School of Informatics

Informatics Research Review
Comparison of Traditional Machine Learning and Deep Learning Algorithm in Supervised Classification of Remote Sensing Images

[redacted]

January 2021

Abstract

Remote sensing has provided its values in many fields. The supervised land use and land cover (LULC) classification is an essential application in remote sensing. Regarding imagery resolution, the traditional machine learnings represented by Random Forest and SVM are optimal for supervised LULC classification using medium-resolution imagery, and the deep learning represented by CNN is more suitable when using high-spatial-resolution imagery as the data source. To solve the poor performance of CNN on medium-resolution images, a novel improved CNN algorithm was proposed. Despite the improved performance, the improved CNN has not reached the level that can be used in real practices.

Date: Friday 29^{th} January, 2021

Supervisor: [redacted]

Contents

1 Introduction

Remote sensing is a mainstream method for the Earth's surface monitoring, and its popular applications include agriculture, environmental science, mineralogy, urban, ocean, and lakes (Kellenberger et al., 2018). Remote sensing is the acquisition of an object's image without any physical connections and extracts useful information from the images. Actually, remote sensing is similar to computer vision, except remote sensing includes more information on wavebands. Remote sensing has been utilized in many applications. Land use and land cover (LULC) classification is an essential field in remote sensing, and it is fundamental for other advanced remote sensing applications (Ma et al., 2019). Besides, remote sensing image analysis tasks also include scene classification, image fusion, image registration, object detection, etc. (Lary et al., 2016).

Supervised classifiers are more robust than model-based methods, so they are widely used in the recent two decades (Niemeyer et al., 2014). The supervised classification utilizes the training data to learn about the characteristics extracted from target data and classify the unknown data according to identifying the learned characteristics (Belgiu and Drăguţ, 2016). Since remote sensing are usually used to identify and classify specific targets, supervised classification is more common than unsupervised classifications in the field of remote sensing. Therefore, this review will specifically focus on LULC supervised classification algorithms. Since 2000, land use and land cover supervised classification of remote sensing are usually generated by traditional machine learning because machine learning an effective empirical approach for supervised classification of nonlinear systems (Lary et al., 2016). The commonly supervised methodologies of machine learning for many geoscience applications are support vector machines (SVM), decision trees (DT), and its ensemble methods, such as random forests (RF), genetic algorithm (GA), etc. (Shahin et al., 2001; Shahin and Jaksa, 2005; Azamathulla and Wu, 2011). Among many machine learning algorithms, SVM and random forest (RF) are the most commonly used ML methods (Lary et al., 2016; Mountrakis et al., 2011), since they were considered to outperform decision trees, the genetic algorithm (GA), ANN classifier, etc. in terms of classification accuracy (Belgiu and Drăguţ, 2016).

The remote sensing data with a high spatial resolution is steadily becoming popular, as the advances in remote sensing data acquisition technologies (Belward and Skøien, 2015). Since 2016, commercial satellites with a spatial resolution of 1-2 meters and even sub-meter levels have become common. Compared with remote sensing images with a spatial-resolution of more than 10 meters, finer spatial resolution brings challenges to machine learning represented by FR and SVM, resulting in significantly reduced classification accuracy. In recent years, researchers pay attention to deep learning and utilize deep learning algorithms to classify LULC with high spatial resolutions as deep learning becomes increasingly popular, having made significant progress (Ma et al., 2019).

SVM and RF show an outstanding performance in low or median spatial resolution, but their performance reduces dramatically for high-spatial-resolution classification (Ma et al., 2017; Liu et al., 2017). Fortunately, deep learning, represented by CNN, effectively improves the classification accuracy of high spatial resolution images. However, it is interesting that deep learning does not perform as well as traditional machine learning in low or medium resolution remote sensing images, even if it requires massive labeled samples and computational cost, which are much more than machine learning needs.

In this paper, the performances of RF/SVM and CNN for supervised LULC classification will be compared and discussed. Compared with unsupervised or semi-supervised classification, supervised classification is commonly used by remote sensing community, and land-use-and-land-cover is also the most frequent sub-area in remote sensing. Hence, there is a wealth of high-quality articles in this field. Moreover, most papers reviewed by this paper are peer-reviewed articles to ensure the high-quality and reliability. Section 2.1 reviews performances of machine learning represented RF and SVM on supervised LULC classification. Section 2.2 reviews the performance of CNN. Section 2.3 reviews the performance of an improved CNN when using medium-resolution imagery for classification. Lastly, section 3 gives the summary and conclusion of the literature review.

