with initial segmentation. To create a high-purity homogeneous area, the reasonably pure area from the original segmentation block is deleted together with the erroneous portion. The aforementioned mean-shift segmentation and purification method can successfully produce nonhigh-purity homogeneous patches while also obtaining high-purity homogeneous patches in hyperspectral images. The majority of these leftover pieces are either close to the feature's edge or in a mixed area of various feature kinds. This kind of region is complex and subject to change. The cleansed remaining region is segmented in this work using the SEEDS super pixel segmentation technique. Each super pixel can be thought of as a purified patch due to the tiny size of super pixel patches and the rarity of noise points mixing. Each plaque is a high-purity, homogeneous zone after this phase.

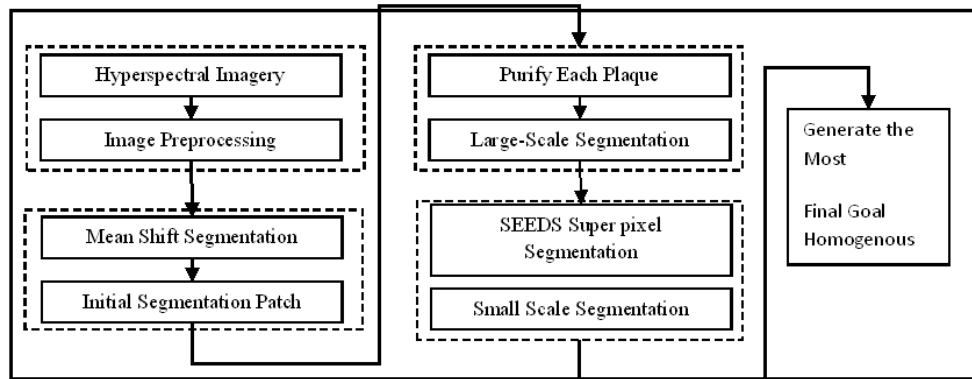


Figure 1. Flow chart of the segmentation method.

The purpose of "pooling" in a CNN is to lower the spatial size of the input image in order to bring down the total number of computations performed by the network. Downsampling is carried out through pooling, which results in a reduction in size, with the critical data being passed on to subsequent layers of the CNN. However, each super pixel is considered a pure patch since it is the only type of pixel that satisfies all of the conditions that are necessary to finish the segmentation process. It denotes that the plaque purification method is created with the assistance of the concept of the mean shift algorithm, which can filter out the initial segmentation patches. This can be accomplished by removing the patches. In order to create a high-purity homogeneous area, the relatively pure area that was contained within the initial segmentation block is eliminated, and the part that was erroneous is eliminated as well. This method has the potential to successfully acquire high-purity homogeneous patches in hyperspectral pictures, but at the same time, it will generate homogenous patches that are not of high purity. The majority of these remaining pieces are located in close proximity to the feature's border or in an area that has a mixture of other types of features. This kind of region is complicated and subject to change. In this article, the purified remaining part is segmented with the help of an algorithm called the SEEDS super pixel segmentation algorithm. Because of the limited surface area of super pixel patches and the infrequent mixing of noise points, each super pixel can be thought of as a separate, purified patch. Each plaque will be a highly pure and consistent region once this process has been completed.

2.2. Segmentation result evaluation index

In this paper, two widely used evaluation metrics are utilized to evaluate the quality of segmentation results, namely Achievable Segmentation Accuracy (ASA) [28] and Under segmentation Error (UE) [29]. The simple linear iterative clustering (SLIC) technique is used by the superpixels function. The programme divides pixels into areas with comparable values. The complexity of these procedures can be decreased by using these regions in image processing techniques like segment. For several image processing and computer vision methods, including image segmentation, picture parsing, semantic labelling, object classification, detection, and tracking, superpixels have emerged as a crucial building element. Among them, ASA represents the maximum segmentation accuracy that super pixel segmentation patches can achieve. It is a method to evaluate super pixels negatively from the final segmentation result, which can also be called purity. On the other hand, UE is a method to directly evaluate the quality of super pixel segmentation from the front. It can measure the edge fit of the super pixel block according to the proportion of the segmentation patch "overflowing" the boundary of the real area. The two evaluation indicators start from different perspectives, and can more comprehensively evaluate the quality and accuracy of the final segmentation results. The specific calculation methods of the two indicators are shown in formulas (1) and (2) respectively: Among them, G represents the real distribution

$$ASA = \frac{\sum_k \max_i |S_k \cap G_i|}{\sum_i |G_i|} \dots \dots \dots (1)$$

$$UE = \frac{\sum_i \sum_{k: S_k \cap G_i \neq \emptyset} |S_k - G_i|}{\sum_i |G_i|} \dots \dots \dots (2)$$

of ground objects, S is the super pixel block obtained by the segmentation algorithm, and "i" represents the number of all pixels contained in these super pixel blocks. There are no real markers in some super pixel blocks, and only some pixels in some super pixel blocks have real markers. The above two indicators are only for the marked samples for statistics and result evaluation.

2.3. The principle of training sample amplification based on homogenous region

The purified initial segmentation plates and super pixel patches are homogeneous regions with high purity, which can be used to expand training samples. The simplest solution for augmenting training samples with homogeneous regions is to select a point in the homogeneous region, obtain its category information, and then assign the same category to all image points in the same region. Figure 2 shows the initial segmentation patch purification. Figure 2(a) is a patch obtained by segmentation, in which "+" indicates that the main category pixels in the patch in this homogeneous area occupy most of the homogeneous area; "x" and "o" indicate different noise points that are misclassified into patches for subject classes. Figure 2(b) is the purified plaque, in which most of the noise points are removed, but there is still a small amount of residual.

