remote sensing

MDPI

Article

A Comprehensive Evaluation of Approaches for Built-up Area Extraction from Landsat OLI Images Using Massive Samples

Tao Zhang [1], Hong Tang [2,*]

[1] Beijing Key Laboratory for Remote Sensing of Environment and Digital Cities, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China; zhangtaobnu@mail.bnu.edu.cn;
[2] State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and Institute of Remote Sensing and Digital Earth of Chinese Academy of Sciences, Beijing 100875, China; tanghong@bnu.edu.cn;
* Correspondence: tanghong@bnu.edu.cn; Tel.: +86-10-5880-6401

Received: date; Accepted: date; Published: date

Abstract: Detailed built-up area information is valuable for mapping complex urban environments. Although a large number of classification algorithms about built-up areas have been developed, they are rarely tested from the perspective of feature engineering and feature learning. Therefore we launched a unique investigation to provide a full test of the OLI imagery for 15-m resolution built-up area classification in 2015, in Beijing, China. Training a classifier requires many sample points, and we propose a method based on the ESA's 38-meter global built-up area data of 2014, Open Street Map and MOD13Q1-NDVI to achieve rapid and automatic generation of a large number of sample points. Our aim is to examine the influence of a single pixel and image patch under traditional feature engineering and modern feature learning strategies. In feature engineering, we consider spectra, shape and texture as the input features, and SVM, random forest (RF) and AdaBoost as the classification algorithms. In feature learning, the convolution neural network (CNN) is used as the classification algorithm. In total, 26 built-up land cover maps were produced. Experimental results show that: (1) the approaches based on feature learning are generally better than those based on feature engineering in terms of classification accuracy, and the performance of ensemble classifiers e.g., RF, is comparable to that of CNN. Two dimensional CNN and the 7 neighborhood RF have the highest classification accuracy of nearly 91%. (2) Overall, the classification effect and accuracy based on image patches are better than those based on single pixels. The features that can highlight the information of the target category (for example, PanTex and EMBI) can help improve classification accuracy.

Keywords: built-up area; classification; Landsat 8- OLI; feature engineering; feature learning; CNN; accuracy evaluation

1. Introduction

Built-up area refers to the land of urban and rural residential and public facilities. Built-up areas are one of the most important elements of land use and play an extremely important role in urban development planning [1]. Extracting built-up areas is crucial for mapping and managing complex urban environments across local and regional scales [7-11]. Landsat images are frequently used and are a good source of data for generating such information over large areas [39, 42].However, mapping built-up land poses a significant challenge for remote sensing due to the high spatial frequency and heterogeneity of surface features. Various algorithms have been applied to extract built-up areas, including supervised classification, unsupervised clustering, and reinforcement learning [15-17]. In the process of built-up area extraction, the most important considerations are

how to design or learn better features to characterize the buildings and how to choose a more appropriate classification strategy. Congcong Li [5] tested two unsupervised and 13 supervised classification algorithms to distinguish urban land in Guangzhou city and then assessed all algorithms in a per-pixel classification decision experiment and all supervised algorithms in a segment-based experiment. Rahman [6] compared the influence of spatial resolution, spectral band set and the classification approach for mapping detailed urban land cover in Nottingham, UK. A WorldView-2 image provide the basis for a set of 12 images with variable spatial and spectral characteristics within three different approaches (maximum likelihood (ML), support vector machine (SVM) and object-based image analysis (OBIA)) to yield 36 output land cover maps. D. LU [53] summarized the major advanced classification approaches and the techniques used for improving classification accuracy.

In this paper, we focus on the following three aspects: (1) In the large region of mapping built-up areas, Google Earth Engine (GEE) is used to obtain high-quality images. (2) Using existing built-up area data products and open map data, a large number of samples are selected quickly and automatically, and then, the samples are filtered and corrected. (3) From the viewpoint of feature engineering and feature learning, the influence of the classification strategy and the features on the result of built-up area extraction is synthetically analyzed.

Next, we review the methods and techniques of feature engineering and feature learning and the classification strategies based on single pixels and image patches. The application and advantages and disadvantages of these methods and techniques in built-up area extraction are discussed. We also discuss the progress and shortcomings of built-up area mapping based on low and medium resolution images.

1.1.Feature Engineering versus Feature Learning

A key step for pattern recognition and classification is to select independent and measurable features with a large amount of information, distinction and independence. According to prior knowledge, feature engineering performs mathematical operations on the image to obtain the typical and iconic features that can represent the extracted object. This is equivalent to the realization of feature transformation, mapping from the original image feature space to a new feature space after feature engineering. Different combinations of bands can highlight different surface features. The simplest feature transformation of remote sensing images is the band operation. In [2], [3] and [4], normalized difference building index (NDBI), an index-based build-up index (IBI) and a texture-derived built-up presence index (PanTex) were proposed to characterize buildings. However these methods based on the remote sensing index have a strong dependence on threshold selection, and finding a suitable threshold is very difficult. In recent years, very high resolution and hyperspectral images have been gradually used in building extraction. The texture, shape, geometry and three-dimensional features of images have been applied to recognize and distinguish objects. Many methods based on morphological filtering [7], spatial structure features [8], grayscale texture features [9], image segmentation [10], geometric features [11] and three-dimensional modeling [12] have increasingly been applied to building extraction. Pattern classification based on feature engineering has a good advantage in extracting certain ground objects (vegetation and water). However, for the recognition and classification of built-up areas, because of uneven distribution and fragmentation of buildings, large surface spectral heterogeneity and morphology characteristics without fixed pattern, it is difficult to find a suitable feature in a wide range of urban and suburban areas.

In recent years, with the increase of artificial intelligence and large data, pattern classification based on feature learning has become a popular research topic, especially in in-depth learning [13-17] and reinforcement learning [18-23]. Feature learning automatically learns and utilizes features from raw data. Deep learning (DL) can automatically extract hierarchical data features by unsupervised or semi-supervised feature learning algorithms. In contrast, traditional machine learning methods require manual design features. DL is a representation learning algorithm based on large scale data in machine learning. Modern DL methods have often been applied successfully in

the field of feature learning, such as self-encoder [24], Restricted Boltzmann Machine [25] and Generative Adversarial Networks [26]. These implement automatic learning abstract feature representation in an unsupervised or semi-supervised manner, and their results support advanced achievements in areas such as speech recognition, image classification [27], and object recognition [28]. With the rapid development of CNNs, especially the excellent performance of deep convolution neural networks on the ImageNet contest [29-32], CNN has shown great advantages in image pattern recognition, scene classification [33], object detection and other issues. An increasing number of researchers have applied CNN to remote sensing image classification. In [34], [35] and [36], CNNs of different structures were used for building extraction. Yang [37] showed that the combination of a subset of spectral bands can promote the classification accuracy of convolution neural networks.

1.2. Pixel–Based versus Patch-Based Classification

With the improvement of image spatial resolution, the basic unit of remote sensing land cover mapping has undergone a transformation from image pixels to image objects (segments and patches). The goal of remote sensing land cover mapping is usually to obtain the semantic category of each pixel. Traditionally, built-up extraction has been conducted using pixel-based approaches, where land cover classes are allocated to each individual pixel. In a feature space, a classifier (e.g., SVM, KNN) is used to separate the feature space into several regions. In the transformation of the feature space and the image plane space, there is the problem of the same object with different spectra and different objects with the same spectrum because the spatial relationship cannot be considered. Therefore, the classification results of the image plane will exhibit salt-and-pepper noise and fragmentation. Rahman Momeni [6] compared the influence of spatial resolution, spectral band set and the classification approach for mapping detailed urban land cover based on WorldView-2 images. Their results demonstrate that spatial resolution is clearly the most influential factor when mapping complex urban environments. Lin Wang [38] identified and inspected the urban built-up area boundary based on the temperature retrieval method, and used qualitative and quantitative analysis methods to analyze the spatial-temporal characteristics of the Jingzhou urban built-up area expansion from 1990 to 2014.

With the improvement of image spatial resolution, especially the launching of SPOT, QuickBird and Worldview, a large number of high-resolution images are publicly available. The era of large-scale remote sensing data has come. High-resolution images, hyperspectral images and radar data are widely used to extract built-up areas. Relying on single pixel spectral information cannot adequately describe and reflect the feature information of ground objects. Instead of the pixels' features, one might use image patches as the features of geo-objects. In one image patch, the spatial relations and semantic links between pixels are considered, regarding a patch as one or more target objects, such as scene recognition, semantic segmentation, object detection. Ping Zhong [39] presented a multiple conditional random fields (CRFs) ensemble model to incorporate multiple features and learn their contextual information, and the experiments on a wide range of images show that their ensemble model produces higher built-up extraction accuracy than single CRF. Xiaogang Ning [40] presented a method for extracting built-up areas from VHSR remote sensing imagery using feature-level-based fusion of right angle corners, right angle sides and road marks. On average, the completeness and the quality of their proposed method are 17.94% and 13.33% better than those of the PanTex method.