2 Literature Review

2.1 Classification by Machine Learning – RF and SVM

2.1.1 Support Vector Machine

In the past 20 years, a wide range of methods were used to be analyzed remote sensing imagery. Since support vector machines can generalize well even with limited training samples which is a general limitation of remote sensing applications, they are particularly attractive in the field of remote sensing (Mountrakis et al., 2011; Chi et al., 2008). There is limited amount of training data is often provided for remote sensing, especially for applications in the real world. The labeling of the training dataset relies on manual labor, which is very time-consuming and costly.

Support vector machine (SVM) is a supervised non-parametric statistical machine learning. This method is presented with a set of labeled data instances. In remote sensing classification, an instance of a data sample to be labeled is usually a single pixel. Each pixel is represented as a vector, and each image band is considered as a dimension. The SVM training algorithm aims to find a hyperplane that can divide the training data into two discrete classes. Because typical remote sensing classification is a multiple classes problem (usually more than two classes), binary SVM must be made adjustments. Firstly, a multi-classes classifier assigns data samples into one category or the other and then assigns the rest of the samples into two categories until all categories are finished. Then, predicted data are predicted to belong to a category based on which sides of hyperplanes they fall in (Liu et al., 2017). Figure 1 illustrates a simple case with a two-classes separable classification in a two-dimensional space.

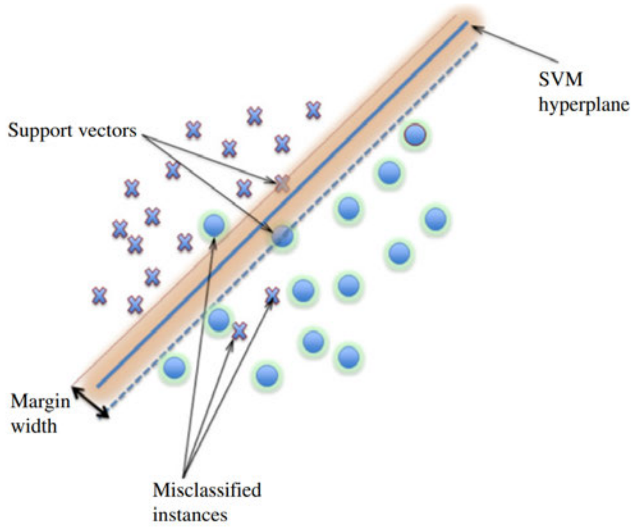

Figure 1: Linear support vector machine in two-dimensional space (Source: adapted from (Burges, 1998))

In practice, it is difficult for linear SVM to classify remote sensing patterns with high accuracy since remote sensing data points of different class clusters often overlap on one another, so that the basic linear decision boundaries have linear separability difficult (Mountrakis et al., 2011). So, when it comes to remote sensing, the SVM usually refers to the kernel methods. The kernel aims to solve the inseparability issue on SVM by using additional variables in SVM optimization and mapping nonlinear hyperplanes into vector space. A kernel SVM used to solve optimization issue usually is written as:

$$MAX_{\alpha}\left\{\sum_{i\,=\,1}^{n}\alpha_i\;+\sum_{i,\;\;j\,=\,1}^{n}\alpha_i\alpha_j y_i y_j K(x_i x_j)\right\}$$
$$s.t.\;\;0\;\leq\;\alpha_i\;\leq\;C$$
$$\sum_{i=1}^{n}\alpha_i y_i = 0$$

where (xi, yi) are s set of given training instance; K represents the kernel function K; α_i are the Lagrange coefficients; and C is a penalty consistent used to penalize errors of training instances.

