

Airborne Object Detection Using Hyperspectral Imaging: Deep Learning Review

T. T. Pham[1,4(✉)], M. A. Takalkar[1,4], M. Xu[1,4], D. T. Hoang[2], H. A. Truong[3], E. Dutkiewicz[1], and S. Perry[1,4]

[1] Faculty of Engineering and IT, University of Technology Sydney, Ultimo, Australia
thuy.pham@uts.edu.au
[2] Hanoi University of Science and Technology, Hanoi, Vietnam
[3] Vietnam National University, Hanoi, Vietnam
[4] DMTC, Hawthorn, Australia

Abstract. Hyperspectral images have been increasingly important in object detection applications especially in remote sensing scenarios. Machine learning algorithms have become emerging tools for hyperspectral image analysis. The high dimensionality of hyperspectral images and the availability of simulated spectral sample libraries make deep learning an appealing approach. This report reviews recent data processing and object detection methods in the area including hand-crafted and automated feature extraction based on deep learning neural networks. The accuracy performances were compared according to existing reports as well as our own experiments (i.e., re-implementing and testing on new datasets). CNN models provided reliable performance of over 97% detection accuracy across a large set of HSI collections. A wide range of data were used: a rural area (Indian Pines data), an urban area (Pavia University), a wetland region (Botswana), an industrial field (Kennedy Space Center), to a farm site (Salinas). Note that, the Botswana set was not reviewed in recent works, thus high accuracy selected methods were newly compared in this work. A *plain* CNN model was also found to be able to perform comparably to its more complex variants in target detection applications.

Keywords: Hyperspectral imaging · Classification · Remote sensing · Deep learning

1 Introduction

Hyperspectral imaging (HSI) techniques gathers and processes data from across the electromagnetic spectrum. Each pixel in a hyperspectral image is obtained with several spectral bands that can be used for object/material detection. The spatial and spectral properties of specific objects show similarities or differences from one another, thereby allowing the discrimination of different objects in the same perspective based on the image data analysis.

© Springer Nature Switzerland AG 2019
S. Misra et al. (Eds.): ICCSA 2019, LNCS 11619, pp. 306–321, 2019.
https://doi.org/10.1007/978-3-030-24289-3_23

Developing efficient methods to process hyperspectral images with hundreds of channels is often a challenging task due to several factors such as high dimensionality, the lack of training samples, mixed pixels, light-scattering mechanisms during acquisition, nonlinear and complex data due to different atmospheric and geometric distortions [22]. One example, the "Hughes phenomenon" [25] showed the overall mean accuracy is a function of three parameters: measurement complexity, data set size, and the prior probability of the pattern classes. Thus, high dimensionality can influence the accuracy performance. Furthermore, [32]. the position of the sun, imaging angle and direction may cause intra-class differences [1,21,45]. Therefore, besides appropriate classification techniques, data reduction and feature extraction have been found crucial to the accuracy performance of object detection [21,22]. For example, among attempts using support vector machine (SVM) [2,32,46], a more recent work [36] was based on independent component analysis and morphological features. Meanwhile, sparsity-based algorithms [15,16,20] showed that the sparse representation of a pixel can predict the class label of the test sample better than classical SVMs. Recently, deep-learning approaches [14,17,28,35,39–41,43,47,49–52] make use of hierarchically extracted deep features. The framework of [14] was a combination of principle component analysis (PCA), stacked autoencoders architecture, and logistic regression. While most works used convolutional neural networks (CNN) excessively increasing network depth, the more recent learning [43] used a deep feature fusion network that utilised the correlated information among different hierarchical layers, thus more discriminative features.

Existing surveys focussed on challenges of HSI processing as a comprehensive tutorial/overview (e.g., [10,22]). While the recent review [22] covered broad topics including classification, unmixing, dimensionality reduction, resolution enhancement, denoising, change detection, and fast computing, this work analyses the latest methods (since 2014) from the object detection application point of view. The main contributions of this work are:

- We review methods that relate to data pre-processing and object detection algorithms using HSI data.
- We systemically summarise the accuracy performance according to existing individual comparisons.
- We implemented and re-evaluated deep-learning methods using a larger set of popular HSI images while existing comparisons only used one or two datasets in common with each other. Thus, this review includes new test trials for several compared methods.
- Our observations suggest important directions in applying deep-learning approaches to target detection scenarios with HSI data.

2 Data Pre-processing Methods

2.1 Hyperspectral Data Representations

Spectral and spatial information are two fundamental representation types for HSI [27]. When performing spatial and spectral sampling, the information is

sampled at the sensor's spatial and spectral resolution; i.e., a 3D "hypercube" $X \in R^{n_1 \times n_2 \times n_b}$ is obtained, containing $n = n_1 \times n_2$ and n_b bands [10]. In the spectral representation, each pixel is defined in the spectral space $x \in R^{n_b}$. In the spatial representation, each image band is a matrix $X_i \in R^{n_1 \times n_2}$. Because of the high spatial correlation within bands, neighbouring pixels likely represent similar material. Spatial (or *contextual*) information can provide the adjoining pixel relationships and thus may improve the classification accuracy [27].

A combined representation of these two types was called spatial-spectral feature. There are two common strategies to combine: extracting spectral and spatial features separately or directly from sample cubes of the HSI. In spatial-spectral representation, spectral information processing incorporates adjacent relationships of pixels, while spatial information analysis for a single band considers the relationships with other bands [26,41,49].