The sample points with real marks selected by random means must be in a high-purity homogeneous region, such as the points surrounded by the red circle in Figure 3(a). This point belongs to the main category in the plaque. According to the simple amplification scheme, all points in the homogenous region will be assigned the main category (as shown in Figure 3(b)), and will be regarded as amplified sample points with pseudo-labels due to plaque most points are labeled in the same category as their true category, so such augmented sample points are beneficial and harmless for supervised training. In Figure 3(c), the noise points in the high-purity homogeneous region are randomly selected. According to the simple amplification principle, the amplification

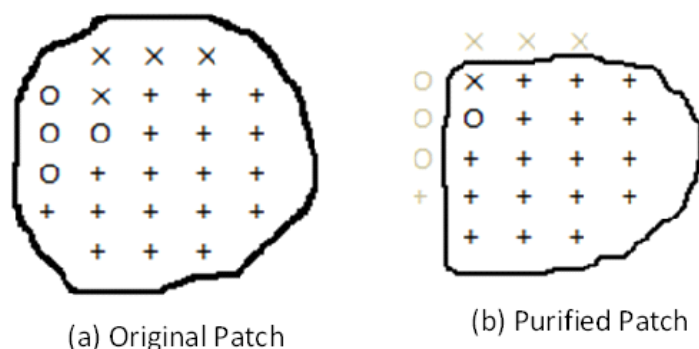


Figure 2. Segmentation patches purification.

result at this time should be as shown in Figure 3(d), and most of the image points are assigned an error label that is different from the actual situation. Pseudo-label samples like this used to train the model will undoubtedly destroy the quality of the classification model, such an augmentation is better than no augmentation.

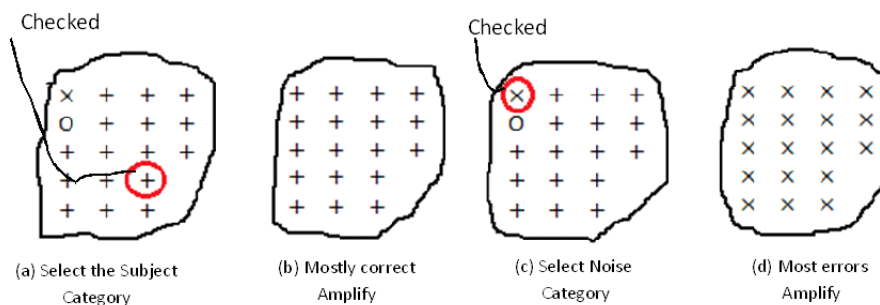


Figure 3. Simple criterion amplification effect.

Although there are very few noise points in the high-purity homogeneous area, it is still possible to select the noise points. To minimize the possibility of false amplification, a classification model is trained using actual labeled samples in the source domain and in the target domain and predicts the class of all pixels in the target domain. The prediction category with the most pixels in the homogeneous patch is called the patch prediction category. If the category of the randomly selected real sample point is consistent with the predicted category of the patch where it is located, the simple scheme amplification is performed; otherwise, the sample amplification is not performed. The situation where the selected sample point category is consistent with the patch prediction category, and finally the amplification result is given. In the high-purity homogeneous region in each of the above schematics, the actual categories of all image points are drawn, but except for the selected marked points, this category information cannot be used, and can only be regarded as unmarked points. The purpose of drawing the actual classes is for the convenience of understanding the principles of the algorithm.

3. Semisupervised migration based on homogeneous regions and Siamese convolutional neural networks mobile learning

3.1. Construction of Siamese convolutional neural network model

The network structure with dual branches and weight sharing is called Siamese neural network (SNN) [30]. When the branch of the Siamese neural network is a convolutional neural network, it is called a Siamese convolutional neural network (SCNN). Compared with the traditional convolutional neural network, the biggest feature of the SCNN is that the original single series network is changed to a new network with two parallel subbranches with the same structure and shared weights, as shown in Figure 4. X_1 and X_2 are the input data of

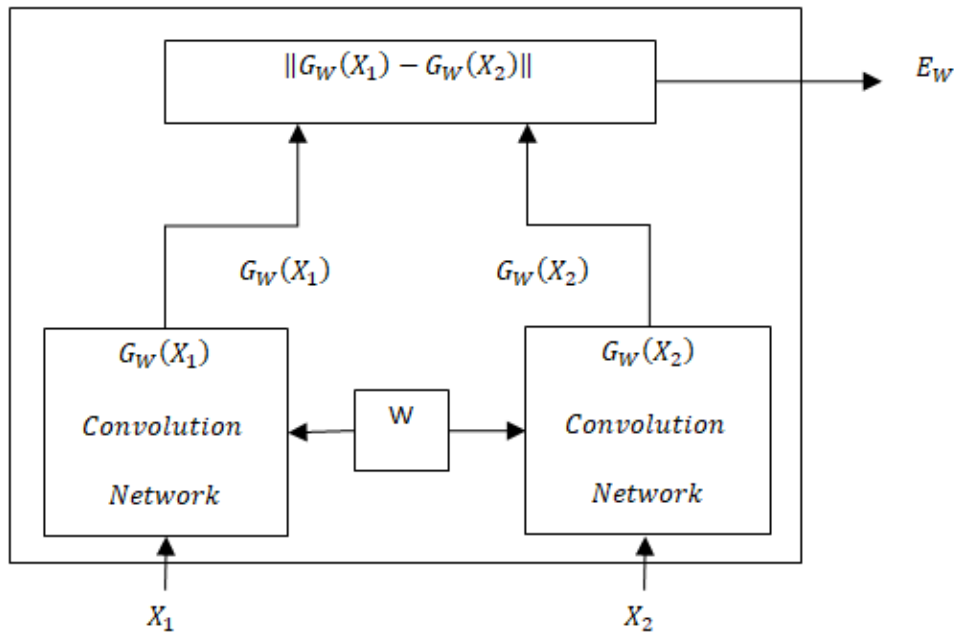


Figure 4. Siamese convolutional neural network diagram[31].

the network model, also known as paired data. $G_W(X_1)$ and $G_W(X_2)$ represent the mapping functions of X_1 and X_2 , respectively, and $E_W = G_W(X_1) - G_W(X_2)$ represents the similarity measure of the two-input data after the output of the twin network, such as Euclidean distance. The network generally uses the contrast loss function as the loss function in its training process [32], so that the original similar samples can maintain their original characteristics in the new target feature space, and the samples with larger differences can be separated farther. This feature is suitable for feature dimensionality reduction, but it is also suitable for transfer learning after improvement. PU classification problem is one where just a portion of the training set's data is labelled as positive but the other data is left unlabeled and might be either positive or negative. Datasets are groups of textual or quantitative data that are organised in tables. They are commonly used by computer scientists to test new software programmes and by statisticians to practise analytic techniques. Models based on SCNN require fewer training samples because the input data to the network are homogeneous or heterogeneous sample pairs, and a small number of labeled samples can be combined into a large number of training sample pairs. For a SCNN, there are four different sorts of layers: Convolutional, pooling, ReLU correction, and completely

connected. Because the input data to the network are homogeneous or heterogeneous sample pairs and a small number of labelled samples may be merged into a large number of training sample pairs, models based on SCNN require fewer training samples. The embedding function g is modelled by a CNN in the SCNN model architecture, which primarily comprises of an initial convolutional layer and a fully connected layer.