In the earlier research on medium resolution imagery (15–30 m spatial resolution), Huang et al. (2002) compared the LULC classification accuracy of SVM and three other classifiers - MLC, neural network classifier (NN), and decision trees. SVM results showed the highest accuracy, and the author thought its high classification accuracy benefits from its ability to locate the best separating hyperplane. The overall accuracy of SVM for LULC classification has reached 84%. Only need some accurate training data, the position of the hyperplane can be identified. Twenty years ago, there were only few open-source labeled data sets for scientific research in remote sensing. Moving towards high-spatial-resolution imagery (1-2 meters), Tuia et al. (2009) proposed an improved SVM by combining morphological filters to classify land uses using QuickBird, a high spatial resolution satellite with 1-2 meters pixel size. They tested multiple morphology-based features within various regions. According to different regions, the optimal method that results in the best performance only showed fluctuant accuracy between 45% and 75%. Li et al. (2010) provided a novel SVM-based algorithm using QuickBird data. The SVM was integrated with a scene segmentation algorithm in order to perform a better classification result on high-spatial-resolution imagery. The overall accuracy of LULC classification was approximately 86.5%.

SVM can produce a comparable accuracy only using limited labeled samples as training data. This is consistent with the concept of "support vector", which relies on only a few data points to define the classifier's hyperplane (Mountrakis et al., 2011). The origin SVM perform well for medium-resolution images, but does not for high-resolution images. With the increase of dimensionality, typical dimensionality issues, such as increased unusual outliers and computational demands, also appeal within SVM. It means SVM does not work well for training hyperspectral data since the dimensionality of the originally hyperspectral data is high (Mountrakis et al., 2011). (Notice: the hyperspectral imagery refers to the remote sensing image with more than 15 bands, its spatial resolution is usually greater than 10 m.)

2.1.2 Random Forest

The random forest is an ensemble classifier that randomly selected subsets from training data and variables as actual training samples to generate multiple decision trees. Then, these multiple decision trees vote for the category of the testing/predicted datasets. Due to its high classification accuracy, the remote sensing community has shifted its attention to the random forest in the past years (Miao et al., 2012; Belgiu and Drăguţ, 2016). The random forest, as a successful classifier, has been widely used in LULC classification. Simultaneously, the random forest also has the characteristics of fast computational speed and a small number of samples. Compared with SVM, random forest results perform slightly better for high-dimensional training data than SVM, meaning RF is an optimal classifier for hyperspectral imagery classification (Ghosh and Joshi, 2014).

The random forest randomly selects a subset of training samples by way of replacement to create the tress. Among the training samples, approximately 2/3 of the sample (as known as in-bag samples) would be utilized to randomly generate the trees, and the rest 1/3 of samples (as known as out-of-the bag samples are used to conduct internal cross-validation to assess the performance of the RF outputs. The pruning that is common in the decision trees should be ignored when each independent tree is being produced. In addition, users should define the number of features expected to be used (Mtree), and each node would be split according to it. The trees would not stop generating until reaching a user-defined number of trees (Ntree). The tree produced in this way has the characteristics of high variance and low deviation (Belgiu and Drăguţ, 2016). Finally, the class with the highest average probability calculated from all trees is the classification decision. Lawrence et al. (2006) proposed that Ntree should be set as 500, because the errors usually reach stable before the iteration of 500 finished. This number of trees has been accepted by the majority of remote sensing researchers as a default. As to Mtrees, this parameter is usually set to the square root of the number of input variables (Gislason et al., 2006). Figure 2 shows the training and classification phases of the Random Forest classifier below.

Due to its outstanding performance, the RF classifier has been successfully used in LCLU classification for many years. Deng and Wu (2013) compared the effects of spectral mixture analysis (SMA) and random forest about the land-cover classification within an urban area. They use a medium-spatial-resolution imagery, called MODIS, as the data source. The random forest outperforms SMA in any sample size, and its best classification accuracy has been up