1.3. Built-up Area Extraction from Medium-Resolution Images

Although there are an increasing number of high-resolution images, medium-resolution images are still the most widely used for a wide range of ground object extraction, because of the limited computer performance and considerable data mining technology. Worldwide, the spatial resolution of built-up area data ranges from low to high with 500 m, 250 m, 38 m and 30 m. The IGBP scheme was classified using the C4.5 decision tree algorithm that ingested a full year of 8-day MODIS Nadir BRDF-Adjusted Reflectance [41]. Jie Wang [42] utilized a random forest classification algorithm to map global land cover in 2001 and 2010 with spatial-temporal consistency based on MODIS data and

Landsat images. Peng Gong [43] produced the first 30 m resolution global land-cover maps based on four classifiers(maximum likelihood, J4.8 decision tree, random forest and support vector machine) using Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) data. Based on TM and ETM + images, Jun Chen [44] applied the pixel-object hierarchical classification method to extract the global man-made surface, and the user accuracy reached 80%. Using the symbolic classification algorithm [45], ESA processed massive Landsat images and high-resolution images to extract 38-meter resolution global residential areas in 1975, 1990, 2000 and 2014, with an overall accuracy of more than 85%. Based on the Google Earth Engine (GEE) platform [46], Xiaoping Liu proposed the urban comprehensive land use index [47], and found the appropriate threshold in the global sub-climate areas and extracted multi-temporal urban built-up areas [48].

In the smaller region, research on built-up area extraction methods has become a popular topic. Ping Zhang [49] proposed an empirical normalized difference of a seasonal brightness temperature index (NDSTI) for enhancing a built-up area based on the contrast heat emission seasonal response of a built-up area to solar radiation, and adopted a decision tree classification method for the rapid and accurate extraction of the built-up area. Ran Goldblatta [10] presented an efficient and low-cost machine-learning approach for pixel-based image classification of built-up areas at a large geographic scale using Landsat data. Their methodology combines nighttime-lights data and Landsat 8 and overcomes the lack of extensive ground reference data. Xiaolong Ma [50] presented a sample-optimized approach for classifying urban area data in several cities of western China using a combination of the DMSP-OLS for nighttime-light data, Landsat images, and GlobeLand30. Ran Goldblatt [51] applied a classification and regression tree, SVM and random forests to extract urban areas in India based on a single pixel using the GEE platform.

In this paper, we compare the accuracy and efficiency of the approaches for built-up area extraction from Landsat 8-OLI images based on single pixels or image patches in two perspectives of feature engineering and feature learning. We systematically and comprehensively compare the impact of features and classifiers on built-up area extraction results using 15-meter resolution OLI-images. Moreover, given the influential role that the classification approach plays on output accuracy, and how this is linked intrinsically with image specifications, all image data sets are classified using parametric and non-parametric pixel-based and patch-based, classifiers. This enables a fuller and more robust assessment of the Landsat 8 data, but also transmits helpful and practical information for urban planners and other user communities on the level of thematic detail that can be achieved when mapping complex built-up areas. Finally, an analysis is conducted using a relatively large image covering of approximately 32400 km^2 of the city of Beijing, China and its environs. This means that built-up area extraction is generated at a scale of practical value and relevance (the whole city-scale), unlike the earlier experiments of Congcong Li [5] and Rahman Momeni [6], which were limited to very small, local areas.

2. Study Site and Data

2.1. Study Area and OLI Image

The study area is the city of Beijing, the capital of China, located at 107E longitude, 36N latitude. Beijing has a population of slightly more than 21 million. The climate is a typical north temperate semi-humid continental monsoon climate, with a hot and rainy summer, a cold and dry winter, and a short spring and autumn. The landscape consists of 62% mountains and 38% plains. The topography of Beijing is high in the northwest and low in the southeast, with an altitude of approximately 43.5 m. Beijing is a typical international metropolis with prosperous business circles and developed transportation systems. The objects on the ground surface are complex and heterogeneous. Within the Fifth Ring, the buildings are densely distributed, while the buildings are sparse in the suburb outside the Fifth Ring. Therefore, we determined that choosing Beijing as an experimental area is typical, scientific and reasonable.

The Landsat 8-OLI land imager has 9 bands and the imaging width is 185x185 km. The resolution of Band1 to Band7 is 30 meters, and Band8 is a panchromatic band with 15-meter

resolution. Compared with the ETM sensor on Landsat-7, the OLI terrestrial imager has made the following adjustments: (1) The wavelength of Band 5 is adjusted to 0.845 - 0.885 um, eliminating the influence of water vapor absorption at 0.825 um. (2)The band 8 panchromatic wave band is narrow, so the vegetation and non-vegetation areas can be distinguished better. (3) The newly added blue band of Band 1 (0.433-0.453 um) is mainly used for coastal zone observation.

Table 1. Comparison of Landsat 7 and Landsat 8 satellite bands

Landsat 8-OLI				**Landsat 7-ETM**			
Band Index	Band Name	Bandwidth (μm)	Resolution (m)	Band Index	Band Name	Bandwidth (μm)	Resolution (m)
Band 1	COASTAL	0.43 – 0.45	30				
Band 2	BLUE	0.45 – 0.51	30	Band 1	BLUE	0.45 – 0.52	30
Band 3	GREEN	0.53 – 0.59	30	Band 2	GREEN	0.52 – 0.60	30
Band 4	RED	0.64 – 0.67	30	Band 3	RED	0.63 – 0.69	30
Band 5	NIR	0.85 – 0.88	30	Band 4	NIR	0.77 – 0.90	30
Band 6	SWIR 1	1.57 – 1.65	30	Band 5	SWIR 1	1.55 – 1.75	30
Band 7	SWIR 2	2.11 – 2.29	30	Band 7	SWIR 2	2.09 – 2.35	30
Band 8	PAN	0.50 – 0.68	15	Band 8	PAN	0.52 – 0.90	15

We selected OLI images on GEE, taking into account the large amount of cloud cover in spring and autumn, so the dates of the images are mainly in summer. To ensure data quality, we utilized the minimum cloud cover synthesis algorithm provided by GEE to preprocess and generate the required images [46]. To facilitate built-up area extraction at 15 m resolution, the first seven bands (Band 1 to Band 7) of the Landsat 8 OLI images are up-sampled to 15 meters using the nearest neighborhood sampling. Then we clipped the image with a size of 12000*12000 pixels. As shown in figure 1, the false color (7, 6, 4 band combination) shows that the quality of the data is good and meets the requirements. To facilitate the mapping display and more clearly express the details, we choose two representative regions A and B with sizes of 1000 * 1000. B is an urban central region with a dense distribution of buildings, while A is in the suburban region and the distribution of buildings is sparse. However we still consider the whole research area as the analysis object when we conduct the experiment.

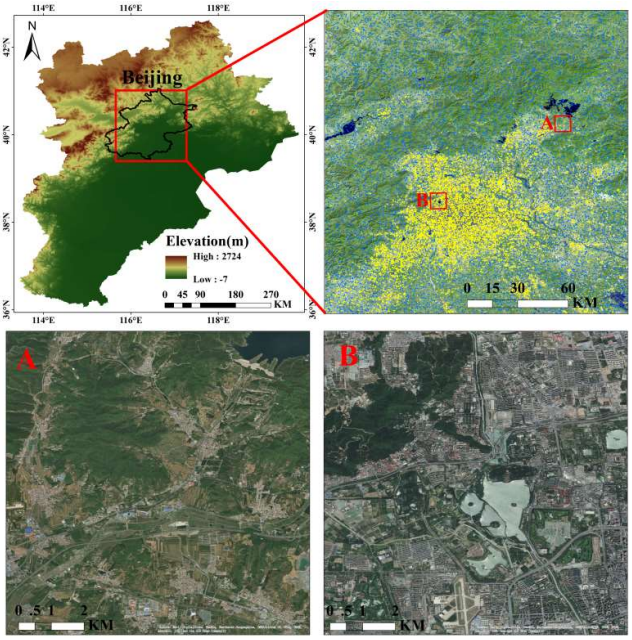

Figure 1. Research area and samples

2.2. Massive Samples Automatically Selected from Built-up Production

The training and testing samples are automatically selected from the 38-meter global built-up production of ESA in 2014 [45]. A large number of sample points are automatically generated, filtered and corrected. As shown in figure 2, the detailed process includes three steps: (1) Randomly selecting 20 thousand sample points in each experimental area; (2) Using the buildings and water data sets of Open Street Map (OSM) in China and the MOD13Q1-NDVI data to filter and correct the selected sample points. The aim is to modify the built-up sample points in the vegetation area and the water body into non-built-up sample points, and to modify the non-built-up sample points in the built-up area into the built-up sample points; (3) Combining with ArcGIS Online Image (ESRI image) for manual correction. Finally, sample points of built-up area and non-built-up area are obtained. The sample points after filtration and correction are hierarchically divided into training samples and test samples at a proportion of 6:4. Finally, there are 11499 training samples and 7667 test samples ultimately. In figure 1, the yellow points represent built-up samples, and the blue points represent non built-up samples.