3.2. Network model construction

For the purpose of transfer learning, the SCNN model architecture as shown in Figure 5 is used. The embedding function g is modeled by a CNN, which mainly consists of an initial convolutional layer and a fully connected layer. Among them, the two branches of the Siamese network, one is used for training the source domain and the other is used for the target domain. Since $g_s = g_t = g$, the parameters of CNN will be shared in both branches. Furthermore, the source domain branch uses an additional fully connected layer to model h ,

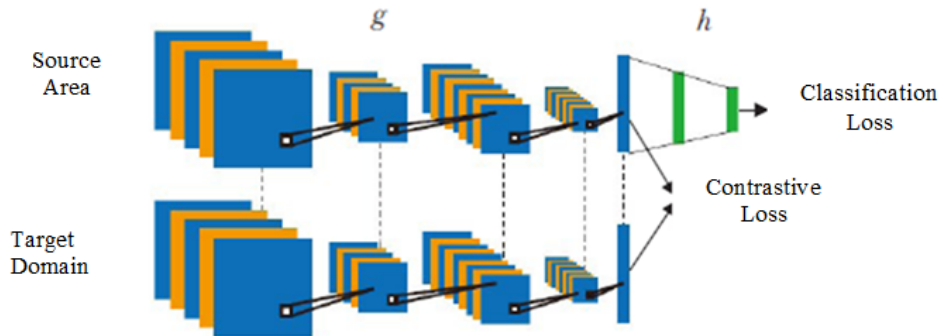


Figure 5. Proposed Siamese convolutional neural network model[33].

building a classification loss function. There are two main types of loss functions in the model: (1) classification loss function as shown in equation 3, (2) contrastive loss function as shown in equation 4. Contrastive loss function is further divided into the alignment loss function between similar samples and the separation loss function between different classes of samples as shown in equation 5. The final overall loss function as shown in equation 6 is divided into the following three parts. Where $E[.]$ represents the statistical expectation, α can be

$$L_c(f) = E[\ell(f(X^s), Y^s)] \dots \dots \dots (3)$$

$$L_s(g) = \sum_{a=1}^c d(p(g(X_a^s)), p(g(X_a^t))) \dots \dots \dots (4)$$

$$L_D(g) = \sum_{a,b|a \neq b} k(p(g(X_a^s)), p(g(X_b^t))) \dots \dots \dots (5)$$

$$L_{total}(f) = \gamma L_c(f) + (1 - \gamma)(L_s(g) + L_D(g)) \dots (6)$$

any suitable loss function (such as the cross-entropy function used in multi-classification problems), when the distributions of X_s and X_t are different, only utilize source domain data to train the resulting deep model f will cause performance degradation of the target domain. d is a distance metric (such as Euclidean distance) in the embedding space of the two domains, and once aligned, the model has the characteristic of ambiguous sample domain properties. k is the similarity measure of the distribution of X_s and X_t in the embedding space. When $p(g(X_a^s))$ and $p(g(X_b^t))$ are similar, the classification accuracy will be reduced, so k acts as a penalty. γ is

the trade-off coefficient that is used to control the respective weights of the classification loss function and the contrast loss function.

Because homogeneous or heterogeneous sample pairs are used as input to the network and because a small number of labelled samples may be merged into a large number of training sample pairs, models based on SCNN require fewer training samples. The SCNN model architecture is utilised for the aim of transfer learning. An initial convolutional layer and a fully connected layer make up the majority of a CNN, which is used to simulate the embedding function g . One of these, the source domain, is trained using the other of the two branches of the Siamese network. According to the specific situation of this study, the concrete network structure mainly consists of three parts: feature extraction, classification and recognition, and similarity calculation as illustrated in Figure 6. In the feature extraction part, the structure of the left and right branches of the SCNN is exactly the same, consisting of a series of convolutional, pooling, and fully connected layers respectively. Through the above network structure, the input source and target domain images are mapped to the corresponding feature space. The Euclidean Distance is used as the similarity index of the image output features of the two fields, and the distance between the same samples in the two fields is shortened by the alignment loss function, and the separation loss function increases the distance between samples of various classes. In order to ensure that the source domain branch can achieve accurate classification after training, an additional network layer is introduced after the source domain branch to build a classification model. The expanded training samples can be employed with the homogenous, very pure starting segmentation plates and superpixel patches. Selecting a point in the homogeneous region, obtaining its category information, and then assigning the same category is the easiest way to add homogeneous regions to training samples. In order to allow the original comparable samples to retain their original properties in the new target feature space, the network often utilises the contrast loss function as the loss function in its training process. This research conducts related experiments using two sets of hyperspectral test datasets with varying degrees of similarity (varying degrees of data offset) in order to fairly assess the efficacy of this approach and its adaptability to various data and geographies.

This transfer idea drawn on SCNN has two advantages: firstly, the principle is intuitive and easy to understand. Through paired training of samples in the source and target domain, the samples of the same type in the two domains closely resemble each other in the feature space, while the samples of different types closely resemble each other in the feature space. The distance between them is increased as much as possible, which makes it easy to classify the target domain data with the source domain information in the feature space; the second is the general practicability of this migration idea. Many existing transfer learning methods often have no obvious effect when applied to remote sensing images, while the method based on the Siamese network can improve the classification effect of almost all experimental images to a certain extent. Therefore, based on this network structure, improvement of the classification effect of transfer learning is done by the expansion of target domain training samples and the change of training sample pairing measures.