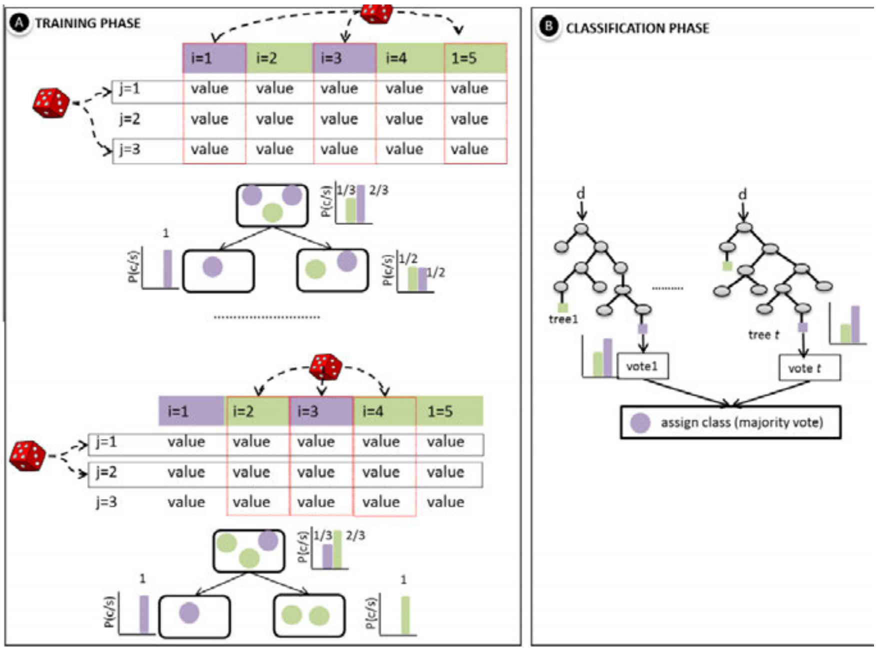

Figure 2: The training and classification phases of Random Forest classifier (Source: adapted from (Belgiu and Drăguţ, 2016))

to 85%. In recent years, a number of novel approaches have been developed for improving the random forest classification. Zhang and Suganthan (2014) proposed a novel method to improve the diversity of the RF tree, thereby increasing the overall classification accuracy. They concatenated various rotation spaces into a higher space at the root node of individual trees, aiming to increase individual tree diversity in random forest during the training process. The results showed that the RF with higher diversity outperform the original RF in most cases, and its overall accuracy of LULC classification was 87.3%. Du et al. (2015) proposed to classify land-cover within arctic pole area using rotation RF classifiers. They compared the results of rotation RF, Random Forest, and using SVM as benchmark classifiers, concluding that the rotation RF had a remarkable improvement in terms of overall classification accuracy within the North Pole Area, but at an increased computational cost. In the meantime, Random Forest is much faster than rotation RF.

Since the trees in RF are created by randomly selecting a subset from training data and a subset of variables to split at each tree node, the Random Forest shows insensitivity to the training data (quality and overfitting). SVM and other typical machine learning classifiers are more sensitive to the quality of data. Belgiu and Drăguţ (2016) illustrated that Random Forest shows a better performance for classification results when high dimensional data, such as hyperspectral imagery, is used, and RF is faster than other mainstream machine learning methods, including SVM.

In conclusion, both Random Forest and SVM are reliable in terms of LULC classification in remote sensing, widely used by many researchers worldwide. By a meta-analysis, Ma et al. (2017) collected statistics from more than 220 LULC classification studies. They found that RF usually has the highest average classification accuracy (85.81%), followed by SVM classifier (85.19%) for LULC classification. It should be noted that due to the limitations of sensor technology, most of the studies used low or medium-resolution images in the early classification studies.

2.2 Classification by Deep Learning (CNN) using High-Resolution Imagery

Over the past several years, Deep Learning algorithms have been increasingly popular in the remote sensing community. Deep learning is an algorithm based on neural networks (NN), proposed as an explicit research field in the 1990s (Hochreiter, 1991). Machine learning researchers ignored the neural network at that time due to the limitation of the hardware. The other machine learning methods represented by RF and SVM are more attractive to remote sensing communities for a relatively long period. Since 2014, the remote sensing community shifted its attention from traditional machine learnings to deep learning because deep learning outperforms traditional machine learning in numerous applications, such as land use and land cover classification using high-spatial-resolution imagery (Ma et al., 2019). So far, the DL is most frequently used to classify land cover types with high spatial resolution images (spatial resolutions of 10 m or finer than 2 m). Through the meta-analysis of remote sensing applications, Ma et al. (2019) reviewed 171 peer-reviewed articles mainly published from 2014 to 2018. They illustrated that more than 50% of articles use Convolutional Neural Network (CNN) model, followed by AE and RNN models. At the same time, about 70 papers focus on LULC classification based on high-spatial-resolution imagery. Therefore, it concludes that CNN has been the most popular DL algorithm for LULC classification based on high-spatial-resolution imagery.