2.2 Challenges and Pre-processing Techniques

Redundancy Reduction: Because the dimension of HSI data is often large (tens of thousands), this causes a high computational cost in object detection applications, especially in deep learning approaches [14]. On the other hand, the sparseness of HSI data has been demonstrated in earlier works [15,16,20]. Therefore, selecting an appropriate subset of bands was considered an efficient process. Band selection can be done using the highest class-separability criterion [2,19] or information theory-based methods [23]. Other techniques include a kernel method for the selection based on nonlinear dependence between spectral bands and class labels [13] and minimized the error probability using a Bayes classifier [11].

Limited Annotated Training Samples: Collecting HSI data and annotating each pixel are labour-consuming tasks thus the training sets found in existing works were often in limited size. Recently, Li et al. [28] used *pixel pair features* (PPF) to mitigate the problem of training label shortage. PPF was based on *combined pairs* of pixels in a training set. Specifically, two pixels are randomly chosen from a labelled training set, and a subtle rule deduces the label of each *PPF pair* based on the labels of both pixels. If two pixels have same labels, the pair is assigned the same label. If the pixels do not have the same labels, the pair is labelled as "extra", a newly added auxiliary class. The random combination of pixels significantly increases the number of labelled training instances.

3 Object Detection Methods

Detection approaches can be grouped into two directions: "two-step" that consists of a hand-crafted feature extraction step with classification afterwards, and "one-step" (i.e., feature extraction is integrated within the classification model). The former group include conventional methods (e.g., SVMs [2,32,46], sparsity-based [15,16,20]) while the latter was recently introduced in deep learning (DL) models (e.g., [28,35,39,41,43,49–51]).

3.1 Two-Step Approaches

Most traditional feature extraction methods for HSI have been using statistical theory, fuzzy theory, and machine learning [10,22]. For example, linear extraction techniques include PCA [42], minimum noise fraction (MNF) [38], independent component analysis (ICA) [8], and linear discriminant analysis (LDA) [6]. In the spatial domain, existing works used Gabor filter banks [33]. Meanwhile, several approaches used nonlinear transformations such as morphological analysis [9,42], kernel methods [12], and manifold regularisation [3–5,29].

Then, hand-crafted features were used as inputs of conventional classifiers such as the K-nearest neighbour classifier (KNN) [29] and SVM [2,26,37,46]. For instance, Ma et al. [29] deployed a manifold structure from the pixel values then utilised a weighted KNN classifier. In another work, Tuia and Camp-Valls [37] employed SVM with a kernel to train directly from images. Meanwhile, Ji et al. [26] addressed both the pixel spectral and spatial constraints then formulated the relationships among pixels in a hypergraph structure. Their hyperedge is generated by using distance among pixels, where each pixel is connected with its K-nearest neighbours in the feature space.

Recently, deep learning approaches were used for automated feature extraction step from internal layers of a neural network; then these *features* were used as inputs of conventional classifiers. Auto-encoders and deep belief networks were typical examples of this direction [14,24,40]. Auto-encoder (AE) is an unsupervised learning method to learn fewer representations from high dimensional data space and not require annotated training sets [40]. The approach reduced the reconstruction error between the input data at the encoding layer and its reconstruction at the decoding layer [40]. Stacked Auto-encoder (SAE) is a model comprising a number of AE layers with a greedy layer-wise training scheme and a logistic regression layer for classification [14]. Deep belief network (DBN) [24] is a another DL model in which nonlinear description of objects can be analysed. DBN combines the advantages of unsupervised and supervised learning. It is also an automated feature extraction process. DBN was introduced to apply for HSI data using a back-propagation network and a SVM classifier as the final step of classification [48].

3.2 One-Step Approaches

In another approach, DL has also been suggested in one-step solutions (i.e., automatically extract *deep* information from pre-processed HSI data and blindly feed this information into the classification layer of the network). These *deep* features were considered high-level and abstract representations, thus, could be more robust and efficient than lower-level hand-crafted features. Existing works followed this direction include CNN variants [41,50,51], auto-encoders [14,40]. CNN structures were typically designed to process data that come in the form of multiple arrays, e.g., a multispectral image composed of many 2-D arrays containing pixel intensities in the multiple band channels. Due to the properties of natural signals, namely, local connections, shared weights, pooling, and the use

of multiple layers, CNN has been suggested for the HSI classification applications as a *one-step* strategy [17,39,41,43,50,51]. The architecture of a typical CNN is integrated as a series of layers for different assignments including input, convolution, pooling, normalization, *drop-out* and output. Convolutional layers were considered the most important layers that extract features. Specifically, a few first layers provided low-level information such as edges, lines and corners while the latter ones described more abstract features such as structures, objects, and shapes. Typically, after each convolutional layer, there exist pooling layers that are created by computing some local non-linear operation of a particular feature over a region of the image. The process guarantees that the similar outcome can be obtained, even when image features have small translations or rotations, which is essential for scene classification and detection. Then, normalization layers aim to improve generalisation inspired by inhibition schemes presented in the real neurons of the brain. In the last few layers of the network. A *dropout* training method [44] has been recently suggested to reduce over-fitting effects [39,41,43].