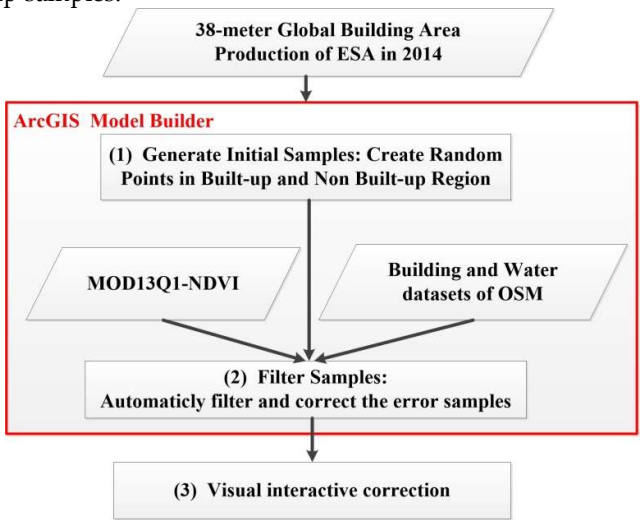

Figure 2. Sample generation and correction

3. Research Methods

In this paper, the accuracy and efficiency of extracting a 15-m resolution built-up area based on a single pixel and image patch are compared and analyzed comprehensively in the two perspectives of feature engineering and feature learning. As shown in table 2, we proceed from four aspects: (1) single-pixel classification under feature engineering, that is, pre-set features, using the original 8-band spectrum, building remote sensing index (NDBI and IBI), morphological building index (EMBI), building area presence index (PanTex), texture feature mean of these five features, the classifier is SVM, RF and AdaBoost. (2) Image patch classification under feature engineering, the original 8-band features, considering the single pixel within the neighborhood of 3*3, 5*5, 7*7 pixel patches, that is, generate a new feature vector, and then classify. The classifier is still SVM, RF and AdaBoost. (3) Single pixel classification under feature learning. For eigenvectors of 8 bands on a single pixel, one-dimensional CNN is used to learn the spectral features, and then classification is realized. (4) Image patch classification under feature learning, the original 8 bands, considering the 5*5 neighborhood pixel block, using two-dimensional CNN, while learning the spectral and plane spatial location relationship features, and then built-up area is distinguished.

Table 2. Overall framework of method and technology

	Feature Engineering				Feature Learning	
	Feature description		Abbreviations	Classifier	Network Architecture	Abbreviations
Pixel	Spectrum	Original 8 Bands	**OS8**	**SVM** / **RF** / **AdaBoost**	CNN with one dimensional Convolution on inputting bands of each pixel	**CNN_1D**
		Pan+NDBI+IBI	**RSBI**			
	Morphology	Pan+ EMBI	**EMBI**			
	Texture	Pan+Build-up presence index	**PanTex**			
		Texture from GLCM	**TEGL**			
Patch	Original 8 Bands	3*3 neighborhood	**P3**		CNN fed with an image patch of size 5*5	**CNN_2D**
		5*5 neighborhood	**P5**			
		7*7 neighborhood	**P7**			

3.1. Feature Engineering

3.1.1. Pixel-Based classification

Based on single pixel classification, spectral (OS8 and RSBI), morphological (EMBI), texture features (PanTex and TEGL) are considered in this paper. EMBI and texture features are computed mathematically from panchromatic band, while NDBI and IBI in RSBI are calculated from multispectral bands. The panchromatic band is very important for 15-m-resolution built-up area extraction, and to make the feature dimension greater than 1, RSBI, EMBI and PanTex include panchromatic band. The range of values of each feature is different, to train better classifiers, so all the features are normalized to the range of 0 to 1, and then linearly stretched to 0-255 with the type of UINT8. We apply SVM, RF and AdaBoost to realize classification for a single pixel. There will be 15 classification results based on 3 kinds of classifiers and 5 kinds of features.

(1) **Spectrum**

OS8: The first seven bands, which were sampled up to 15 meters, were stacked with panchromatic band to form 8-band data as the original spectral feature.

RSBI: Buildings have unique spectral characteristics. Through the combination and operation of different bands, the remote sensing index that can characterize the building information is obtained. In [2], Y. ZHA et al. proposed a method based on Normalized Difference Built-up Index (NDBI) to automate the process of mapping built-up areas. Built-up areas are effectively mapped through arithmetic manipulation of NDBI (see Equation (1)) derived from near infrared (NIR) and short wave infrared band (SWIR1).

$$\mathrm{NDBI} = \frac{SWIR1 - NIR}{SWIR1 + NIR} \tag{1}$$

where *SWIR1* is Band 6 of Landsat 8, and *NIR* is Band 5.

In [3], a new index derived from existing indices – an index-based built-up index (IBI) is proposed for the rapid extraction of built-up land features in satellite imagery. The IBI is distinguished from conventional indices by its first-time use of thematic index-derived bands including RED, GREEN, near-infrared (NIR) and short wave infrared (SWIR1) to construct an index rather than by using original image bands. Built-up areas are effectively extracted by setting the appropriate threshold for IBI. The IBI is calculated using Equation (2).

$$\mathrm{IBI} = \frac{2 * SWIR1 / (SWIR1 + NIR) - [NIR / (NIR + RED) + GREEN / (GREEN + SWIR1)]}{2 * SWIR1 / (SWIR1 + NIR) + [NIR / (NIR + RED) + GREEN / (GREEN + SWIR1)]} \tag{2}$$

where *SWIR1* is Band 6 of Landsat 8, *NIR* is Band 5, *RED* is Band 4, and *GREEN* is Band 3.

(2) **Morphology: Pan + EMBI**

Referring to the study of Huang Xin [52], EMBI (see Equation (3)), regarded as a characteristic feature of a building object, is the mean value of the multi-directional and multi-scale differential morphological sequence.

$$\mathrm{EMBI} = \frac{\sum_{d_i}^{D_N} \sum_{s_j}^{S_N} \mathrm{DMP}_{W-TH_{OFC}}(d_i, s_j)}{D_N \times S_N} \tag{3}$$

where, $\mathrm{DMP}_{W-TH_{OFC}}(d_i, s_j)$ denotes the different morphological characteristics of the size and direction of structural elements, D_N and S_N denote the number of directions and dimensions of structural elements respectively.

Considering the building size on 15-meter resolution panchromatic image, we set the size of the linear structure element from one pixel to six pixels, and the direction from 10 degrees to 180 degrees, so D_N =18 and S_N = 6, and there are 108 linear structure elements. EMBI is calculated based on these linear structure elements.

(3) **Texture**

Built-up presence index: Pan + PanTex

Based on the high local contrast of buildings, a texture calculation method of the building area existence index (PanTex) was proposed by Pesaresi [4]. The method is slightly adjusted in this paper. For the panchromatic image, the grayscale co-occurrence matrix (GLCM) of the 12 displacement vectors shown in figure 3 is calculated in the sliding window of 5*5. Then, based on each GLCM, the contrast texture statistical features are computed. Finally, 12 contrast features of all displacement vectors are maximized as the pixel values of the center pixel in the sliding window. The PanTex is calculated using Equation (4).

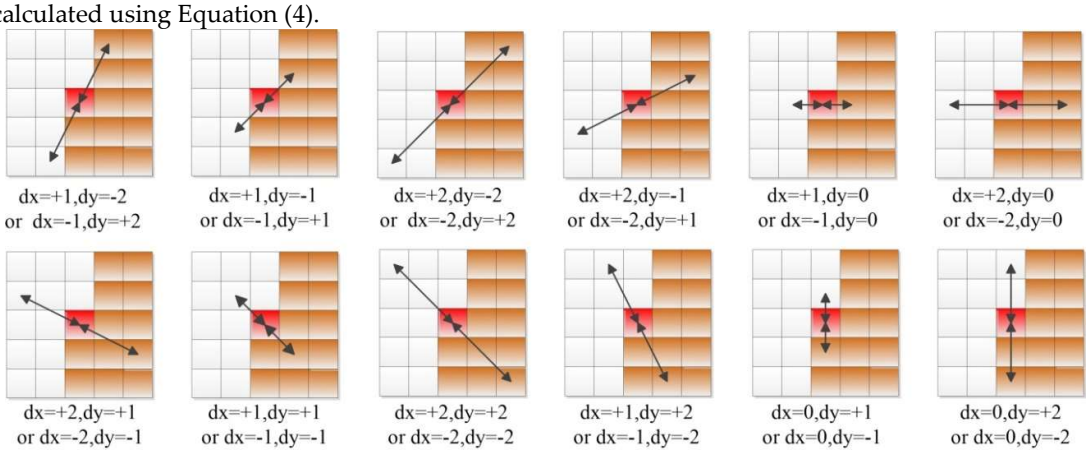

Figure 3. 12 Displacement vectors in the 5*5 window

$$PanTex = \max(w = 5, V_i, CON) \tag{4}$$

where, w = 5 indicates the size of the analysis window as 5 * 5, V_i represents the displacement vector used to calculate the GLCM, the maximum value of i is 12, and CON represents the contrast characteristics calculated based on the GLCM. The CON is calculated using Equation (5).

$$CON = \sum_i \sum_j (i - j)^2 * P(i, j) \tag{5}$$

where, i and j are the discrete values of the row and column directions in GLCM, and $P(i, j)$ is the corresponding value of i and j in GLCM.