Convolutional Neural Network is a kind of feed-forward neural network (Zhang et al., 2019), specially designed to deal with data in the form of multiple arrays (LeCun et al., 2015). Usually, the input data for CNN is color images with pixels. In fact, remote sensing images are similar to normal color images where pixels are arranged regularly but contain much more bands than normal color images. CNN usually comprises three different layers – convolutional layer, pooling layer, and fully-connected layer. Each convolutional layer comprises several convolution units where each unit is optimized through the backpropagation algorithm. The convolution operation aims to extract different features of the input. Only some low-level features may be extracted after the convolutional layer's first layer. However, more complex features can be extracted iteratively based on low-level features. After the convolutional layer, a feature with a large dimension is usually obtained. The polling layers aim to obtain a new feature with a smaller dimension by computing the maximum or average value within a sub-region (Zhang et al., 2019). The fully-connected layer generates global features according to all local features, which are used to calculate the final probability of each category. The similarity to the neural network, each neural unit in CNN has a set of K kernel weights and added biases, but the CNN's neural unit is 3D volumes of neurons. The reason is that both input image and output results of CNN have three dimensions, where the third dimension represents the number of bands in the pixel.

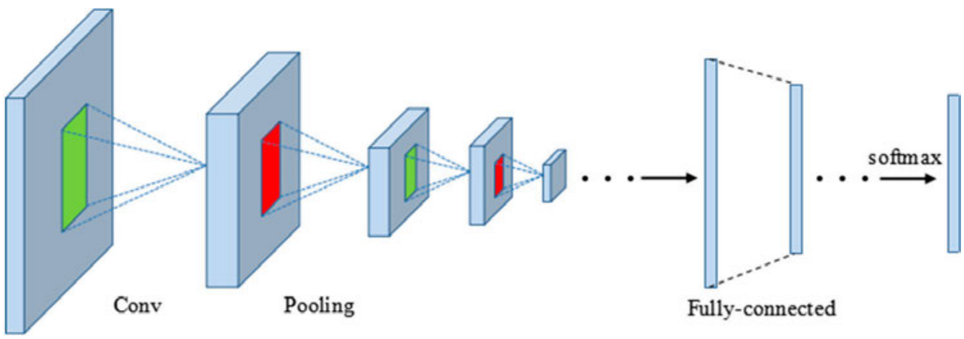

Figure 3: The Convolutional Neural Network architecture (Source: adapted from (Zhang et al., 2019))

Generally, it has been found that DL algorithms are commonly used to classify high-resolution-images that have fine structural information (Ma et al., 2019). Maggiori et al. (2016) proposed an end-to-end framework for the classification of high-spatial-resolution satellite imagery based on the CNNs algorithm, aiming to solve imperfect training data. The framework included two steps: they first initialize the CNN using a large amount of inaccurate training data and then optimize it based on a small amount of accurately training data. They tested the framework in a large-scale rural region, and the results show the accuracy of classification is up to 99.3%, yielding a much better accuracy than the traditional SVM method. Huang et al. (2018) applied CNN to classify urban land-use types from a high-spatial-resolution WorldView-3 image of 143 km2 area of Hong Kong. Finally, the result showed they obtained an overall accuracy of 91.25% for Hong Kong urban land-use classification. Marcos et al. (2018) proposed an improved CNN algorithm (RotEqNet) that can maintain outstanding performances for classification in sub-meter resolution images using a very small amount of training data. The overall accuracy for classification using the improved CNN is 87.5% and the overall accuracy for classification using original CNN is 87.4%, but both of them outperform SVM with an overall accuracy of 76.8%.

According to the statistics proposed by Ma et al. (2019), among 171 peer-reviewed papers about remote sensing DL published between 2014 -2018, 70 papers focused on LULC classification regarding sub-areas of remote sensing, and more than 130 papers used high-spatial-resolution imagery regarding the spatial resolution of images. This means LULC classification studies mostly focused on high-spatial-resolution images, in the meantime, the CNN model has been the most commonly used method for remote sensing analysis in deep learning. Therefore, compared with other deep learning studies in remote sensing, the LULC classification of high-spatial-resolution images using CNNs is the most frequent research. What's more, DL algorithms' overall accuracy for LULC classification exceeds 91% (Ma et al., 2019), significantly higher than the accuracy of other traditional supervised classifiers, such as FR and SVM. As mentioned in the last section, neither RF and SVM have an overall accuracy of more than 90% for supervised LULC classification in remote sensing. However, the authors did not mention that most DL often use high-resolution images while most ML often use medium-resolution images.