4 HSI Datasets

Due to the application-oriented purpose of this survey, we categorised scenarios of target detection scene as follows: rural area, urban, industrial field, and natural reserves. We selected popular public datasets for each scene type as in Table 1. Indian Pines data [7] (DS1) represents a rural area. Pavia University campus site [31] (DS2) represents an urban area. Botswana image [18] (DS3) illustrates a swarm and delta region. Kennedy Space Center HSI [34] (DS4) is an example for an industrial field. Other datasets in Table 1 including Salinas Valley, Houston, and Pavia City share a near similar scene and settings, thus, are briefly compared according to previous results (Table 2). Specific details of datasets tested in this work are depicted as follows.

Rural Area (Dataset 1): This scene shot was taken in 1992 at a 2×2-mile area at 20 m spatial resolution of Northwest Tippecanoe County, Indiana (USA) by AVIRIS sensor (224 spectral reflectance bands ranging 0.4–2.5 μm). This is so-called *Indian Pines data* and are provided by Purdue University [7] as a subset of the original capture. The public part consists of 145×145 pixels and covers agriculture land, forest, and other natural perennial vegetation. In the human living area, the scene contains roads, major dual lane highways, a rail line, low density housing, and other built structures. In the farm field, as in June, the field is in early stages of growth. The manual labels mark 16 object classes (Fig. 1). The image was removed the portion with water absorption, specifically in bands of 104–108, 150–163, and 220 [7]. Thus, there are 200 bands used out of 224 in total of the raw shot.

Table 1. Specifications of publicly available datasets.

Name	Scene type	Location	Size (Pixels)	Spatial resolution (m/pixel)	Spectral band (used/total)
DS1	Rural	Northwestern Indiana, USA. 1992	145 × 145	20	200/224
DS2	Urban	Pavia University Campus, Pavia, Italy	610 × 340	1.3	103/116
DS3	Wetlands	Okavango Delta, Botswana (Africa)	1476 × 256	30	145/242
DS4	Industrial field	Kennedy Space Center, FL, USA	512 × 614	18	176/224
DS5	Farm	Salinas Valley, California, USA	512 × 217	3.7	204/224
Other	Urban	University of Houston campus	–	2.5	114
Other	Urban	Pavia, Italy	1096 × 1096	1.3	102/114

Urban Area (Dataset 2): The Pavia University scene was captured at the University of Pavia campus, Pavia, Italy, from a reflective optics system imaging spectrometer (airborne by the German Aerospace Agency, sponsored by the European Union). The spectrometer has 115 band channels ranging 0.43–0.86 µm. The spatial resolution is 1.3 m per pixel. The publicly portion of the dataset has the size of 640 × 340 pixels provided by Pavia University (Italy) [31]. There are 9 classes of objects in the image: asphalt, meadow, gravel, trees, painted metal sheet, bare soil, bitumen, self-blocking bricks, and shadows (Details of number samples for each object are listed in Fig. 2). Due to noise, only 103 channels were further processed. The image was corrected atmospherically, but not geometrically [31].

Wetland Area (Dataset 3): In 2001–2004, National Aeronautics and Space Administration (NASA) collected from the Earth Observer-1 satellite a sequence of HSI images over the Okavango Delta, Botswana. This scene is one of the world's largest freshwater wetlands (approximately 15,000 km^2). The Hyperion sensor on the satellite has 30 m pixel resolution and 242 bands covering the 400–2500 nm (10 nm windows) [18]. The public data portion was a 7.7 km strip

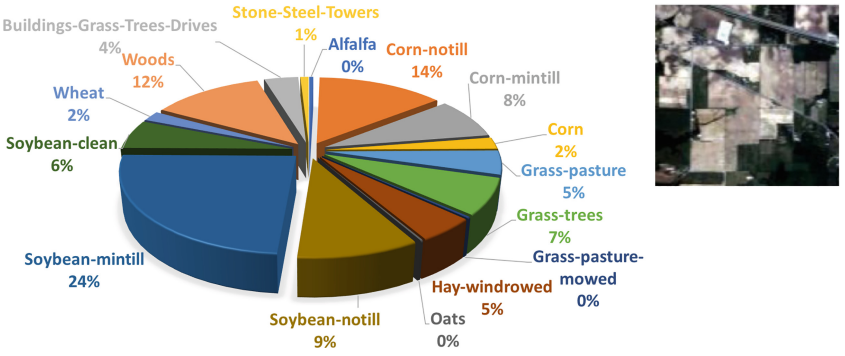

Fig. 1. Objects distribution of the scene and RGB image for Indian Pines.

