

THE POTENTIAL OF BUILDING DETECTION FROM SAR AND LIDAR USING DEEP LEARNING

Z. Nordin¹, H. Z. M. Shafri^{1,*}, A. F. Abdullah², S. J. Hashim³,

¹ Department of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia –
zuhaidah_nordin@yahoo.com, helmi@upm.edu.my

² Department of Biological and Agricultural Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang,
Selangor, Malaysia – ahmadfikri@upm.edu.my

³ Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400
Serdang, Selangor, Malaysia – sjh@upm.edu.my

KEY WORDS: Synthetic Aperture Radar (SAR), LIDAR, Orthophoto, Feature Extraction, Building Extraction, Deep Learning

ABSTRACT:

The introduction of airborne Synthetic Aperture Radar (SAR) approach has successfully addressed several challenges for mapping and surveying applications. Unlike other conventional sensors, airborne SAR mapping approach offers practicality and significant cost savings for the nation minimizing the need for ground control points on the ground in addition to providing high-resolution, day-and-night, cloud coverage and weather independent images, which in turn provides faster turnaround times for creation of large area geospatial data. Up-to-date building map is necessary to guide the decision making in many fields to understand the urban dynamics such as in disaster management, population estimation, planning and many other applications. Whilst mapping and surveying work using airborne SAR have started to capture many interest among surveyors, professionals and practitioners abroad, Malaysia however is still lacking behind in term of the knowledge and the usage of this technology together with Deep Learning, Machine Learning approach especially in building extraction for topographic mapping and urban planning and development. Deep learning is a subset of the machine learning algorithm. Recently, Deep Learning has been proposed to solve traditional artificial intelligent problems. In order to develop a sustainable national geospatial infrastructure for years to come, the integration between airborne SAR and other sensors as such LIDAR is therefore essential in Malaysia and in high demand for urban planning and management. Thus, this paper reviews current techniques and future trends of multi-sources Remote Sensing for building extraction.

1. INTRODUCTION

Data is the fuel that powers geospatial systems. Acquisition of accurate and reliable data is becoming increasingly important in which sensors and processing approach selection play a signature role. Over the last decade, Remote Sensing approach has been increasingly used in Malaysia especially in urban/city planning, monitoring, disaster management, mapping and many more (Abd Mubin et al., 2019; Hamedianfar et al., 2014; Meesuk et al., 2015; Shaharum et al., 2018). Remote sensing platform usually has a wide coverage, fast development and high frequency and provide high accuracy data. The sensors were mounted on various platforms such as space, air and ground to collect the data and perform the features extraction. Big company like Google employs high-resolution (HR) Remote Sensing images to provide vivid pictures of the earth's surface. Many government agencies also rely on Remote Sensing images weather forecasting and meteorological application.

As in the case of topographic mapping, high quality data acquisition plays a major part towards acquiring results at the highest level. Traditional practices in topographic mapping include surveying techniques coupled with rigorous geodetic networks on the ground for georeferencing purposes to create topographic sheets of large areas. Example of common issues faced by several countries including Malaysia whilst developing their topographic maps is its high cost of operations in recording the physical changes of the ever-expanding and continuously-developing nation. But even as the maps are being developed, they are already perversely out-dated, so the

financial investment and the value of that investment are completely misaligned. This is especially true with ground-based mapping and surveying campaign that historically requires a large sum of expenditure and highly laborious especially towards producing high accuracy elevation models used in creating contours for topographic maps. Since there have been rapid changes especially in urban area, it is necessary to produce very latest and up-to-date map for urban area for urban management and development. The speed of change in urban areas demands regular monitoring to identify areas of high population density and plan for sustainable urban development. Thus combining multispectral aerial imagery and digital surface models to extract urban buildings is needed for rapid monitoring.

Remote Sensing object detection plays an important role toward producing an up-to-date map for environmental monitoring, geological hazard detection, land use land cover (LULC) mapping, geographic information system (GIS) update, precision agriculture, and urban planning. Recently, Deep Learning methods have shown tremendous success for Remote Sensing object detection interpretation problems in which algorithmic methods have deficits. Deep Learning are being used to work with remotely sensed imagery and geospatial data to solve problems in agriculture, utilities, transportation, defence, disaster response, and other industries. A prominent example is the classification and interpretation of images, where Deep Learning approaches outperform the traditional computer vision methods. This paper will highlight some of the previous project in building detection using conventional method and Deep Learning.