Siamese neural network is the name given to the topology of a network that consists of dual branches and weight sharing (SNN). When a convolutional neural network serves as the Siamese neural network's branch, the resulting network is referred to as a Siamese convolutional neural network (SCNN). When compared to the conventional convolutional neural network, the most notable characteristic of the SCNN is that the original single series network is converted into a new network that consists of two parallel subbranches that have the same structure and share the same weights. This is the case in contrast to the traditional convolutional neural network. Because the input data to the network are homogeneous or heterogeneous sample pairs, models that are based on SCNN require less data for training purposes. This is due to the fact that a small number of labelled samples can be combined into a large number of training sample pairs using only a small number of

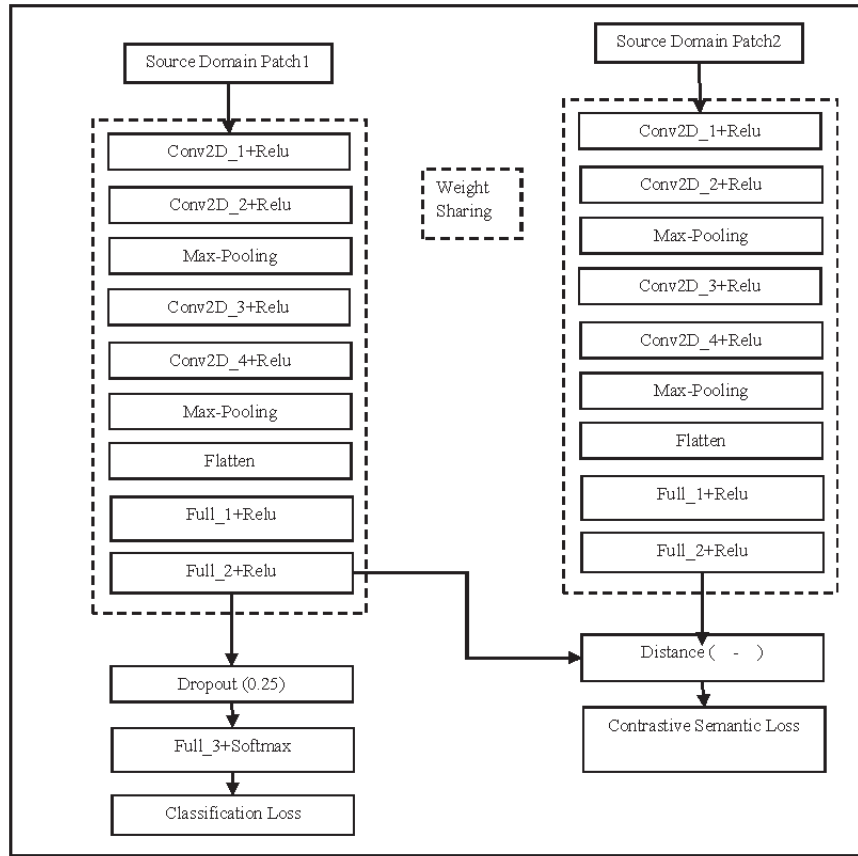


Figure 6. Schematic diagram of the structure of the Siamese convolutional neural network used in this paper.

unlabeled samples. The SCNN model architecture (see Figure 5 for reference) is utilised so that transfer learning may be accomplished. A CNN, or convolutional neural network, is used to model the embedding function g . A CNN is primarily made up of an initial convolutional layer and a fully connected layer. One of the two branches of the Siamese network is assigned to the task of training the source domain, while the other is assigned to the task of training the target domain.

3.3. Generation of paired samples

Training a Siamese network demands a sufficient number of training sample pairs, including homogeneous sample pairs and heterogeneous sample pairs. For convenience, the source and target domain sample label sets are denoted as $S_y = y_i^S (i = 1)^{(N_S)}$ and $T_y = y_j^T (j = 1)^{(N_T)}$, respectively, assuming that the two domains share C-type features.

First, the classical sample pair construction method is described. The construction of homogeneous sample pairs is relatively simple. A total of N_{same} pairs of homogeneous samples are constructed by randomly sampling $N_{perclass1}$ samples from each class of the source and target domain sample sets.

4. Experimental study and result analysis

4.1. Experimental data and environment

To reasonably evaluate the effectiveness of this method and its adaptability to different data and regions, in this study, related experiments are carried out using two different collections of hyperspectral test datasets, each of which has a different degree of similarity (a different degree of data offset). Source and Target are two subregions in the same scene image that are not connected to one another. PC (Photo Courtesy) and PU (positive and unlabeled) are two cross-scene images that are not connected to one another. The initial group of information consists of hyperspectral remote sensing photos obtained by the ROSIS-3 sensor while it was flying over the city. Figure 7 depicts the entire picture of these data in their entirety. In order to conduct the migration experiment, it was decided to involve two distinct subregions, which will be referred to as the Source image and the Target picture, respectively.



Figure 7. Full-image false-color image of the first dataset.

Their respective sizes are 172×123 pixels and 350×350 pixels, where the spectral range lies from 0.43 micrometre to 0.86 micrometre, and the spatial resolution is equal to 1.3 m. The data contains 115 spectral bands. In order to avoid the influence of noise bands during actual classification, 12 noise bands are removed, and 103 bands are reserved for classification. Figure 8 shows the Source and Target false color image maps and the corresponding actual ground reference data. The two images share four types of objects: roads, vegetation, shadows, and buildings.

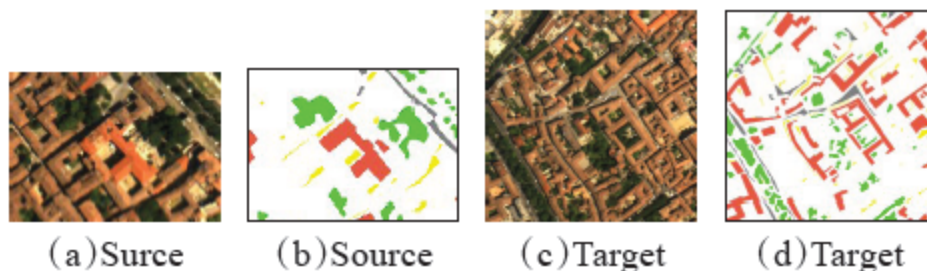


Figure 8. Source and Target's false-color images and real-world distributions.

As shown in Figure 8, the Source and Target images share four categories of objects: roads, vegetation,

shadows, and buildings. Table 2 counts the number of different samples of the Source and Target images. The Target image has a significantly higher number of labelled samples than the Source.

Table 2. Statistics of the number of samples of the first group of data sets.

S. No.	Feature Category	Source	Target
1	The Way	362	2547
2	Vegetation	1756	6302
3	Shadow	503	1298
4	Architecture	1211	16881
Total	—	3832	27028

The second set of data sets are hyperspectral images of two different regions captured by the same sensor, from Pavia Center and Pavia University in Italy, both captured by imaging spectrometer ROSIS-3, the resulting, referred to as PC and PU images. The spectral range, spatial resolution, and spectral bands of this dataset are the same as the first set of data. The sizes of the two images are 1096×492 and 610×340 pixels respectively, and there are seven types of public features.