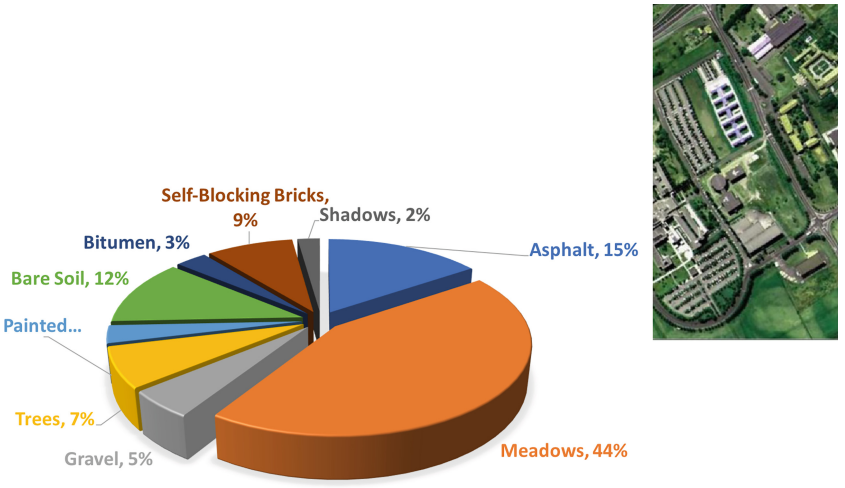

Fig. 2. Objects distribution of the scene and RGB image for Pavia University Campus.

acquired on May 31, 2001 (Fig. 3). The set includes seasonal swamps, and woodlands of the Delta. The data was pre-processed by the provider to reduce the effects of miscalibration, and intermittent anomalies [18]. Noisy bands (due to water absorption) were removed, and the remaining 145 bands were: [10–55, 82–97, 102–119, 134–164, 187–220].

Industrial Field (Dataset 4): Kennedy Space Center (KSC) located on Merritt Island, Florida is one of ten field centres of NASA. The AVIRIS instrument (Airborne Visible/Infrared Imaging Spectrometer) was used to capture KSC site on March 23, 1996. The specifications of the sensor were: 224 bands of 10 nm width; wavelengths from 400–2500 nm [34]. The KSC data, collected from an altitude of approximately 20 km, have a spatial resolution of 18 m (Fig. 4). After

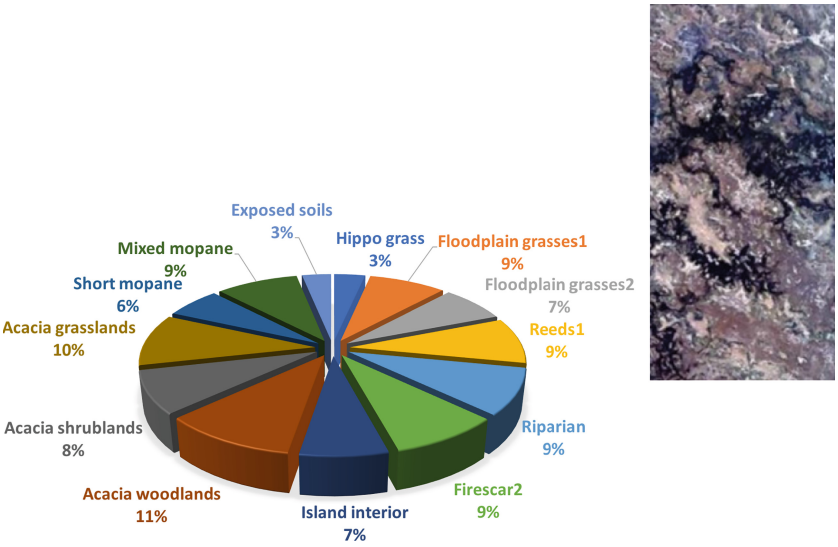

Fig. 3. Objects distribution of the scene and a section of RGB for Botswana delta.

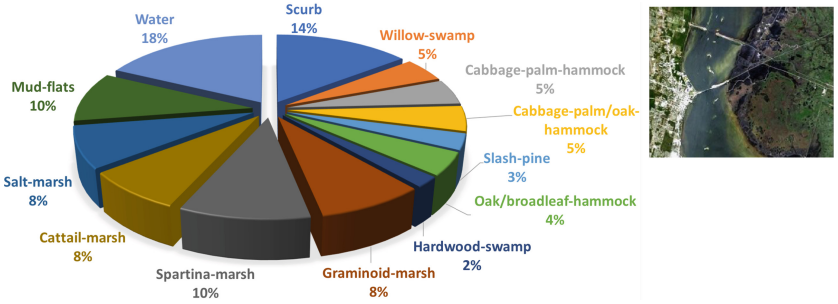

Fig. 4. Objects distribution of the scene and a section of RGB for KSC.

being pre-processed (e.g., water absorption and low SNR bands removed), 176 bands were used further. Class labels of pixels were derived by KSC personnel using colour infrared photography and Landsat Thematic Mapper (TM) imagery.

Farm (Dataset 5): Salinas Valley HSI data was captured by the 224-band AVIRIS sensor over Salinas Valley, California. The spatial resolution was 3.7 m. The data comprises 512×217 samples. Noisy bands with water absorption were [108–112], [154–167], 224. There are 16 objects labelled in this data set including vegetable matter, bare soils, and vineyard fields (Fig. 5).