Five mean texture features: Texture from GLCM

ASM: The Angular Second Moment (ASM) reflects the distribution of the grayscale and the size of the texture. The value of ASM depends on the distribution of elements in GLCM. If the values of all elements tend to be the same, the ASM values are smaller; however, if the values of the elements

are more different and distributed more centrally, the ASM values are larger. In addition, the large ASM value means that the distribution of the texture patterns is more uniform and regular.

$$ASM = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} P(i,j)^2 \tag{6}$$

CON：The contrast (CON) can show us the clarity of the image and the depth of the texture. It reflects how the values of the GLCM elements are distributed and the local variation information of the image, that is, the moment of inertia near the main diagonal of GLCM. The greater the value of the element from the diagonal in GLCM, the greater the contrast.

$$CON = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} (i-j)^2 P(i,j) \tag{7}$$

COR：The correlation (COR) measures the similarity of the spatial gray level co-occurrence matrix elements in row or column directions, thus the magnitude of correlation value reflects the local gray level correlation in the image. When the values of matrix elements are uniform and equal, the value of the correlation is large; on the contrary, if the pixel values of the matrix differ greatly, the value of the correlation is small. If there are horizontal directional textures in the image, the COR of the horizontal direction matrix is larger than the COR value of the other matrix.

$$COR = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} \frac{ijP(i,j) - \mu_1\mu_2}{\sigma_1\sigma_2} \tag{8}$$

where μ_1 , μ_2 , σ_1 and σ_2 respectively are given by:

$$\mu_1 = \sum_{i=0}^{L-1} i \sum_{j=0}^{L-1} P(i,j)$$

$$\mu_2 = \sum_{i=0}^{L-1} j \sum_{j=0}^{L-1} P(i,j)$$

$$\sigma_1 = \sum_{i=0}^{L-1} (i-\mu_1)^2 \sum_{j=0}^{L-1} P(i,j)$$

$$\sigma_2 = \sum_{i=0}^{L-1} (j-\mu_2)^2 \sum_{j=0}^{L-1} P(i,j)$$

ENT: Entropy (ENT) is the measure of the amount of information that an image has. Texture information is also a random measure of the image information. Entropy is larger when all elements in GLCM have the largest randomness, all values are almost equal, and the elements are dispersed. ENT represents the degree of non-uniformity or complexity of the texture in images.

$$ENT = -\sum_{i=0}^{L-1}\sum_{j=0}^{L-1} P(i,j) \lg P(i,j) \tag{9}$$

HOM: The homogeneity (HOM) reflects the homogeneity of the image texture and measures the local variation of image texture. A large value indicates that there is a lack of variation among different regions of the image texture, and the local distribution is very uniform.

$$HOM = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} \frac{P(i,j)}{1+(i-j)^2} \tag{10}$$

For each pixel of a 15-meter resolution panchromatic gray image, all GLCMs of all displacement vectors are calculated by considering the neighborhood window of 5 * 5. Then the ASM, CON, COR, ENT, HOM corresponding to each GLCM are calculated. In the end, the average value is determined. Five texture features based on GLCM in the 5 * 5 neighborhood are finally obtained: mean-ASM, mean-CON, mean-COR, mean-ENT and mean-HOM.

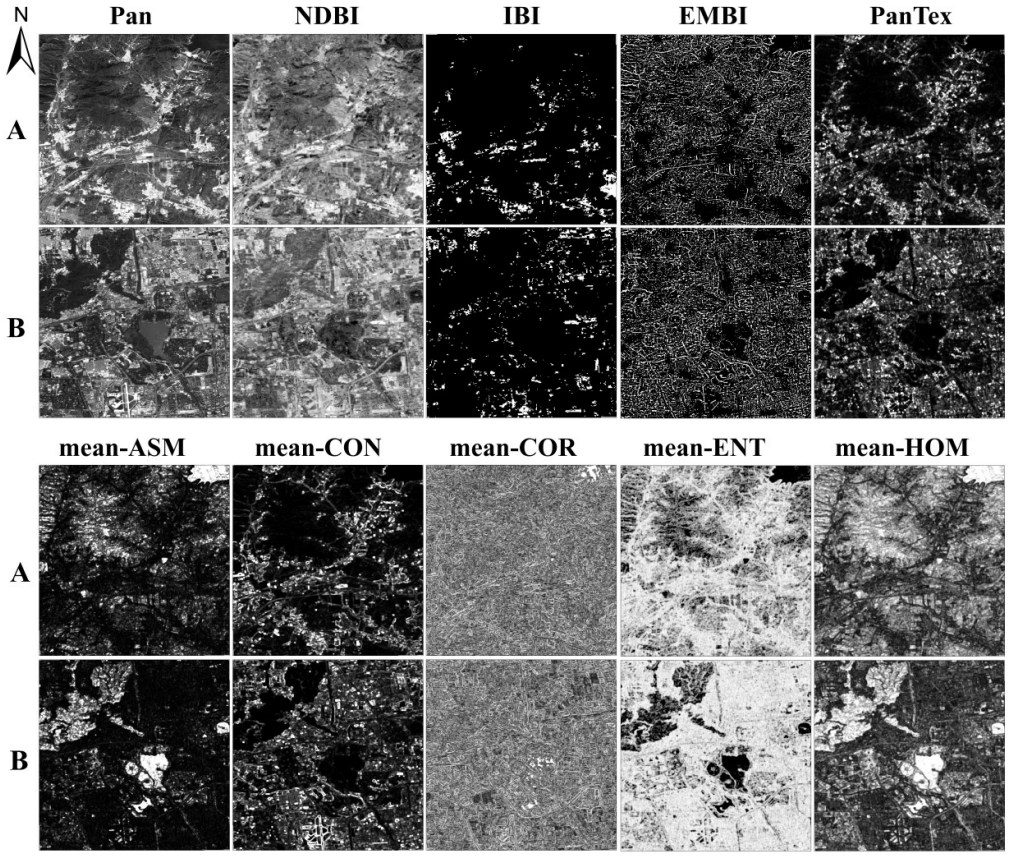

Figure 4. Feature maps of different feature descriptors

3.1.2. Patch-based classification

For each pixel, we consider its neighborhood windows of 3 * 3, 5 * 5 and 7 * 7, and input the pixel patch into the classifier, which is equivalent to the increase of feature dimension. In the original 8-band images, the feature dimensions of patches in 3, 5 and 7 neighborhoods are 72, 200 and 392 respectively. We also apply SVM, RF and AdaBoost to realize classification for image patches. There are 9 classification results based on 3 types of classifiers and 3 types of image patch size.

3.1.3. Classification algorithm

The main idea of SVM [50] is to establish an optimal decision hyperplane to maximize the distance between the nearest two classes of samples on both sides of the plane, to provide good generalization ability for classification problems. RF is a parallel ensemble classification algorithm, but AdaBoost is a serial classifier. The essence of RF is an improvement on the decision tree algorithm, which merges multiple decision trees, and the establishment of each tree depends on the samples extracted independently [51]. The core idea of AdaBoost is to train different weak classifiers using the same training set, and then assemble these weak classifiers to form a stronger final classifier. In this paper, we use the three classifiers (svm.SVC, ensemble.RandomForestClassifier and ensemble.AdaBoostClassifier) provided by the Python sklearn module. Referring to [5, 6, 43, 53], and through theoretical analysis and experiment, the parameters of the three classifiers are shown in table 3.

Table 3. Parameters of classifiers used in our experiments (Python sklearn)

Algorithm	Abbreviation	Parameter Type	Parameter name (sklearn)	Parameter Set
Support Vector Machine	SVM	kernel type	kernel	rbf
		penalty coefficient	C	10
		gamma	gamma	100
Random Forests	RF	base classifier	base_estimator	decision Tree
		number of trees	n_estimators	60
AdaBoost Classifier	AdaBoost	base classifier	base_estimator	decision Tree
		number of trees	n_estimators	60
		learning rate	learning_rate	10^{-3}

3.2. Feature Learning

The original image is input into the CNN model in the form of a three-dimensional pixel patch or a single pixel sequence. For the input layer, the up-sampled seven bands are stacked with the panchromatic band. After the convolution layer and pooling layer of the CNN, the multi-level features of the buildings and non-buildings can be automatically learned. For single-pixel classification, a one-dimensional convolution (CNN_1D) is utilized to learn the spectral features. For image patch classification, the two-dimensional convolution (CNN_2D) is applied to learn spectral and spatial relations simultaneously. The loss function is cross-entropy, and the categories are determined by the softmax layer. We compare the accuracy and efficiency of the one-dimensional temporal convolution on a single pixel with that of the two-dimensional spatial convolution in the neighborhood. We apply Python-Keras module to build the CNN and combine sklearn module to realize classification and accuracy evaluation.

3.2.1. CNN_1D classification

For each pixel, only spectral information is taken into account without considering the spatial relationship between pixels. Within the convolutional layer, there are two Conv1D layers realizing one-dimensional band directional convolution, which is equivalent to complex band operation. Then we use three fully-connected layers. To prevent overfitting, one batch-normalization layer and a dropout layer are added. The output layer consists of a soft max operator, which outputs two categories. In the whole network, we use the popular function called Rectified Linear Unit (ReLU) to solve the vanishing gradient problem for the training epochs in the network.

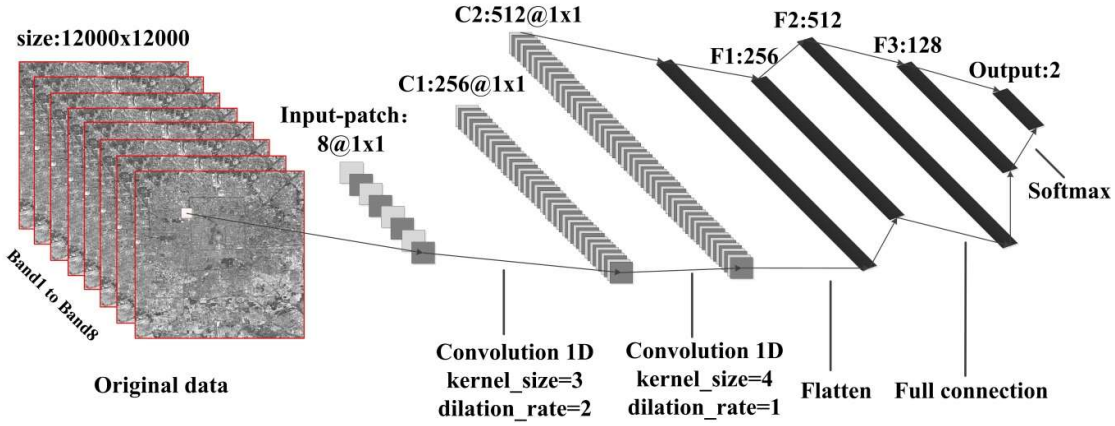

Figure 5. Network structure of CNN_1D

3.2.2. CNN_2D classification

In the 15-meter resolution image, the size of the building is generally less than 5 pixels. For each pixel, the 5-neighborhood is considered, which means that the size of image patch is 5*5*8.