Table 3 shows various sample statistics of the second set of data sets.

The experiments in this paper are implemented based on Python3 and MATLAB under Windows10 system. In the Python3 environment, the Keras library (with TensorFlow as the backend) is used to train and classify the SCNN, and the rest are done in the MATLAB environment. The results of the classification experiments below are the average of 10 independent experiments.

4.2. Segmentation results display and quality evaluation of target images

The experiments in this section demonstrate the segmentation and purification process in this paper and the effects of the three image homogeneous regions obtained based on the multi-scale segmentation method. The impacts of the three picture homogenous zones created using the multi-scale segmentation approach, as well as the segmentation and purification procedure used in this research. The three pictures of Target, PC, and PU after segmentation and purification, as well as the final segmentation findings. The figure shows that after the three images are cleaned up, more homogeneous large-area blocks are produced, the over segmented areas around many objects are removed, and SEEDS superpixel segmentation is then carried out in the residual

Table 3. Training and testing samples of the second set of data sets.

S. No.	Feature Category	PC		PU	
		Training Samples	Test Sample	Training Samples	Test Sample
1	Tree	763	6406	512	3044
2	Asphalt Road	649	7385	534	6531
3	Self-Sealing Brick	485	2360	503	3582
4	Asphalt Construction	798	7167	391	1220
5	Grassland	762	2895	557	17549
6	Bare Soil	817	6499	523	5229
7	Shadow	205	2205	277	977
Total	—	4479	34917	3297	38132

image area, which can be accurate and detailed. During superpixel segmentation, the number of patches is set to make the average area of patches about 40 pixels. Figures 9, 10, and 11 show the segmentation and purification process and the final segmentation results of the three images of Target, PC, and PU, respectively. It can be observed from the figure that after the three images are purified, more large-area homogeneous blocks are obtained, and the over-segmented areas at the edges of many objects are eliminated, and SEEDS superpixel segmentation is performed in the remaining residual image area, which can be precise and detailed. This part of the complex and changeable area is divided.

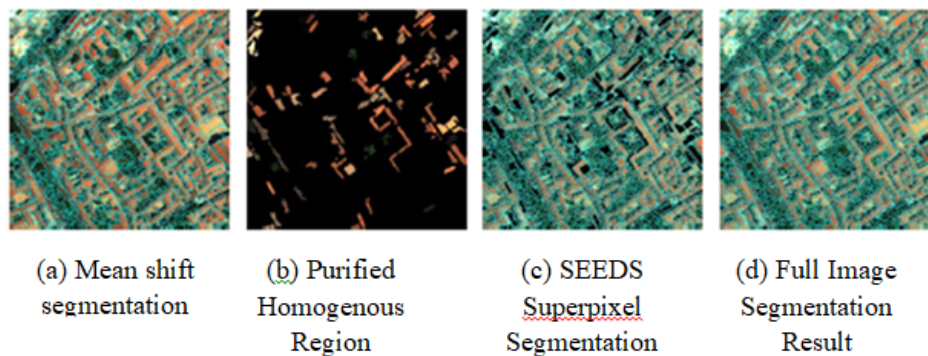


Figure 9. Target image segmentation result.

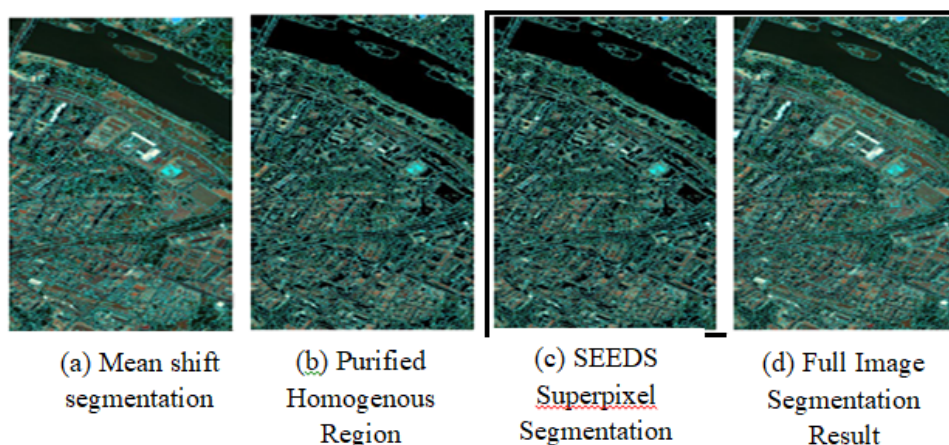


Figure 10. PC image segmentation results [41].

Table 4 shows the number of patches in the homogenous area of the three image segmentation results and the values of the two segmentation evaluation indicators. From the table, it is clear that the under-segmentation error rates of the three images are all lower than 0.05%; the attainable segmentation accuracy rates are all over 99.7%, which verifies the potential of the proposed segmentation-purification scheme.

4.3. Experimental results and analysis of object image classification

In the classification experiment, 300 labeled samples are randomly selected for each type of ground object in the source domain image, and 1 labeled sample is randomly selected for each type of target domain image as

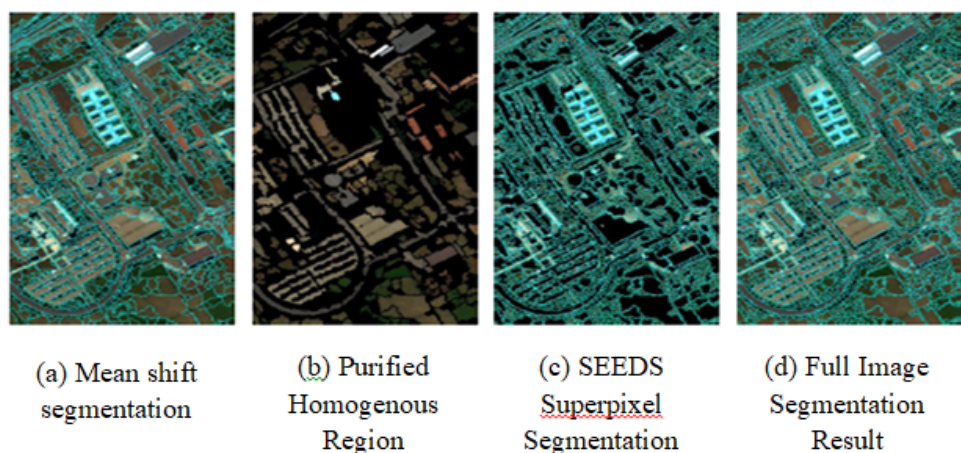


Figure 11. PU image segmentation result.