Other Datasets: There are other popular HSI collections that were used in existing comparisons. Houston campus site image was provided by the University

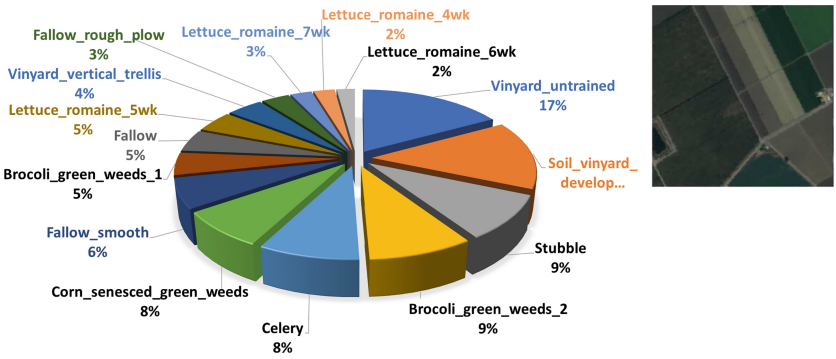

Fig. 5. Objects distribution of the scene and a section of RGB for Salinas.

of Houston in June 2012. The image has 349 × 1,905 pixels (spatial resolution of 2.5 m, 144 spectral bands ranging from 0.38 → 1.05 μm). There are parking lots, highway, railway, tree, soil, water, grass, residential and commercial blocks in the scene. Pavia city data set was collected by the similar instrument with the Pavia University campus. The image was of the center of Pavia, Italy. The data contains 103 bands and 1096 × 1096 samples.

5 Existing Performance Comparisons

From the previously reported results, Table 2 lists latest noticeable works (since 2017) in deep learning neural network approaches and a few of baseline methods. The performance was reported as percentage % for overall accuracy (i.e., averaging for all classes) across several HSI collections. We noticed that most of deep learning works achieved much higher performances against non-deep-learning ones such as random forest and support vector machine based. Moreover, among recent deep learning methods, most of CNN variants performed comparably. However, not all of works used the same experiment datasets.

6 Experiments

6.1 Settings

In response to the aforementioned reviews of literature reports, we carried out our own experiments to re-evaluate a selected groups of works with a broader set of HSI. First, we implemented the method of FDSSC [47] as it was the latest and had the highest accuracy (over 99%). As FDSSC is a CNN variant, we also implemented a plain CNN model as a baseline. We selected a simple CNN in [50] as it was well described and recently reported but was missed in state-of-art comparisons. Then we deployed the SAE-LR [14] approach because it is an unsupervised learning type of detection while CNNs are supervised ones.

Table 2. Comparisons of accuracy performance across public HSI images as reported in existing works. Overall accuracy performance is calculated for all classes (%) in each dataset used or '-' if not used. IndianP: DS1. PaviaU: DS2. KSC: DS3. PaviaC: Pavia City data. Houston: Houston University campus (Sect. 4).

Methods	IndianP	PaviaU	PaviaC	KSC	Salinas	Houston	Reported work (reference, year)
SAE-LR [14]	-	-	98.52	98.76	-	-	Proposed vs. RBF-SVM, EMP RBF-SVM ([14], 2014)
DC-CNNaug [51]	98.76	99.68	-	-	-	-	Proposed vs. DC-CNN, CNN, SSDL ([51], 2017)
CNN [50]	64.19	67.85	-	-	85.24	-	Proposed vs. SVM, KNN ([50], 2017)
RNN-GRU-PRetanh [35]	88.63	88.85	-	-	-	89.85	Proposed vs. RNN-LSTM, RNN GRU-tanh, CNN, SVM-RBF, RF 200 trees ([35], 2017)
RNN-GRU-tanh [35]	85.71	80.70	-	-	-	85.73	Baseline ([35], 2017)
RNN LSTM [35]	80.52	77.99	-	-	-	85.41	Baseline ([35], 2017)
CNN [35]	84.18	80.51	-	-	-	85.42	Baseline ([35], 2017)
SVM-RBF [35]	72.78	78.82	-	-	-	77.09	Baseline ([35], 2017)
RF-200 trees [35]	69.79	71.37	-	-	-	72.93	Baseline ([35], 2017)
FDSSC [47]	99.75	99.97	-	99.96	-	-	Proposed vs. CNN, SAE-LR, 3D-CNN-LR, SSRN ([47], 2018)
CNN [30]	95.96	99.38	-	99.31	-	-	Baseline ([47], 2018)
SAE-LR [14]	96.53	98.46	-	92.99	-	-	Baseline ([47], 2018)
DFFN [43]	98.52	98.73	-	-	98.87	-	Proposed vs. DRNN, DCNN, RVCANet, SVM ([43], 2018)
DRN [43]	98.36	98.52	-	-	98.48	-	baseline ([43], 2018)
DCNN [43]	97.93	97.19	-	-	95.05	-	baseline ([43], 2018)

SAE-LR: Stacked Autoencoder with Logistic Regression. RF: Random Forest. SVM-RBF: Support Vector Machines - Radial Basis Function. EMP: Extended Morphological Profiles CNN: Convolutional Neural Network. DC-CNNaug: Dual Channel CNN + Augmentation. FDSSC: Fast Dense Spectral–Spatial CNN DFFN: Deep feature fusion network RNN: Recurrent Neural Network. DRN: Deep residual CNN. DCNN: Deep plain CNN.

From Table 2, the SAE-LR was reported to have high accuracy but did not perform consistently in later works as a baseline (e.g., in the report of [47]).

These three methods were re-evaluated on Indian Pines data [7] for a rural area; Pavia University campus site [31] for an urban area; Botswana image [18] for a wetland region; Kennedy Space Center [34] for an industrial field; and Salinas data for farming areas (Sect. 4). Note that, the Botswana data was not reported recently (Table 2) thus it is newly tested for all of three implemented methods. For each dataset, we repeated and recorded the accuracy performance

metric 10 times. Therefore, the experiment results were listed as mean and standard deviation format in Table 3.