Therefore, an image patch with 8 bands and 5*5 neighborhood centered on each sample is input into the neural network. Within the convolutional layer, there are two Conv2D and two max-pooling layers, which aim to extract spectral features and spatial features, and more high-grade features. In the fully-connected layer, we use three fully-connected layers. Meanwhile, one batch-normalization layer and a dropout layer are added to prevent overfitting. The output layer consists of a soft max operator, which outputs two categories. In the whole network, we also use ReLU as the activation function.

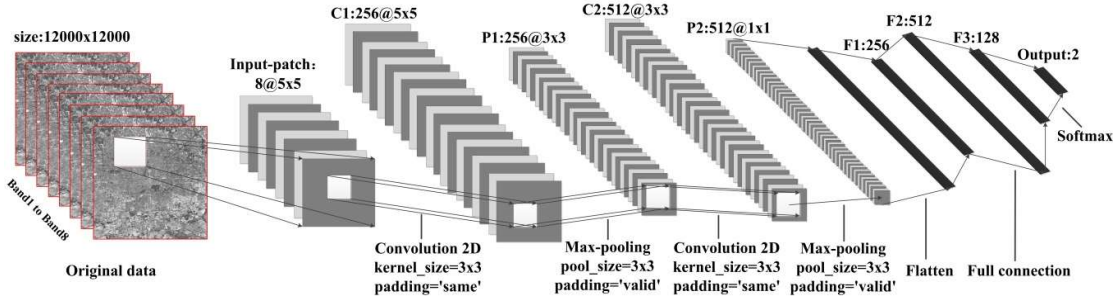

Figure 6. Network structure of CNN_2D

4. Experimental results and evaluation

In total, 26 built-up land cover results were produced. However, considering the large range of research, the large amount of map data, and the reduction of mapping resolution, so we cut out the results of two small areas (A and B) for map display. As shown in figure 7, for feature engineering based on a single pixel, 30 maps were produced using a combination of 5 features (OS8, RSBI, EMBI, PanTex, TEGL) and three classifiers (SVM, RF, AdaBoost). For feature engineering based on image patches, 18 maps were produced, using a combination of 3 kinds of neighbourhoods (P3, P5, P7) and three classifiers (SVM, RF, AdaBoost). For feature learning, there were 4 maps, including the result of one-dimensional convolution on a single pixel and the result of the two-dimensional convolution on an image patch. The main aim of this paper is to compare the accuracy and efficiency of extracting a 15-meter resolution built-up area based on a single pixel and image patch in two cases of feature engineering and feature learning, for completeness, all 52 classified maps for sub-regions A and B were extracted and are provided in figure 7 and figure 9. For the sake of qualitative comparison, we compared the results of all conditions with those of Global-Urban-2015 [48] and GlobalLand30 [44].

A total of 7667 test samples were used for accuracy evaluation. The test samples were classified to obtain the predictive label of each sample, and then the confusion matrix shown in table 4 was obtained according to the real label and the predictive label. Then, the overall accuracy (OA), recall and precision were calculated based on the confusion matrix. OA represents the correctly predicted sample size for all samples. Recall indicates the size of the predicted built-up sample in all true built-up samples. Precision indicates the size of the true built-up sample in all predicted built-up samples. These three precision indices can be used to comprehensively evaluate the accuracy of the built-up area extraction.

Table 4. The representation of confusion matrix for the test samples

		Prediction		
		Non Built-up	**Built-up**	**Sum**
Ground Truth	**Non Built-up**	True Negative (TN)	False Positive (FP)	Actual Negative(TN+FP)
	Built-up	False Negative (FN)	True Positive (TP)	Actual Positive(FN+TP)
	Sum	Predicted Negative(TN+FN)	Predicted Positive(FP+TP)	TN+ TP+ FN+ FP

The *OA*, *Recall* and *Precision* are calculated using Equation (11), Equation (12) and Equation (13) respectively.

$$OA = \frac{TP + TN}{TP + TN + FN + FP} \tag{11}$$

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

where the meanings of *TP* , *TN*, *FN* and *FP* are shown in table 4.

4.1. Feature engineering and feature learning

From the perspective of feature engineering and feature learning, based on feature engineering classification, the classification results are highly correlated with the setting of features, and appropriate features are conducive to improving classification accuracy. However, feature learning does not need to consider manual feature setting. CNN can automatically learn multi-level features from the original image and then achieve classification by black box operation. As shown in figure 7 and figure 9, table 5 and table 6, the classification accuracy based on feature learning is generally better than that based on feature engineering. However, in feature engineering, when the original 8 bands consider the neighborhood and the classifier is RF, the overall accuracy reaches 90%, which is comparable to the results of CNN_2D.

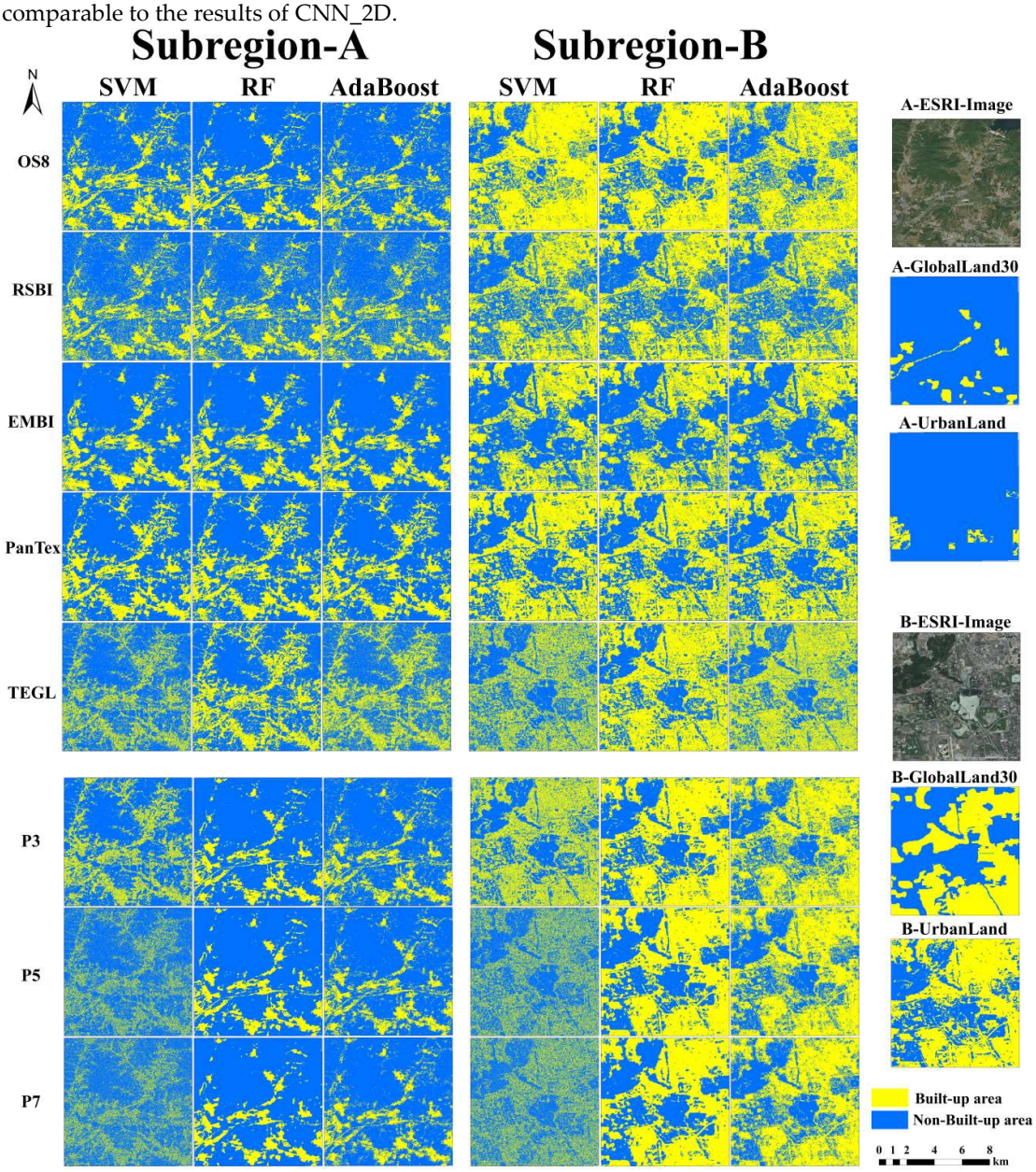

Figure 7. Classification results based on feature engineering

Table 5. Accuracy evaluation based on feature engineering

Feature Engineering	SVM			RF			AdaBoost		
	OA	Recall	Precision	OA	Recall	Precision	OA	Recall	Precision
OS8	0.849	0.922	0.794	0.887	0.904	0.864	0.841	0.834	0.831
RSBI	0.782	0.762	0.774	0.805	0.781	0.802	0.780	0.745	0.781
EMBI	0.810	0.760	0.825	0.810	0.760	0.825	0.809	0.755	0.827
PanTex	0.824	0.825	0.808	0.821	0.820	0.805	0.817	0.811	0.804
TEGL	0.662	0.517	0.693	0.758	0.733	0.752	0.696	0.664	0.686
P3	0.730	0.960	0.644	0.900	0.902	0.889	0.850	0.838	0.841
P5	0.525	0.001	0.429	0.903	0.906	0.891	0.855	0.839	0.853
P7	0.525	0.000	0.000	0.906	0.907	0.896	0.855	0.846	0.848