2. BUILDING DETECTION FOR MAPPING APPLICATION

Object detection is the procedure of determining the instance of the class to which the object belongs and estimating the location of the object. Object detection is the foremost step in any activity. Being a cost-effective data resource, Remote Sensing data have been considered as the primary resource to generate and detect the object for mapping purposes. There are various methods for the detection of different types of objects in imagery (satellite and aerial images), such as trees, man-made object (roads, building, vehicles and many more) by using traditional method, machine learning and Deep Learning. Detecting single instance of class from image is called as single class object detection, whereas detecting the classes of all objects present in the image is known as multi class object detection.

2.1 Building Detection Using Traditional Technique

In last decade, there are various methods for the detection of different types of objects in RS imagery (satellite and aerial images), such as trees, man-made object (roads, building, vehicles and many more). In general, this task has been performed by human experts manually, so that it is very costly and time consuming. The workflow traditionally used for object detection is relatively straight-forward in Remote Sensing, the images was transform to raster image, and then inspected by human editors for building detection. If a building is found, the building is manually digitized and drawn by the editor. This is most time-consuming and expensive step in the manual search and digitization of polygons. Accurate pixel labelling of images is also complex attentional task for a human, because terrestrial objects have a great deal of variation in their shapes, and an object may be occluded by other objects such as trees and buildings' shadows. This will lead the problem of building generalization using traditional approaches and tends to eliminate the small object or building. Therefore, automatic extraction of buildings and roads is highly demanded and many attempts have been proposed in the Remote Sensing literature. To overcome this, both feature extractors and classifiers are automatically constructed using Deep Learning.

2.2 Building Detection Using Machine Learning Technique

With the advance of machine learning techniques, especially the powerful feature representations and classifiers, many recent approaches regarded object detection as a classification problem and have achieved significant improvements. Object detection can be performed by learning a classifier that captures the variation in object appearances and views from a set of training data in a supervised or semi-supervised or weakly supervised framework. The input of the classifier is a set of regions (sliding windows or object proposals) with their corresponding feature representations and the output is their corresponding predicted labels, i.e., object or not. Feature extraction, feature fusion, dimension reduction and classifier training play the most important roles in the performance of object detection and hence we mainly focus on reviewing these three crucial steps.

Bellakaout, et al., (2016) conducted a research on an approach for automatic classification of aerial LIDAR data solely into five groups of items: buildings, trees, roads, linear object and soil using single return LIDAR and processing the point cloud without generating DEM. Topological relationship and height variation analysis were adapted to segment, preliminary, the

entire point cloud preliminarily into upper and lower contours, uniform and non-uniform surface, non-uniform surfaces, linear objects, and others. This primary classification was used on to identify the upper and lower part of each building in an urban scene, needed to model buildings façades; and to extract point cloud of uniform surfaces which contain roofs, roads and ground used in the second phase of classification.

R. Huang et al., (2018) proposed a method to extract buildings solely based on airborne LIDAR point clouds using a top-down strategy. First, the ground and non-ground points are separated. Second, a top level processing is used to recognize building regions via surface characteristics and penetrating capacities, which are calculated based on the object entity replacing the point and segment entities. Finally, a down-level processing is removed a non-building points from building regions. The results demonstrate that the proposed method could robustly extract the buildings with details (e.g., door eaves and roof furniture) and has good performance in distinguishing buildings. The average values of the area-based quality and object-based quality indicates that the proposed method has good performance for extracting buildings.

LIDAR generates point clouds for digital surface models, digital elevation models and light intensity models. LIDAR has been used in the specially designed algorithms instead of optical sensor images to solve automatic building extraction problems (Canaz et al., 2015, Karsli et al., 2016) LIDAR provides more accurate height information but less accurate boundary lines. Considering the complementary advantages of LIDAR data and high-resolution optic image data, the fusion of two data sources is regarded as a promising procedure to detect the building boundaries (Awrangjeb et al., 2010; (Awrangjeb et al., 2013; Cheng et al., 2013; Li et al., 2013).