2.3 Classification by Improved CNN using Medium-Resolution Imagery

For LULC classification, there is no doubt that deep learnings proved they have super-precision performance compared with traditional machine learning, for example, RF and SVM (Ma et al., 2019). However, it was found that DL algorithms typically use high-resolution images for LULC classification; in the meantime, traditional classifiers presented by RF and SVM usually use medium- or low-resolution images for LULC classification. There have been many free medium-resolution (10m – 30m) satellite images for LCLU mapping since the 1980s, so researchers often utilize these free images to conduct studies. Also, the medium or low-resolution images make direct observations across large areas of the land surface possible. In fact, in large-scale remote sensing researches, medium-resolution images are more common and useful than high-resolution images due to their small storage space and fast computation speed that can cover larger regions (Sharma et al., 2017). Unfortunately, because of the lack of fine structures, it is challenging to apply deep learnings to these medium-resolution images directly, and even the overall accuracy of DL algorithms is much lower than some traditional classifiers (Sharma et al., 2017). It is these fine structures that make the DL's classification effect outstanding.

Sharma et al. (2017) proposed a novel improved CNN to develop land cover classification of medium-resolution satellite imagery – a patch-based CNN system. The input of CNN has to be an image-like multidimensional data, but the medium-resolution image is organized based on pixel-based single vector samples. Therefore, the principle of patch-based CNN is to zoom in individual pixel to be a patch sample for supplementing "fine structures". The patch-based CNN extracted samples as patches with size $5 \times 5 \times 8$ out of multidimensional data and labeled patches using each patch's center pixel. Sharma et al. (2017) found that the optimal size of a patch may vary on the source of medium-resolution imagery; however, they proposed the size of $5 \times 5 \times 8$ (pixels $\times$ pixels $\times$ bands) is the optimal size for medium-resolution images Landsat 8 (30 m), which can capture the locally spatial correlation between the center pixel and surrounding pixels and limits heterogeneous pixels. To overlap with neighboring patches, the stride value is set to one to extract all potential patches and valid locations in the data. In the following, the steps are similar to the normal CNN algorithm for training model and predicting results.

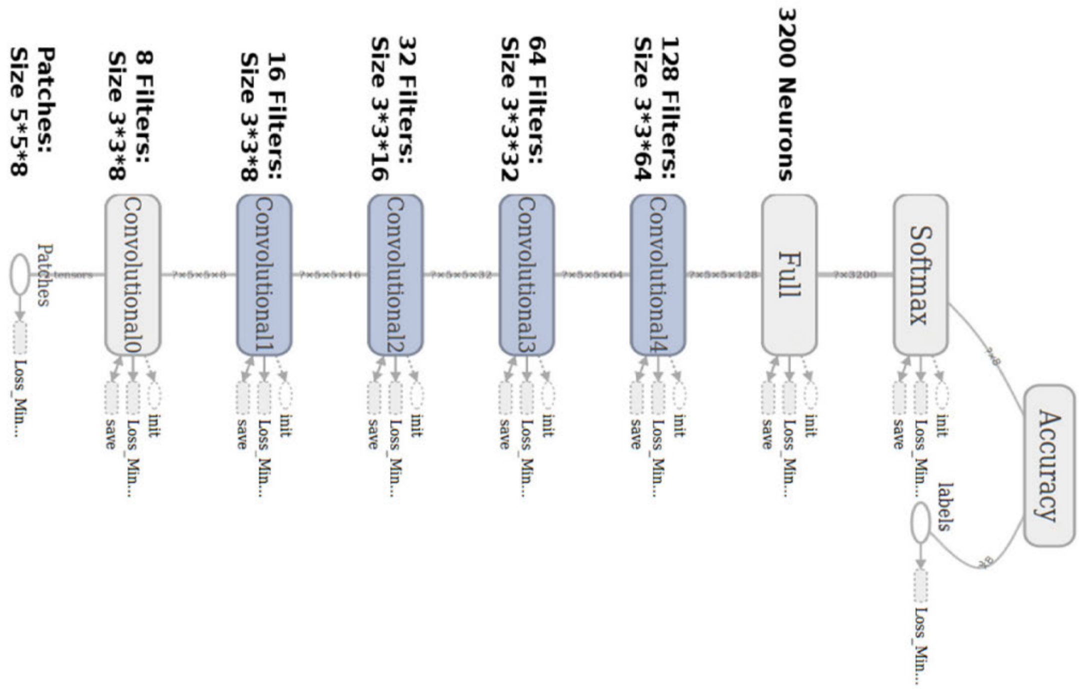

Figure 4: Architecture of patch-based CNN system (Source: adapted from (Sharma et al., 2017))