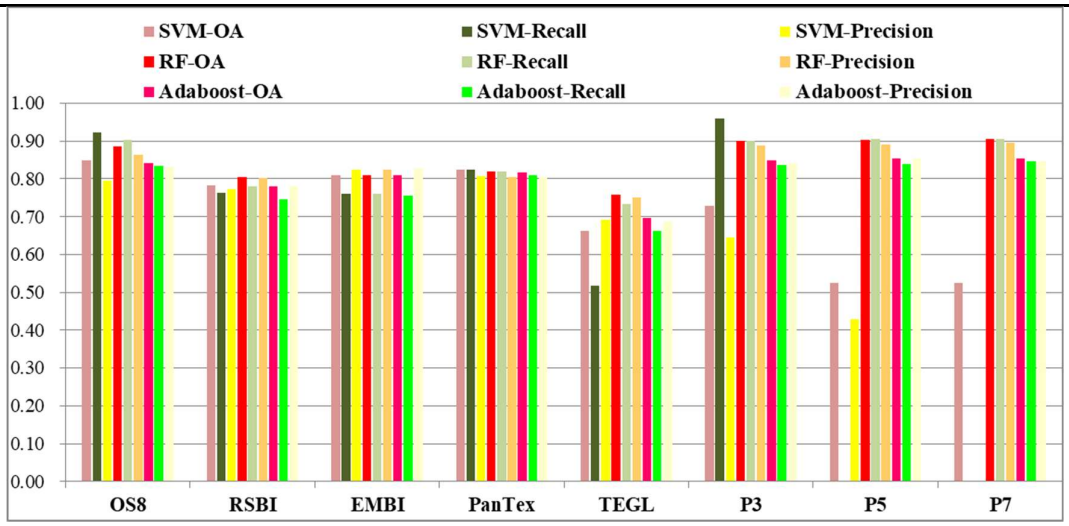

Figure 8. Classification accuracies of built-up area based on Feature Engineering

4.2. Single pixel and image patch

Considering a single pixel and image patch, the classification based on a single pixel only considers the feature vectors of the pixel, ignoring the spatial relationship between pixels in the image spatial plane. As shown in figure 7 and table 5, overall, the classification effect and accuracy based on the image patch are better than those based on a single pixel, but the feature dimension of the image patch is large, and there may be feature redundancy. When training samples are large, a more complex classification model is needed. As shown in figure 8 and table 6, CNN based on the image patch has a significant advantage over CNN based on a single pixel.

Under feature engineering, the accuracy of classifications based on a single pixel is significantly lower than those based on an image patch. Comparing OS8, RSBI, EMBI, PanTex and TEGL, the order of OA and Recall are OS8, PanTex, EMBI, RSBI, TEGL from high to low. The original spectrum (OS8) has the best effect, and the OA of OS8 and PanTex is higher than 82%. The analysis shows that the feature dimension is not necessarily related to the improvement of classification accuracy. Highlighting the characteristics of the target category information can help improve classification accuracy. Combining figure 7 and figure 4, PanTex and EMBI can efficaciously distinguish built-up areas, while RSBI and TEGL cannot reflect buildings well. In particular, the five texture features under TEGL have redundancy and conflict. In the 5-dimensional feature space, it is difficult to learn the appropriate classification boundary, which leads to a poor classification effect.

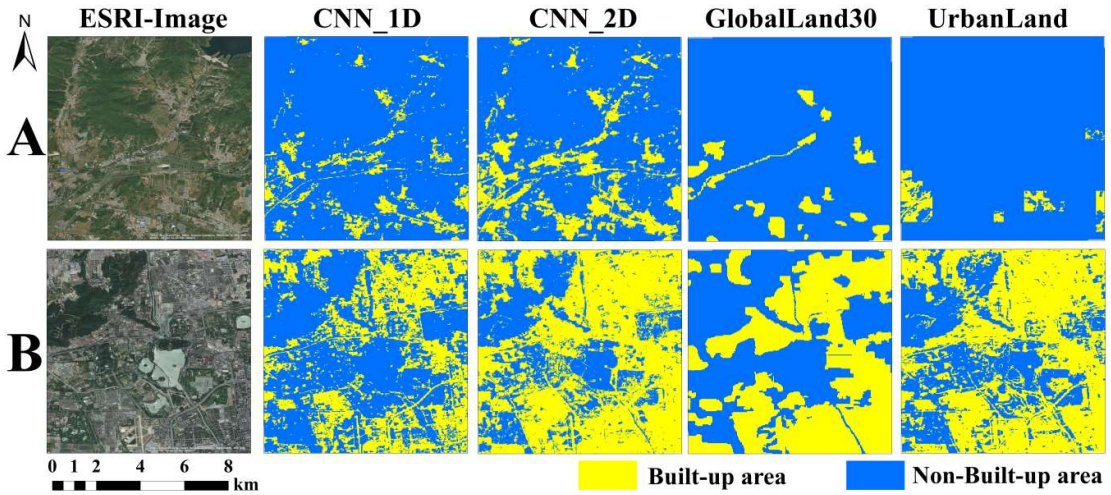

Figure 9. Classification results based on feature learning

As shown in figure 9 and figure 10, the built-up area extraction effect based on CNN_2D is the best. We found many details in the results of CNN_2D, which are missing in the other two productions (GlobalLand30 and Global-Urban-2015). One of the reasons is that the result of CNN_2D is produced from satellite images with higher spatial resolutions. Consequently, within the urban area, non-built-up areas, e.g., water and vegetation in dense buildings, can be discriminated from built-up areas within the Landsat 8 images. Meanwhile, in the suburbs, small built-up areas and narrow roads become distinguishable from the background. Another reason is the higher classification accuracy of the CNN.

Table 6. Accuracy evaluation based on feature learning

CNN model	Training accuracy	OA	Recall	Precision
CNN_1D	0.872	0.823	0.675	0.935
CNN_2D	0.968	0.901	0.915	0.882

4.3. Classification strategy

From the perspective of classification strategy, compared with traditional machine learning algorithms such as SVM, RF and AdaBoost, CNN has the advantages of autonomous learning, stability and robustness. In addition, CNN can learn the dual characteristics of the spectrum and spatial structure in a black box. Users can migrate and use the trained network structure, and only need to focus on input and output. As shown in figure 7 and figure 9, the classification accuracy of P5-RF and P7-RF differs little from that of CNN_2D and is far superior to other classification results. CNN has the structure of batch-normalization and dropout, which can prevent over-fitting. With the increase of the convolutional layer and pooling layer, the network becomes increasingly complex and has stronger fitting and predicting ability. Integrated classifiers (such as RF and AdaBoost), which synthesize the prediction results of all base classifiers and determine the final category by a voting method, can effectively prevent over-fitting, and still have higher classification accuracy when the feature dimension is high. However SVM is more suitable for small sample learning. When the number of samples is too large and the feature dimension is high, most of the training samples are regarded as support vectors, resulting in over-fitting, and the final classification accuracy is very low, even worse than a random guess. The OA of P3-SVM, P5-SVM and P7-SVM were 0.730, 0.525 and 0.525, respectively, and the training accuracy was 1 for all measures. Overfitting occurred obviously, which led to the classification failure.

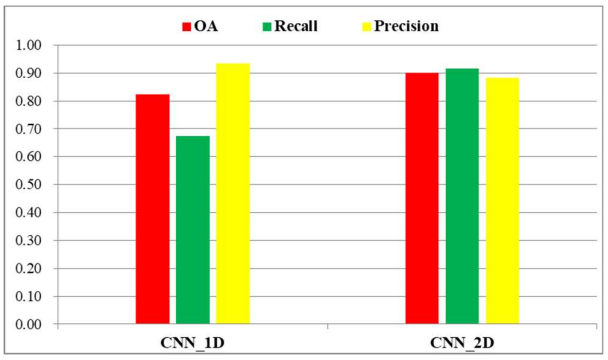

Figure 10. Classification accuracies of the built-up area based on feature learning

5. Discussion

5.1. Support vectors of SVM

As shown in figures 7 and 8, tables 5 and 7, P3-SVM, P5-SVM, and P7-SVM were over-fitted, and the number of support vectors in 11499 training samples was 11431, 11485, and 11493, respectively. Therefore, all training samples are regarded as support vectors, so the model trained is too complex with poor generalization ability and is unable to accurately predict the unclassified data. In the above experimental analysis, the penalty coefficient (C) and Gamma, which are the key parameters, are set to 10 and 100, respectively. The setting of these two parameters is reasonable and scientifically based on prior knowledge and experimental attempts. When the features were OS8, RSBI, EMBI, PanTex and TEGL, the classification results met expectations, and no overfitting was observed. When considering 3, 5, 7 neighborhoods, the number of features is 72, 200, and 392 respectively, the feature dimension increases significantly, and these high-dimensional features have greater correlation and redundancy, resulting in over-fitting in SVM.