Table 4. Evaluation of segmentation results of three experimental images.

Image	Homogenous Region	Under Segmentation	Maximum Split
	Number of Plaques	Error Rate/%	Accuracy/%
Target	4655	0.04	98.79
PC	6923	0.05	99.00
PU	5287	0.04	99.61

the original training data; all remaining target domain labeled sample images are used for testing. The training samples are used to train the method in this paper and several reference methods, the training model is used to predict the test sample category, and the performance of various methods is evaluated by analyzing and comparing various indicators. On the ImageNet dataset (ImageNet large scale visual recognition challenge; ILSVRC), they underwent pretraining using millions of real-world pictures [34][35]. The principle of CNN is the transmission of information at the parametric level. Convolutional layer parameters are used by well-trained CNN models to perform a new task in the field of medicine. In particular, in TL with CNN for medical image classification, a medical image classification (target job) may be trained by making use of the generic features acquired from the natural image classification (source task), when labels are present in both domains.

In addition to the semisupervised transfer learning (SCNN-HR) of homogeneous regions and SCNN proposed in this paper, the contrasting methods in the experiments include SRC, SVM1, SVM2 and CCSA based on linear support vector machines. Among them, SRC only uses the training samples from source domain images and does not use the labeled samples in the target domain, trains the support vector machine model based on the spectral features of the samples and directly classifies the target images without transfer learning. The training samples of SVM1 add the labeled single sample of the target domain; SVM2 adds the single sample of the target domain and the amplified sample based on the homogenous region to the training samples of the source domain; CCSA is the latest semisupervised transfer learning method, using Source domain training samples and target domain labeled single samples, based on the empty-spectrum joint features of the samples train the Siamese neural network and classify the target image. CCSA is the closest to the method in this paper, but it

Table 5. Evaluation of the classification results of the first group of datasets (Source to Target)

Evaluation Indicators	SRC	SVM1	SVM2	CCSA	SCNN-HR
OA/%	75.63	75.87	78.38	79.14	82.35
AA/%	79.84	80.19	80.57	80.14	82.38
Kappa	0.6271	0.6310	0.6740	0.6720	0.7150

utilizes one labelled sample in each class in the target domain, and the pairing strategies of samples w.r.t. source and target domain respectively also differ. Deep learning (DL) algorithms' recent development has ushered in a move away from manual engineering, allowing for automated picture analysis. Convolutional neural networks (CNN) in particular have emerged as the dominant DL approach for image processing. All of the top-ranked teams in recent data challenges for medical image processing used CNN. Still, DL algorithms, like CNN, need a sizable quantity of data for training, which leads to the data scarcity issue. Some well-known difficulties are the small size of medical cohorts and the expense of expert-annotated data sets. Transfer learning (TL) approaches have been used in several research projects in an effort to solve this issue. These seek to maximise efficiency on target activities by utilising the information gained from source tasks.

Both the method in this paper and CCSA use the joint feature of the space spectrum, that is, the neighborhood subimages of all bands of a pixel are all used as their features. The size of the neighborhood window is 5×5 , which can be used for almost all images, and can achieve the purpose of utilizing spatial information without causing the edge blurring of the classification results. Each band of the neighborhood is expanded into a column of vectors, and a 5×5 neighborhood is expanded into a 25×1 vector. The data used for the transfer learning experiment is 102 bands, so the feature data of each pixel is a 25×102 matrix, and the input layer size of each branch of the Siamese network is $25 \times 102 \times 1$. Table 5 shows the experimental results and classification effects of proposed SCNN-HR method and comparison methods on the first set of data, respectively.

The SRC method directly uses the labeled samples of the source domain for training without any transfer learning method, and the classification accuracy is the lowest. However, when there exists exactly one labeled sample per class in the target domain, that is, the SVM1 method, the classification accuracy cannot be improved, because the data information of the source domain in the training sample is dominant, and a small amount of feature information in the target domain cannot affect the training of the model [36] [37]. There is a certain difference between the domain and the target domain, so the ideal classification results cannot be obtained. Due to the addition of more amplified target domain labeled sample information, SVM2 does not transfer between the two domains, and its classification accuracy can also be improved by about 3%. The classification results of the semisupervised transfer learning method CCSA are roughly comparable to those of SVM2. The method in this paper uses the homogenous region to amplify the single sample of the target domain, uses the Siamese convolutional neural network to shorten the distance between the source domain data and the target domain data, and achieves the best results among the three classification evaluation indicators. The overall accuracy and average accuracy are 82.35% and 82.38%, respectively, which are about 2% to 3% greater than the CCSA method, and the Kappa coefficient is 0.04 greater. Accuracy measures how frequently an ML model for classification is overall right. Precision measures how often a machine learning model correctly predicts the target class. Recall demonstrates if an ML model is able to locate every item in the target class. A machine

Table 6. Evaluation of the classification results of the second group of datasets (PC to PU).

Evaluation Indicators	SRC	SVM1	SVM2	CCSA	SCNN-HR
OA/%	56.41	54.97	70.42	71.13	72.20
AA/%	71.88	71.88	72.22	69.14	79.45
Kappa	0.502	0.498	0.541	0.532	0.599

Table 7. Evaluation of the classification results of the second group of datasets (PU to PC).

Evaluation Indicators	SRC	SVM1	SVM2	CCSA	SCNN-HR
OA/%	72.66	74.25	83.25	85.44	87.58
AA/%	73.24	74.92	83.17	87.42	89.63
Kappa	0.701	0.674	0.800	0.821	0.848

learning assessment statistic called the F1 score assesses a model's accuracy. It combines a model's accuracy and recall ratings. How many times a model is correctly predicted throughout the full dataset is determined by the accuracy statistic. We may really improve the forecasting ability of our model and get a significant competitive advantage by selecting the appropriate measure. The method that is presented in this study makes use of the homogeneous region in order to amplify the one sample that is taken from the target domain. In addition to this, it makes use of the Siamese convolutional neural network in order to reduce the gap in time and space that exists between the data coming from the source domain and the data coming from the target domain. As a consequence of this, the approach accomplishes the highest attainable levels of success across all three categorization assessment indicators. Both the overall accuracy and the average accuracy come in at 82.35 %, which is around 2% to 3% than the CCSA technique. Additionally, there is a 0.04-point increase in the Kappa coefficient.