Our experiments were executed on two separate computing systems: a server of Ubuntu (16.04 LTS, Intel Core i5-5200U processor, 4 cores of 2.20 GHz, Nvidia GeForce 940M GPU) and a Windows 10 (Home 64-bit; Intel(R) Core(TM) i7 NVIDIA GeForce GTX 1070 GPU). Algorithms were implemented in Python 3.6.6; Tensorflow (GPU version) 1.12.0 Keras 2.2.4. We used PyTorch environment to more quickly run models for various HSI inputs. We utilised a newer optimizer to optimize the loss function than the ones in earlier works. Specifically, we used the Adam optimizer with 1000 epochs. The optimization is stopped when the loss value does not decrease any more.

We split the training, validation, and test sets as the ratio of 3 : 1 : 6 consistently across five datasets for the experiments with the first two works: SAE-LR [14] and CNN [50]. Note that this splitting might be different from the original settings of the CNN work [50]. They suggested 15 training samples per classes after varying the size in a range from 3 to 15. For the work of FDSSC [47], we kept the same settings as it was reported for the sake of reproducibility (i.e., 2 : 1 : 7 for Indian Pines and KSC; 10 : 5 : 85 for Pavia University; and 2 : 1 : 7 for new tests).

6.2 Experimental Results

Table 3 summarises our experimental results for deep learning methods across public HSI images. Overall accuracy performance is calculated for all classes (% mean ± standard deviation). Three of methods achieved very high accuracy levels for all classes. The auto-encoder approach performed dramatically up on each individual dataset. For instance, the model reached 98% of accuracy for the Pavia University, Salinas, and Botswana data but yielded only 58% for the KSC collection. Meanwhile, the two CNN variants performed consistently over 97% of overall accuracy.

Table 3. Our test results of emerging models across public HSI images. Accuracy performance is calculated for all classes (% mean ± standard deviation).

Methods (reference, year)	IndianP	PaviaU	Salinas	KSC	Botswana
SAE-LR [14] 2014	84.65 ± 2.70^{a}	98.32 ± 0.09^{a}	97.79 ± 0.11^{a}	57.29 ± 1.23^{a}	97.40 ± 0.18
CNN [50] 2017	99.14 ± 0.23	97.86 ± 0.19	96.50 ± 0.21	97.02 ± 0.30^{a}	99.85 ± 0.15^{a}
FDSSC [47] 2018	99.77 ± 0.11	99.98 ± 0.01	99.95 ± 0.03^{a}	99.96 ± 0.07	99.78 ± 0.24^{a}

[a]: Newly tested datasets comparing with earlier works.
IndianP: DS1. PaviaU: DS2. KSC: DS3. Botswana: DS4. Salinas (Sect. 4).
SAE-LR: Stacked Autoencoder with Logistic Regression [14]. CNN: Convolutional Neural Network. FDSSC: Fast Dense Spectral–Spatial CNN [47].

7 Conclusion

In this report, hyperspectral image analysis was reviewed particularly for object detection applications. Several HSI data representations were summarised from spectral/spatial characteristics to hand-crafted or automated feature extraction using deep-learning methods.

In terms of data sources, we reported publicly available datasets commonly used for HSI analysis. We found limited research on off-nadir data for our target detection focus but some available in other areas of HSI applications such as material discrimination [45], atmospheric compensation [1,2]. We recommend that future works should investigate this scenario of data collection to expand further and more efficient HSI-based target detection applications.

Regarding to technical approaches for HSI object detection, we described both hand-crafted feature extraction in-cooperated conventional methods (e.g., support vector machine or decision trees) and automated feature extraction using deep learning neural networks. According to existing comparisons (summarised in Table 2), deep learning neural networks performed much better than the conventional ones. Unsupervised features (e.g., intermediate information from auto-encoders (SAE-LR [14])) may act as mid-level representations and hence provide more semantic features and more robust detection accuracy than the low-level ones (i.e., hand-crafted).

8 Discussion

Based on our observations for a selected group of works, we found CNN models provided comparatively reliable performance of over 97% detection accuracy across a large set of HSI collections. We included data for a rural area (Indian Pines data [7]), an urban area (Pavia University campus site [31]), a wetland region (Botswana image [18]) an industrial field (Kennedy Space Center [34]), and a farm site (Salinas data, Sect. 4). Note that, the Botswana data was not reported recently (Table 2) thus it is newly tested for all of three implemented methods.

According to our experimental experience, we confirmed the non-consistent performance of SAE-LR [14] found in earlier comparisons (Tables 2 and 3). This is probably due to the fact that the algorithm depends heavily on the reconstruction error between the input data at the encoding layer and its reconstruction at the decoding layer which in turns relates much on the quality of each dataset. We hypothesize that with a robust pre-processing procedure, the SAE-LR can work well for unsupervised HSI data. It should be noted that we used the Adam optimizer when deploying the work of SAE-LR [14] that used a basic optimizer of stochastic gradient descent thus it took the original experiments four to six hours to complete a trial. Furthermore, we observed CNN models detected objects comparably despite of a *plain* design (i.e., without using residual learning and feature fusion) or a more complex architecture in recent reports except for one report of a simple CNN [50]. The authors of [50] only yielded less than 67% for

Indian Pines and Pavia University sets (Table 2). Therefore, we re-evaluated this design in our experiments and found it actually worked well (above 97% across all five datasets). Hence, we suggest that a simple CNN model can perform well for target detection rather than complex variants.