The patch-based CNN achieved significant success in both overall and categorial classification accuracies. In terms of overall accuracy, the patch-based CNN achieves an accuracy of 85.6%, comparing with pixel-based CNN with an accuracy of 62.34%. It is noticed that the overall accuracy of path-based CNN has not reached the same level as when using high-spatial-resolution images (over 90%). In addition, RF and SVM, with lower cost and faster computational speed, can also achieve an overall accuracy of 85% approximately for LULC classification when using medium-resolution imagery. Hence, the patch-based CNN is of no practiced use in real-world for LULC classification using medium-resolution images.

3 Summary & Conclusion

Remote sensing technology has been widely used in many fields for monitoring the Earth's surface, and the land use and land cover (LULC) classification is one of the streamlined applications for remote sensing. Because of the characteristic of recognition for specific land use, LULC classification usually uses a supervised classifier. Over the past two decades, remote sensing communities shifted their attention from low/medium-spatial-resolution imagery to high-resolution-imagery due to sensor technology development. Correspondingly, they shifted their attention from traditional machine learning presented by RF and SVM to deep learning such as CNN. The accuracy of classification highly depends on the method applied, so the remote sensing community always develops new methods to make it have better performance (Ma et al., 2019).

Before 2015, the medium resolution images are the primary data source for supervised LULC classification. Random Forest and SVM show super-precision performance than another method, including deep learnings, achieving accuracies of 85.81% and 85.19%, respectively (Ma et al., 2017). With the popularity of higher resolution images, the RF and SVM performance for LULC supervised classification reduces significantly. Fortunately, the emergence of deep learning represented by CNN has filled this gap, and even the accuracy of CNN for supervised LULC classification based on high-resolution image exceeds 90%, which is an accuracy rate that has never been achieved by traditional machine learning. Due to CNN's poor performance for a medium-resolution image, Sharma et al. (2017) proposed an improved CNN algorithm to enhance its performance. The results showed that the improved CNN accuracy outperformed the origin CNN, but only reaches the same level as RF and SVM, which means the novel algorithm is useless since this improved CNN would consume more computational cost and require more labeled data for training. RF and SVM are the most commonly used machine learning algorithms, and CNN is the most commonly used deep learning algorithm. It is suggested that RF and SVM would be approved when handling LULC classification in medium-resolution imagery, but a deep learning method, such as CNN, would be recommended when dealing with high-spatial-resolution imagery for LULC classification.

Compared with medium or low-resolution images that only contain spectrum information within pixels, the high-spatial-resolution imagery can provide rich spatial feature information, such as point and line of targets. Deep learning utilizes this high-level feature information to outperform traditional machine learning (Ma et al., 2019). Compared with DL, the significant advantages of RF and SVM includes fewer training samples, faster computational speed, and fewer parameters needed to be determined (meaning it is easier to use for users). Sample labeling needs to be completed manually, which is expensive and time-consuming. Therefore, in practical applications, people actually try to use traditional machine learning to avoid using deep learning. Additionally, the improved patch-based CNN still needs many training samples and relative parameters to determine, meaning the improved CNN is not actually useful in real practices. Finally, the accuracy of various algorithms for supervised LULC classification is also highly dependent on the number of LULC classes (Ma et al., 2019); therefore, it is necessary to comprehensively consider which algorithm to be used according to the actual situation of the application and various other factors. Deep learning does not mean that it is more advanced than machine learning and vice versa. For supervised LULC classification, only a suitable algorithm is the best.

References

Azamathulla, H. M. and Wu, F.-C. (2011). Support vector machine approach for longitudinal dispersion coefficients in natural streams. *Applied Soft Computing*, 11(2):2902–2905.

Belgiu, M. and Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, 114:24–31.

Belward, A. S. and Skøien, J. O. (2015). Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103:115–128.

Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167.

Chi, M., Feng, R., and Bruzzone, L. (2008). Classification of hyperspectral remote-sensing data with primal svm for small-sized training dataset problem. *Advances in space research*, 41(11):1793–1799.