Table 7. Number of support vectors of SVM

Classifier	OS8	P3	P5	P7
SVM	9612	11431	11485	11493
SVM-L2	6368	5201	4564	4250

We analyze the principle of the SVM classification algorithm. Overfitting is mainly due to the irregular distribution and clustering of training data in the feature space, resulting in a large number of samples as support vectors, and the classification boundary is very complex. Therefore, we utilize the L2 regularization method provided by Python-sklearn to process the original 8-band data, eliminate the noise and scattering of the data, and then use SVM to classify. The results show that the over-fitting is effectively suppressed. As shown in table 7 and table 8, after L2 regularization, the number of support vectors corresponding to OS8, P3, P5 and P7 was significantly reduced, and the OA were 0.800, 0.832, 0.858 and 0.874, respectively. The classification effect significantly improved, which agreed with the logic and expectations.

Table 8. Accuracy evaluation based on SVM-L2

Feature	Training accuracy	OA	Recall	Precision
OS8	0.799	0.800	0.721	0.835
P3	0.833	0.832	0.769	0.862
P5	0.863	0.858	0.821	0.872
P7	0.881	0.874	0.858	0.874

5.2. CNN_2D versus VGG16

We compare the results of CNN_2D with VGG16 [30]. As shown in figure 11, we reserve the weight of the convolution layers and the pooling layers of VGG16 and reset the top layers including the fully-connected layers, the BatchNormalization layer and the softmax layer. Since the original image has 8 bands and cannot be directly input to VGG16, we fuse panchromatic and multispectral

bands by Gram-Schmidt pan sharpening to obtain the fusion image with 15 m resolution having 7 bands. Then, we consider three bands in two ways: (1) The first three principal components are taken after principal component analysis; (2) the 432 bands representing RGB are taken directly. For VGG16, the channels of input data must be 3, and the size must be greater than 48*48, so the neighborhood of 5*5 is up-sampled to 50*50 by nearest neighbor sampling.

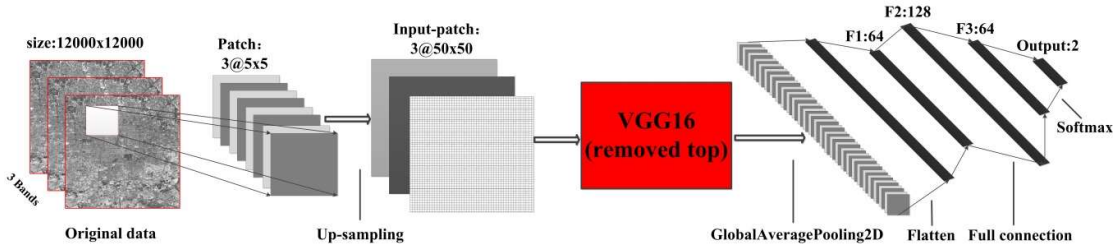

Figure 11. Transfer learning and fine-tuning of VGG16

We set the ratio of training samples and validation samples to 6:4 for training the proposed CNN and VGG16. The accuracy and loss of the training process are shown in figure 12.

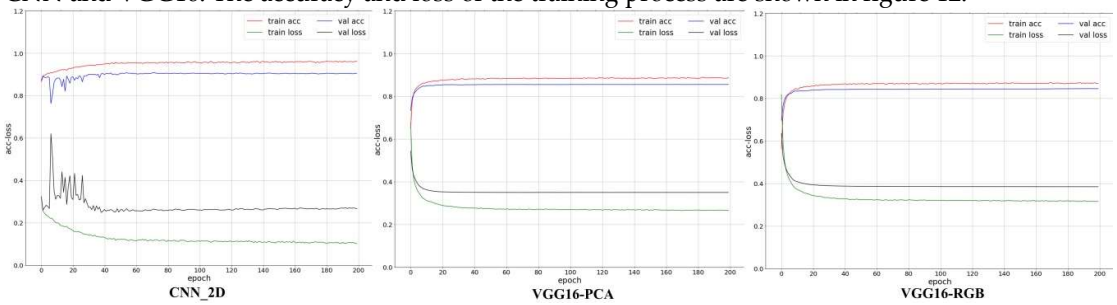

Figure 12. Accuracy and loss in the training process of the proposed CNN_2D, VGG16-PCA and VGG16-RGB

We recorded the training accuracy, training loss, test accuracy, test loss, and the training time of 200 epochs. Table 9 shows that the accuracy of the CNN_2D is significantly better than that of VGG16, and the training time is greatly shortened. In figure 13, the classification effect of CNN_2D is obviously greater than that of VGG16, and the extraction of built-up areas is more detailed and accurate.

Table 9. The accuracy and loss of CNN_2D, VGG16-PCA and VGG16-RGB

CNN-strategy	Training accuracy	OA	Recall	Precision	Training time (s)
CNN_2D	0.968	0.901	0.915	0.882	4000
VGG16-PCA	0.886	0.806	0.782	0.812	36000
VGG16-RGB	0.873	0.790	0.755	0.781	34000

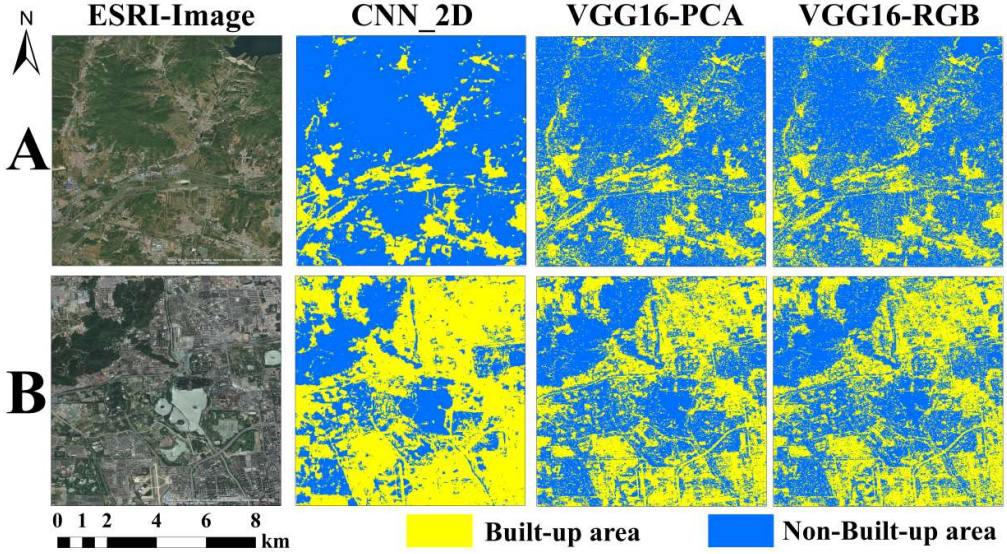

Figure 13. Results of built-up area by the CNN_2D, VGG16-PCA and VGG16-RGB

6. Conclusion

This paper presents a unique investigation to provide a full evaluation of OLI imagery for 15 m-resolution built-up area classification from two viewpoints. First, traditional feature engineering and modern feature learning strategies are compared. Next, the influence of a single pixel and image patch is examined. In contrast, previous studies have generally tended to conduct limited comparisons between, for instance, coarse and fine resolution or pixel- and object-based classification. First, our training samples are automatically selected and filtered based on the existing product data. Then, we make a multi-level and all-around comparison from two different perspectives: single pixel and image patch, feature engineering and feature learning. In feature engineering, we have taken into account the spectrum, morphology, texture and other characteristics. In previous studies, there was no such detailed and comprehensive consideration. Finally, our work is conducted on a relatively large image area, the city of Beijing, China and its environs, ensuring that urban land cover information is generated at a scale of practical value. In contrast, earlier experiments have often been limited to very small, local areas. All tests were evaluated by the same set of test samples with overall accuracy and Kappa coefficient. The results can be summarized as follows:

(1) The classification accuracy based on feature learning is generally better than that based on feature engineering. However in feature engineering, when the original 8 bands consider the neighborhood and the classifier is RF, the overall accuracy reaches 90%, which is comparable to the results of CNN_2D.

(2) The classification effect and accuracy based on the image patch are better than those based on a single pixel, but the feature dimension of the image patch is large, and there may be feature redundancy. When training samples are large, a more complex classification model is needed. CNNs based on image patches have a significant advantage over CNNs based on single pixels. The results of CNN_2D, water and vegetation in dense buildings can be discriminated from built-up area within the Landsat 8 images. Meanwhile, in the suburbs, small built-up areas and narrow roads become distinguishable from the background.

(3) Compared with traditional machine learning algorithms such as SVM, RF and AdaBoost, CNN has the advantages of autonomous learning, stability and robustness. The classification accuracy of P5-RF and P7-RF differs little from that of CNN_2D and is far superior to other classification results. The accuracy of CNN_2D is significantly better than that of VGG16. L2 regularization can eliminate the noise and scattering of the original 8-band data, effectively suppress SVM over-fitting and significantly reduce the number of support vectors.

The research of this paper can be used as a reference for the extraction and mapping of large 15-meter resolution building areas. The comprehensive comparison of classification algorithms can help researchers in remote sensing image pattern recognition to understand the principle and applicability of the algorithm and better carry out scientific research. In this paper, a large number of samples are selected automatically based on existing data products, which is of great significance to improve the efficiency and effectiveness of supervised classification, and can save considerable manpower and time. At the same time, there are some shortcomings to this research, such as not using multi-scale remote sensing data (low, medium and high resolution) for comparative analysis of built-up area extraction, the spatial relationship of the pixels in an image patch is not analyzed in depth, and the hidden layer of CNN is not displayed and analyzed in detail. We will study these problems in follow-up work and hope that more scholars will be involved.