Table 6 shows the classification results obtained by each method in the second set of data sets PC→PU and the corresponding classification effect diagrams.

In this group of experiments, the classification accuracy of the method in this paper is the highest, and the overall classification accuracy reaches 70.10%; compared with the SRC that directly uses the source domain data for classification, it is about 11 percentage points higher; compared to the SVM2 without the transfer method [38] [39]. There is a 4-percentage point improvement in accuracy. Compared with CCSA, the classification accuracy and average accuracy are 5 to 6 percentage points higher, which has obvious advantages. The classification effect chart shows that compared with other methods, the method in this paper is better at distinguishing between bare soil and grassland, and contains fewer noise points. Table 7 shows the classification results and classification effects obtained by different methods when PU to PC. The overall accuracy of the method in this paper reaches 87.58%; it is about 13 percentage points higher than the benchmark method SRC; it is 4 percentage points higher than the SVM2 without the transfer method; at the same time, compared with the semisupervised classification method CCSA, there are also 2 points percent improvement in accuracy [40]. The classification effect diagram shows that the asphalt pouring integrity separated by the method in this paper is the highest, and it is basically close to the real distribution of ground objects.

5. Conclusion

The structure of the Siamese neural network is suitable for transfer learning when the target domain has a few marked points, but the transfer classification effect is not ideal. This study investigates sample augmentation for the one-shot scenario to investigate the best classification accuracy that can be achieved with a single true labelled sample for each class in the study domain. When a CNN scene classification model is being trained, the idea of sample augmentation occurs. Picture segmentation on two regions with distinct characteristics-the original remote sensing picture itself and the irregular area close to the border of the object- is required to produce the homogenous area. In this paper, a semisupervised transfer classification method based on homogenous region and twin neural network structure is studied. The labeled samples are amplified by the homogenous region of the target domain obtained by segmentation, and the network model training is realized by combining the joint features of the empty spectrum. The experimental results show that the improvement measures significantly improve the classification effect of hyperspectral data, and basically meet the needs of ground object recognition and information extraction. The experimental results of the two sets of data also prove that the method explored in this paper is adaptable to hyperspectral images with different degrees of offset between the source and target domains.

References

- [1] Cui Z, Chang H, Shan S, Zhong B, Chen X. Deep Network Cascade for Image Super-Resolution. Cham, Switzerland: Springer, 2014.
- [2] Wang Z, Liu D, Yang J, Han W, Huang T. Deep networks for image super-resolution with sparse prior. In Proc. IEEE Int. Conf. Comput. Vis.; 2015. pp. 370–378. <https://doi.org/10.1109/ICCV.2015.50>
- [3] Li Y, Hu J, Zhao X, Xie W, Li J. Hyperspectral image superresolution using deep convolutional neural network. Neurocomputing 2017; 266 (1): 29–41. <https://doi.org/10.1016/j.neucom.2017.05.024>
- [4] Dong C, Loy CC, He K, Tang X. Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell., 2015; 38 (2): 295–307. <https://doi.org/10.48550/arXiv.1501.00092>
- [5] Ledig C et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit.; 2017. pp. 105–114. <https://doi.org/10.1109/CVPR.2017.19>
- [6] Zeng K, Yu J, Wan R, Li C, Tao D. Coupled deep autoencoder for single image super-resolution. IEEE Trans. Cybern., 2017; 47 (1): 27–37. <https://doi.org/10.1109/TCYB.2015.2501373>
- [7] Tarabalka Y, Tilton JC. Spectral-spatial classification of hyperspectral images using hierarchical optimization. 2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS); 2011. pp. 1-4. <https://doi.org/10.1109/WHISPERS.2011.6080900>
- [8] Myasnikov E. Segmentation of Hyperspectral Images Based on Dimensionality Estimation in Spatial Regions. 2021 International Conference on Information Technology and Nanotechnology (ITNT); 2021. pp. 1-4. <https://doi.org/10.1109/ITNT52450.2021.9649299>
- [9] Tarabalka Y, Chanussot J, Benediktsson JA. Classification based marker selection for watershed transform of hyperspectral images. 2009 IEEE International Geoscience and Remote Sensing Symposium; 2009. pp. III-105-III-108. <https://doi.org/10.1109/IGARSS.2009.5418080>
- [10] Feng J, Liu L, Cao X, Jiao L, Sun T et al. Marginal Stacked Autoencoder With Adaptively-Spatial Regularization for Hyperspectral Image Classification. In IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018; 11 (9): 3297-3311. <https://doi.org/10.1109/JSTARS.2018.2854893>