Acknowledgment. This paper includes research that was supported by DMTC Limited (Australia). The authors have prepared this paper in accordance with the intellectual property rights granted to partners from the original DMTC project.

References

1. Adler-Golden, S.M., Bernstein, L.S., Matthew, M.W., Sundberg, R.L., Ratkowski, A.J.: Atmospheric compensation of extreme off-nadir hyperspectral imagery from hyperion. In: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIII, vol. 6565, p. 65651P. International Society for Optics and Photonics (2007)
2. Archibald, R., Fann, G.: Feature selection and classification of hyperspectral images with support vector machines. IEEE Geosci. Remote Sens. Lett. **4**(4), 674–677 (2007)
3. Bachmann, C.M., Ainsworth, T.L., Fusina, R.A.: Exploiting manifold geometry in hyperspectral imagery. IEEE Trans. Geosci. Remote Sens. **43**(3), 441–454 (2005)
4. Bachmann, C.M., et al.: Bathymetric retrieval from hyperspectral imagery using manifold coordinate representations. IEEE Trans. Geosci. Remote Sens. **47**(3), 884 (2009)
5. Bachmann, C.M., Ainsworth, T.L., Fusina, R.A., Topping, R., Gates, T.: Manifold coordinate representations of hyperspectral imagery: improvements in algorithm performance and computational efficiency. In: 2010 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 4244–4247. IEEE (2010)
6. Bandos, T.V., Bruzzone, L., Camps-Valls, G.: Classification of hyperspectral images with regularized linear discriminant analysis. IEEE Trans. Geosci. Remote Sens. **47**(3), 862–873 (2009)
7. Baumgardner, M.F., Biehl, L.L., Landgrebe, D.A.: 220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3, September 2015. https://doi.org/10.4231/R7RX991C, https://purr.purdue.edu/publications/1947/1
8. Bayliss, J.D., Gualtieri, J.A., Cromp, R.F.: Analyzing hyperspectral data with independent component analysis. In: 26th AIPR Workshop: Exploiting New Image Sources and Sensors, vol. 3240, pp. 133–144. International Society for Optics and Photonics (1998)
9. Benediktsson, J.A., Palmason, J.A., Sveinsson, J.R.: Classification of hyperspectral data from urban areas based on extended morphological profiles. IEEE Trans. Geosci. Remote Sens. **43**(3), 480–491 (2005)
10. Bioucas-Dias, J.M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., Chanussot, J.: Hyperspectral remote sensing data analysis and future challenges. IEEE Geosci. Remote Sens. Mag. **1**(2), 6–36 (2013)
11. Bruzzone, L., Serpico, S.B.: A technique for feature selection in multiclass problems. Int. J. Remote Sens. **21**(3), 549–563 (2000)
12. Camps-Valls, G., Bruzzone, L.: Kernel-based methods for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. **43**(6), 1351–1362 (2005)

13. Camps-Valls, G., Mooij, J., Scholkopf, B.: Remote sensing feature selection by kernel dependence measures. IEEE Trans. Geosci. Remote Sens. **7**(3), 587–591 (2010)
14. Chen, Y., Lin, Z., Zhao, X., Wang, G., Gu, Y.: Deep learning-based classification of hyperspectral data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **7**(6), 2094–2107 (2014)
15. Chen, Y., Nasrabadi, N.M., Tran, T.D.: Hyperspectral image classification using dictionary-based sparse representation. IEEE Trans. Geosci. Remote Sens. **49**, 3973–3985 (2011)
16. Chen, Y., Nasrabadi, N.M., Tran, T.D.: Hyperspectral image classification via kernel sparse representation. IEEE Trans. Geosci. Remote Sens. **51**(1), 217–231 (2013)
17. Chen, Y., Jiang, H., Li, C., Jia, X., Ghamisi, P.: Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Trans. Geosci. Remote Sens. **54**(10), 6232–6251 (2016)
18. Crawford, M.M., Ghosh, J.: Random forests of binary hierarchical classifiers for analysis of hyperspectral data. In: IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data 2003, pp. 337–345 (2003)
19. De Backer, S., Kempeneers, P., Debruyn, W., Scheunders, P.: A band selection technique for spectral classification. IEEE Geosci. Remote Sens. Lett. **2**(3), 319–323 (2005)
20. Fotiadou, K., Tsagkatakis, G., Tsakalides, P.: Spectral super-resolution for hyperspectral images via sparse representations. In: Living Planet Symposium, vol. 740, p. 417 (2016)
21. Gewali, U.B., Monteiro, S.T., Saber, E.: Machine learning based hyperspectral image analysis: a survey. arXiv preprint arXiv:1802.08701 (2018)
22. Ghamisi, P., et al.: Advances in hyperspectral image and signal processing: a comprehensive overview of the state of the art. IEEE Geosci. Remote Sens. Mag. **5**(4), 37–78 (2017)
23. Guo, B., Gunn, S.R., Damper, R.I., Nelson, J.D.: Band selection for hyperspectral image classification using mutual information. IEEE Geosci. Remote Sens. Lett. **3**(4), 522–526 (2006)
24. Hinton, G.E., Osindero, S., Teh, Y.W.: A fast learning algorithm for deep belief nets. Neural Comput. **18**(7), 1527–1554 (2006)
25. Hughes, G.: On the mean accuracy of statistical pattern recognizers. IEEE Trans. Inf. Theor. **14**(1), 55–63 (1968)
26. Ji, R., Gao, Y., Hong, R., Liu, Q., Tao, D., Li, X.: Spectral-spatial constraint hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. **52**(3), 1811–1824 (2014)
27. Landgrebe, D.A.: Signal theory methods in multispectral remote sensing, vol. 29. Wiley, Hoboken (2005)
28. Li, W., Wu, G., Zhang, F., Du, Q.: Hyperspectral image classification using deep pixel-pair features. IEEE Trans. Geosci. Remote Sens. **55**(2), 844–853 (2017)
29. Ma, L., Crawford, M.M., Tian, J.: Local manifold learning-based k-nearest-neighbor for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. **48**(11), 4099–4109 (2010)
30. Makantasis, K., Karantzalos, K., Doulamis, A., Doulamis, N.: Deep supervised learning for hyperspectral data classification through convolutional neural networks. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 4959–4962, July 2015. https://doi.org/10.1109/IGARSS.2015.7326945