Author Contributions: Conceptualization, H.T.; Funding acquisition, H.T.; Investigation, H.T. and T.Z.; Methodology, T.Z.; Software, T.Z.; Supervision, H.T.

Funding: This work was supported by the National Key R&D Program of China (No. 2017YFB0504104) and the National Natural Science Foundation of China (No. 41571334).

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

References

1. Chen X H, Cao X, Liao A P, et al. Global mapping of artificial surfaces at 30-m resolution. *Science China Earth Sciences*. **2016**, *59(12)*, 2295-2306.
2. Zha, Y., J. Gao and S. Ni. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*. **2003**, *24(3)*, 583-594.
3. Xu H. A new index for delineating built-up land features in satellite imagery. *International Journal of Remote Sensing*. **2008**, *29(14)*, 4269-4276.
4. M Pesaresi, A Gerhardinger, F Kayitakire. A Robust Built-Up Area Presence Index by ani-sotropic Rotation-Invariant Texture Measure. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*. **2009**, *1 (3)*, 180-192.
5. Li C, Wang J, Wang L, et al. Comparison of Classification Algorithms and Training Sample Sizes in Urban Land Classification with Landsat Thematic Mapper Imagery. *Remote Sensing*. **2014**, *6(2)*, 964-983.
6. Momeni, R., P. Aplin and D.S. Boyd, Mapping Complex Urban Land Cover from Spaceborne Imagery: The Influence of Spatial Resolution, Spectral Band Set and Classification Approach. *Remote Sensing*. **2016**, *8(882)*.
7. D Chaudhuri, NK Kushwaha, A Samal, et al. Automatic Building Detection From High-Resolution Satellite Images Based on Morphology and Internal Gray Variance. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*. **2016**, *9 (5)*, 1767-1779.
8. Jin X, Davis C H. Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information. *EURASIP Journal on Advances in Signal Processing*. **2005**, *2005(14),* 2196-2206.
9. Pesaresi M, Guo H, Blaes X, et al. A Global Human Settlement Layer From Optical HR/VHR RS Data: Concept and First Results. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*. **2013**, *6(5)*, 2102-2131.
10. Goldblatt, R., et al. Using Landsat and nighttime lights for supervised pixel-based image classification of urban land cover. *Remote Sensing of Environment*. **2018**, *205*, 253-275
11. Yang, J., et al. A New Method Of Building Extraction From High Resolution Remote Sensing Images Based On NSCT And PCNN, *International Conference on Agro-Geoinformatics*. **2016**, 428-432.
12. Lowe, D.G., Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*. 2004, *60(2)*, 91-110.
13. LeCun, Y., Y. Bengio and G. Hinton, Deep learning. *Nature*. **2015**, *521(7553)*, 436-444.
14. Schmidhuber, J., Deep learning in neural networks: An overview. *Neural Networks*. **2015**, *61*, 85-117.
15. Zhu, X.X., et al., Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*. **2017**, *5(4)*, 8-36.
16. Minar, M.R. and J. Naher. Recent Advances in Deep Learning: An Overview. **2018**.
17. Alom, M.Z., et al., The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches. **2018**.
18. Silver, D., et al., Mastering the game of Go without human knowledge. *Nature*. **2017**, *550(7676)*, 354-+.
19. Li, Y., Deep Reinforcement Learning: An Overview. **2017**.
20. Andreas, J., D. Klein and S. Levine. Modular Multitask Reinforcement Learning with Policy Sketches. **2016**.
21. Anschel, O., N. Baram and N. Shimkin. Averaged-DQN: Variance Reduction and Stabilization for Deep Reinforcement Learning. *International Conference on Machine Learning (ICML)*. **2016**.
22. Arulkumaran, K., et al. A Brief Survey of Deep Reinforcement Learning. **2017**.
23. Babaeizadeh, M., et al. Reinforcement Learning through Asynchronous Advantage Actor-Critic on a GPU. **2016**.
24. Hinton, G.E. and R.R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*. **2006**, *313(5786)*, 504-507.
25. Ackley D H, Hinton G E, Sejnowski T J. A learning algorithm for Boltzmann machines. Connectionist models and their implications: readings from cognitive science. *Ablex Publishing Corp*. **1988**, 147-169.
26. Creswell A, White T, Dumoulin V, et al. Generative Adversarial Networks: An Overview. *IEEE Signal Processing Magazine*, **2017**, *35(1)*:53-65.
27. Zhao W, Du S. Learning multiscale and deep representations for classifying remotely sensed imagery. *Isprs Journal of Photogrammetry & Remote Sensing*. **2016**, *113*,155-165.

28. Han J, Zhang D, Cheng G, et al. Object Detection in Optical Remote Sensing Images Based on Weakly Supervised Learning and High-Level Feature Learning. *IEEE Transactions on Geoscience & Remote Sensing*. **2015**, *53(6)*, 3325-3337.
29. Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. International Conference on Neural Information Processing Systems. *Curran Associates Inc*. **2012**, 1097-1105.
30. Simonyan, K. and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *Computer Science*. **2014**.
31. Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, **2015**, 1-9.
32. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, United States, **2016**, 770-778.
33. Castelluccio M, Poggi G, Sansone C, et al. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. *Acta Ecologica Sinica*. **2015**, *28(2)*, 627-635.
34. Vakalopoulou M, Karantzalos K, Komodakis N, et al. Building detection in very high resolution multispectral data with deep learning features. *Geoscience and Remote Sensing Symposium. IEEE.* **2015**, 1873-1876.
35. Huang Z, Cheng G, Wang H, et al. Building extraction from multi-source remote sensing images via deep deconvolution neural networks. *Geoscience and Remote Sensing Symposium. IEEE*. **2016**, 1835-1838.
36. Makantasis K, Karantzalos K, Doulamis A, et al. Deep Learning-Based Man-Made Object Detection from Hyperspectral Data. *Lecture Notes in Computer Science*. **2015**, *9474(1)*, 717-727.
37. Yang N, Tang H, Sun H, et al. DropBand: A Simple and Effective Method for Promoting the Scene Classification Accuracy of Convolutional Neural Networks for VHR Remote Sensing Imagery. *IEEE Geoscience and Remote Sensing Letters*. **2018**, *5(2)*, 257-261.
38. Wang, L., et al. Urban Built-Up Area Boundary Extraction and Spatial-Temporal Characteristics Based on Land Surface Temperature Retrieval. Remote Sensing. **2018**, 10(4733).
39. Zhong P, Wang R. A Multiple Conditional Random Fields Ensemble Model for Urban Area Detection in Remote Sensing Optical Images. *IEEE Transactions on Geoscience & Remote Sensing*. **2007**, *45(12)*, 3978-3988.
40. Ning, X. and X. Lin. An Index Based on Joint Density of Corners and Line Segments for Built-Up Area Detection from High Resolution Satellite Imagery. *ISPRS International Journal of Geo-Information*. **2017**, 6(33811).
41. Friedl M A, Mciver D K, Hodges J C F, et al. Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*. **2002**, *83(1)*, 287-302.
42. Schaaf C B, Gao F, Strahler A H, et al. First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sensing of Environment*. **2002**, *83(1)*, 135-148.
43. Gong, P., Wang, J., Yu, L., et al. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing*. **2013**, *34(7)*, 2607-2654.
44. Chen J, Chen J, Liao A, et al. Global land cover mapping at 30 m resolution: A POK-based operational approach. *Isprs Journal of Photogrammetry & Remote Sensing*. **2015**, *103*, 7-27.
45. Martino P, Daniele E, Stefano F, et al. Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014, *JRC Technical Report EUR 27741 EN*; doi:10.2788/253582 (online).
46. Gorelick N, Hancher M, Dixon M, et al. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*. **2017**, *202*, 18-27.
47. Liu X, Hu G, Ai B, et al. A Normalized Urban Areas Composite Index (NUACI) Based on Combination of DMSP-OLS and MODIS for Mapping Impervious Surface Area. *Remote Sensing*. **2015**, *7(12)*, 17168-17189.
48. Liu X, Hu G, Chen Y, et al. High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine Platform. *Remote Sensing of Environment*. **2018**, *209*, 227-239.
49. Zhang P, Sun Q, Liu M, et al. A Strategy of Rapid Extraction of Built-Up Area Using Multi-Seasonal Landsat-8 Thermal Infrared Band 10 Images. *Remote Sensing*. **2017**, 9(9).
50. Ma X, Tong X, Liu S, et al. Optimized Sample Selection in SVM Classification by Combining with DMSP-OLS, Landsat NDVI and GlobeLand30 Products for Extracting Urban Built-Up Areas[J]. *Remote Sensing*. **2017**, *9(3)*, 236.

51. Goldblatt R, You W, Hanson G, et al. Detecting the Boundaries of Urban Areas in India: A Dataset for Pixel-Based Image Classification in Google Earth Engine. *Remote Sensing*. **2016**, *8(8)*, 634.
52. Huang X, Zhang L. A Multidirectional and Multiscale Morphological Index for Automatic Building Extraction from Multispectral GeoEye-1 Imagery. *Photogrammetric Engineering & Remote Sensing*. **2011**, *77(7)*, 721-732.
53. D. Lu, Q. Weng. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*. **2007**, *28(5)*, 823-870.