- [11] Zhang Y, Huynh CP, Habili N, Ngan KN. Material segmentation in hyperspectral images with minimal region perimeters. 2016 IEEE International Conference on Image Processing (ICIP); 2016; 834-838. <https://doi.org/10.1109/ICIP.2016.7532474>
- [12] Rajan S, Ghosh J, Crawford MM. An Active Learning Approach to Knowledge Transfer for Hyperspectral Data Analysis. 2006 IEEE International Symposium on Geoscience and Remote Sensing; 2006; 541-544. <https://doi.org/10.1109/IGARSS.2006.143>
- [13] Yuan S, Fang Y. ROSS: Robust Learning of One-Shot 3D Shape Segmentation. 2020 IEEE Winter Conference on Applications of Computer Vision (WACV); 2020. pp. 1950-1958. <https://doi.org/10.1109/WACV45572.2020.9093604>
- [14] Datta A, Chakravorty A. Hyperspectral Image Segmentation using Multi-dimensional Histogram over Principal Component Images. 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN) 2018; 857-862. <https://doi.org/10.1109/ICACCCN.2018.8748388>
- [15] Borhani M, Ghassemian H. Hyperspectral image classification based on spatial graph kernel. 2014 22nd Iranian Conference on Electrical Engineering (ICEE); 2014; 1811-1816. <https://doi.org/10.1109/IranianCEE.2014.6999833>
- [16] Golipour M, Ghassemian H, Mirzapour F. Integrating Hierarchical Segmentation Maps With MRF Prior for Classification of Hyperspectral Images in a Bayesian Framework. In IEEE Transactions on Geoscience and Remote Sensing, 2016; 54 (2): 805-816. <https://doi.org/10.1109/TGRS.2015.2466657>
- [17] Hao Q, Sun B, Li S, Crawford MM, Kang X. Curvature Filters-Based Multiscale Feature Extraction for Hyperspectral Image Classification. In IEEE Transactions on Geoscience and Remote Sensing, 2022; 60(5507916): 1-16. <https://doi.org/10.1109/TGRS.2021.3091860>
- [18] L Sun, B Jeon, Y Zheng, Y Xu, Z Wu. Homogeneous region based low rank representation in hidden field for hyperspectral classification. 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2017; 4758-4761. <https://doi.org/10.1109/IGARSS.2017.8128065>
- [19] Tu B, Wang J, Kang X, Zhang G, Ou X et al. KNN-Based Representation of Superpixels for Hyperspectral Image Classification. In IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018; 11 (11): 4032-4047. <https://doi.org/10.1109/JSTARS.2018.2872969>
- [20] Li Z, Yang Z, Xiong H. Homogeneous region segmentation for SAR images based on two steps segmentation algorithm. 2015 International Conference on Computers, Communications, and Systems (ICCCS) 2015; 196-200. <https://doi.org/10.1109/CCOMS.2015.7562900>
- [21] Yue J, Mao S, Li M. A deep learning framework for hyperspectral image classification using spatial pyramid pooling. Remote Sensing Letters, 2016; 7 (9): 875-884. <https://doi.org/10.1080/2150704X.2016.1193793>
- [22] Chen Y, Jiang H, Li C, Jia X, Ghamisi P. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. In IEEE Transactions on Geoscience and Remote Sensing, 2016; 54 (10): 6232-6251. <https://doi.org/10.1109/TGRS.2016.2584107>
- [23] Foody GM. Thematic map comparison. Photogrammetric Engineering & Remote Sensing, 2004; 70 (5): pp. 627-633. <https://doi.org/10.14358/PERS.70.5.627>
- [24] Xu Y, Zhang L, Du B, Zhang F. Spectral-spatial unified networks for hyperspectral image classification. In IEEE Transactions on Geoscience and Remote Sensing, 2018; 56 (10): 5893-5909. <https://doi.org/10.1109/TGRS.2018.2827407>
- [25] Zhong Z, Li J, Luo Z, Chapman M. Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework. In IEEE Transactions on Geoscience and Remote Sensing, 2018; 56 (2): 847-858. <https://doi.org/10.1109/TGRS.2017.2755542>
- [26] Xue Z, Zhou Y, Du P. S3Net: Spectral-spatial Siamese network for few-shot hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 2022; 60 (5531219): 1-19. <https://doi.org/10.1109/TGRS.2022.3181501>

- [27] Feng H, Hou B, Gong M. SAR Image Despeckling Based on Local Homogeneous-Region Segmentation by Using Pixel-Relativity Measurement. In *IEEE Transactions on Geoscience and Remote Sensing*, 2011; 49 (7): 2724-2737. <https://doi.org/10.1109/TGRS.2011.2107915>
- [28] Wan T, Canagarajah N, Achim A. Multiscale Color-Texture Image Segmentation with Adaptive Region Merging. 2007 *IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*; 2007. pp. I-1213-I-1216. <https://doi.org/10.1109/ICASSP.2007.366132>
- [29] Kachouri R, Soua M, Akil M. Unsupervised image segmentation based on local pixel clustering and low-level region merging. 2016 *2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP) 2016*; 177-182. <https://doi.org/10.1109/ATSIP.2016.7523091>
- [30] Ji S, Park HW. Image segmentation of color image based on region coherency. *Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No.98CB36269) 1998*; 1 (1): 80-83. <https://doi.org/10.1109/ICIP.1998.723425>
- [31] Chopra S, Hadsell R, LeCun Y. Learning a similarity metric discriminatively, with application to face verification. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) 2005*; 539-546.
- [32] Mishra A, Eichel JA, Abbott T. DREM: Decoupled region energy model for image segmentation. 2015 *IEEE International Conference on Image Processing (ICIP)*; 2015; 3354-3357, <https://doi.org/10.1109/ICIP.2015.7351425>
- [33] Motiian S, Piccirilli M, Adjeroh DA, Doretto G. Unified deep supervised domain adaptation and generalization. In *Proceedings of the IEEE international conference on computer vision 2017*; 5715-5725.
- [34] Fernandez G, Wittenbrink CM. Coding of spectrally homogeneous regions in multispectral image compression. *Proceedings of 3rd IEEE International Conference on Image Processing*; 1996. pp. 923-926. <https://doi.org/10.1109/ICIP.1996.561055>
- [35] Zhang T, Cheng J, Fu H, Gu Z, Xiao Y et al. Noise adaptation generative adversarial network for medical image analysis. *IEEE Trans Med Imaging*. 2019; 39(1): 1149-59. <https://doi.org/10.1109/TMI.2019.2944488>
- [36] li-jie Y, De-sheng W L. Yue-Zong. A Hybrid Method of Block-Based Region Growing and Visible Color Difference for Color Image Segmentation. 2009 *First International Workshop on Education Technology and Computer Science 2009*; 11-315. <https://doi.org/10.1109/ETCS.2009.595>
- [37] Cardelino J, Randall G, Bertalmio M, Caselles V. Region Based Segmentation Using the Tree of Shapes. 2006 *International Conference on Image Processing 2006*; 2421-2424. <https://doi.org/10.1109/ICIP.2006.312951>
- [38] Chamalis T, Likas A. Region merging for image segmentation based on unimodality tests. 2017 *3rd International Conference on Control, Automation and Robotics (ICCAR) 2017*; 381-384. <https://doi.org/10.1109/ICCAR.2017.7942722>
- [39] Yongming Li, Dongming Lu, Xiquan Lu, Jianming Liu. Interactive color image segmentation by region growing combined with image enhancement based on Bezier model. *Third International Conference on Image and Graphics (ICIG'04) 2004*; 96-99. <https://doi.org/10.1109/ICIG.2004.89>
- [40] Chen J, Xia J, Du P, Chanussot J. Combining Rotation Forest and Multiscale Segmentation for the Classification of Hyperspectral Data. In *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2016; 9 (9): 4060-4072. <https://doi.org/10.1109/JSTARS.2016.2524517>
- [41] Ge Z, Cao G, Li X, Fu P. Hyperspectral image classification method based on 2D-3D CNN and multibranch feature fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020; 13 (1): 5776-5788. <https://doi.org/10.1109/JSTARS.2020.3024841>