31. Marinoni, A., Gamba, P.: A novel approach for efficient p-linear hyperspectral unmixing. IEEE J. Sel. Top. Signal Process. **9**(6), 1156–1168 (2015)
32. Melgani, F., Bruzzone, L.: Classification of hyperspectral remote sensing images with support vector machines. IEEE Trans. Geosci. Remote Sens. **42**(8), 1778–1790 (2004)
33. Mirzapour, F., Ghassemian, H.: Improving hyperspectral image classification by combining spectral, texture, and shape features. Int. J. Remote Sens. **36**(4), 1070–1096 (2015)
34. Morgan, J.: Adaptive hierarchical classifier with limited training data. Ph.D. thesis, Department of Mechanical Engineering, University of Texas at Austin (2002)
35. Mou, L., Ghamisi, P., Zhu, X.X.: Deep recurrent neural networks for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. **55**(7), 3639–3655 (2017)
36. Mura, M.D., Villa, A., Benediktsson, J.A., Chanussot, J., Bruzzone, L.: Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis. IEEE Geosci. Remote Sens. Lett. **8**, 542–546 (2011)
37. Na, L., Wunian, Y.: Hyperspectral remote sensing image feature extraction based on kernel minimum noise fraction transformation. Remote Sens. Technol. Appl. **2**, 013 (2013)
38. Nielsen, A.A.: Kernel maximum autocorrelation factor and minimum noise fraction transformations. IEEE Trans. Image Process. **20**(3), 612–624 (2011)
39. Nogueira, K., Penatti, O.A., dos Santos, J.A.: Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recogn. **61**, 539–556 (2017)
40. Özdemir, A.O.B., Gedik, B.E., Çetin, C.Y.Y.: Hyperspectral classification using stacked autoencoders with deep learning. In: 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), pp. 1–4. IEEE (2014)
41. Paoletti, M., Haut, J., Plaza, J., Plaza, A.: A new deep convolutional neural network for fast hyperspectral image classification. ISPRS J. Photogrammetry Remote Sens. (2017)
42. Plaza, A., Martinez, P., Plaza, J., Perez, R.: Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations. IEEE Trans. Geosci. Remote Sens. **43**(3), 466–479 (2005)
43. Song, W., Li, S., Fang, L., Lu, T.: Hyperspectral image classification with deep feature fusion network. IEEE Trans. Geosci. Remote Sens. **56**(6), 3173–3184 (2018)
44. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)
45. Suen, P.H., Healey, G., Slater, D.: The impact of viewing geometry on material discriminability in hyperspectral images. IEEE Trans. Geosci. Remote Sens. **39**(7), 1352–1359 (2001)
46. Tarabalka, Y., Fauvel, M., Chanussot, J., Benediktsson, J.A.: SVM- and MRF-based method for accurate classification of hyperspectral images. IEEE Geosci. Remote Sens. Lett. **7**, 736–740 (2010)
47. Wang, W., Dou, S., Jiang, Z., Sun, L.: A fast dense spectral-spatial convolution network framework for hyperspectral images classification. Remote Sens. **10**(7), 1068 (2018)
48. Xinhua, J., Heru, X., Lina, Z., Yanqing, Z.: Hyperspectral data feature extraction using deep belief network. Int. J. Smart Sens. Intell. Syst. **9**(4) (2016)

49. Yang, X., Ye, Y., Li, X., Lau, R.Y., Zhang, X., Huang, X.: Hyperspectral image classification with deep learning models. IEEE Trans. Geosci. Remote Sens. (2018)
50. Yu, S., Jia, S., Xu, C.: Convolutional neural networks for hyperspectral image classification. Neurocomputing **219**, 88–98 (2017)
51. Zhang, H., Li, Y., Zhang, Y., Shen, Q.: Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network. Remote Sens. Lett. **8**(5), 438–447 (2017)
52. Zhong, P., Gong, Z., Schnlieb, C.: A diversified deep belief network for hyperspectral image classification. In: ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLI-B7, pp. 443–449 (2016)