Deng, C. and Wu, C. (2013). The use of single-date modis imagery for estimating large-scale urban impervious surface fraction with spectral mixture analysis and machine learning techniques. *ISPRS Journal of Photogrammetry and Remote Sensing*, 86:100–110.

Du, P., Samat, A., Waske, B., Liu, S., and Li, Z. (2015). Random forest and rotation forest for fully polarized sar image classification using polarimetric and spatial features. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105:38–53.

Ghosh, A. and Joshi, P. K. (2014). A comparison of selected classification algorithms for mapping bamboo patches in lower gangetic plains using very high resolution worldview 2 imagery. *International Journal of Applied Earth Observation and Geoinformation*, 26:298–311.

Gislason, P. O., Benediktsson, J. A., and Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern recognition letters*, 27(4):294–300.

Hochreiter, S. (1991). Untersuchungen zu dynamischen neuronalen netzen. *Diploma, Technische Universität München*, 91(1).

Huang, B., Zhao, B., and Song, Y. (2018). Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. *Remote Sensing of Environment*, 214:73–86.

Huang, C., Davis, L., and Townshend, J. (2002). An assessment of support vector machines for land cover classification. *International Journal of remote sensing*, 23(4):725–749.

Kellenberger, B., Marcos, D., and Tuia, D. (2018). Detecting mammals in uav images: Best practices to address a substantially imbalanced dataset with deep learning. *Remote sensing of environment*, 216:139–153.

Lary, D. J., Alavi, A. H., Gandomi, A. H., and Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1):3–10.

Lawrence, R. L., Wood, S. D., and Sheley, R. L. (2006). Mapping invasive plants using hyperspectral imagery and breiman cutler classifications (randomforest). *Remote Sensing of Environment*, 100(3):356–362.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.

Li, H., Gu, H., Han, Y., and Yang, J. (2010). Object-oriented classification of high-resolution remote sensing imagery based on an improved colour structure code and a support vector machine. *International journal of remote sensing*, 31(6):1453–1470.

Liu, P., Choo, K.-K. R., Wang, L., and Huang, F. (2017). Svm or deep learning? a comparative study on remote sensing image classification. *Soft Computing*, 21(23):7053–7065.

Ma, L., Li, M., Ma, X., Cheng, L., Du, P., and Liu, Y. (2017). A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130:277–293.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., and Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, 152:166–177.

Maggiori, E., Tarabalka, Y., Charpiat, G., and Alliez, P. (2016). Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2):645–657.

Marcos, D., Volpi, M., Kellenberger, B., and Tuia, D. (2018). Land cover mapping at very high resolution with rotation equivariant cnns: Towards small yet accurate models. *ISPRS journal of photogrammetry and remote sensing*, 145:96–107.

Miao, X., Heaton, J. S., Zheng, S., Charlet, D. A., and Liu, H. (2012). Applying tree-based ensemble algorithms to the classification of ecological zones using multi-temporal multi-source remote-sensing data. *International journal of remote sensing*, 33(6):1823–1849.

Mountrakis, G., Im, J., and Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3):247–259.

Niemeyer, J., Rottensteiner, F., and Soergel, U. (2014). Contextual classification of lidar data and building object detection in urban areas. *ISPRS journal of photogrammetry and remote sensing*, 87:152–165.

Shahin, M. and Jaksa, M. (2005). Neural network prediction of pullout capacity of marquee ground anchors. *Computers and Geotechnics*, 32(3):153–163.

Shahin, M. A., Jaksa, M. B., and Maier, H. R. (2001). Artificial neural network applications in geotechnical engineering. *Australian geomechanics*, 36(1):49–62.

Sharma, A., Liu, X., Yang, X., and Shi, D. (2017). A patch-based convolutional neural network for remote sensing image classification. *Neural Networks*, 95:19–28.

Tuia, D., Pacifici, F., Kanevski, M., and Emery, W. J. (2009). Classification of very high spatial resolution imagery using mathematical morphology and support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 47(11):3866–3879.

Zhang, L. and Suganthan, P. N. (2014). Random forests with ensemble of feature spaces. *Pattern Recognition*, 47(10):3429–3437.

Zhang, W., Tang, P., and Zhao, L. (2019). Remote sensing image scene classification using cnn-capsnet. *Remote Sensing*, 11(5):494.