Nowadays, the emerging diverse remote sensors are available, obtain complementary information from different sources for materials on the surface of the earth. Such information can vary from spectral information obtained by passive sensors such as multispectral and hyperspectral images, to height and shape information acquired by LIDAR sensors, as well as texture information to amplitude and phase by SAR. The availability of data coming from these multiple sources now allows researchers worldwide to integrate such diverse information to improve object detection ability and classification performance.

Many literatures on LIDAR applications emphasized the needs for data fusion in the processing phase of LIDAR data as a method to improve various feature extraction task. LIDAR alone however can give better accuracy in vertical component in comparison to Airborne SAR which is imperative to produce 3D position of the ground. Point cloud produced by LIDAR also has been used in the specially designed algorithms instead of optical sensor images to solve automatic building extraction problems (Karsli et al., 2016; Michaelsen et al., 2005; Michaelsen, 2010)

The LIDAR data can be used to characterize the elevation and object height information of the scene. (Karsli et al., 2016) proved that multi-features derived from combination of optical and LIDAR data can be successfully applied to solve the problem of automatic detection of buildings by using the proposed approach. LIDAR data provide a source of complementary information that can greatly assist in the classification of hyperspectral data, in particular when it is difficult to separate complex classes. This is because, in

addition to the spatial and the spectral information provided by hyperspectral data, LIDAR can provide very valuable information about the height of the surveyed area that can help with the discrimination of classes and their separability (Khodadadzadeh et al., 2015).

Wegner et al., (2009) performed the study focuses on the combination of features from InSAR data and optical aerial imagery for building recognition in dense urban areas that combines line features from mono aspect InSAR data with classification results from one optical aerial image. The result shows that a combined analysis of InSAR and optical data improves detection results compared to building recognition based on merely a single data source.

Hyperspectral image may not be able to precisely differentiate object composed of the same material which the same spectral characteristics (Plaza et al., 2007). It is difficult to differentiate such categories in the feature space. Rasti et al. (2017) performed the fusion of hyperspectral and LIDAR data called SLRCA fusion. The fusion methodology consists of two main phases. At phase I, spatial and elevation information from hyperspectral and LIDAR data sets is extracted using EPs. At phase II, the SLRCA fusion utilizes an SLRCA to fuse extracted features. The resulting fused features are of lower dimension than the profiles. Both RF and SVM classifiers were used to perform the classification task. It has been shown that applying EPs considerably improves the classification accuracies due to effectively extract spatial and contextual information.

Daneshtalab, S., & Rastiveis, (2017) conducted the research on an automatic urban objects extraction using airborne Remote Sensing data to process and efficiently interpret the vast amount of airborne imagery and LIDAR data available today. The aim of this study is to propose a new approach to integrate high resolution aerial imagery and LIDAR data to improve the accuracy of classification using Support Vector Machine algorithm. In this study, extracted Normalized Digital Surface Model (nDSM) and pulse intensity were used for LIDAR classification together with three spectral visible bands (Red, Green and Blue) for feature vector for the orthoimage classification. The outputs of these classifications were then integrated in a decision level fusion system based on confusion matrices focusing on the urban area of Zeebrugge, Belgium. Based on the output, several advantages can be seen on the image fusion with respect to a single shot dataset. With the capabilities of the proposed decision level fusion method, most of the object extraction difficulties and uncertainty can be minimized and, the overall accuracy and the kappa values can be improved between 7% and 10%, respectively.

X. Huang et al., (2011) investigated information fusion approaches of high-resolution aerial images and elevation data from LIDAR for urban environment mapping. Three feature fusion methods have been proposed and compared: (1) The vector-stacking approach that combines spectral and LIDAR features in one classifier; (2) The re-classification approach that firstly processes spectral signals in a classifier and then integrates its output with LIDAR features to obtain the final result and (3) The post-processing approach that uses the LIDAR data to refine the results of spectral classification. The height features used in the above three algorithms are extracted from the LIDAR digital surface model (DSM) image; these include elevation difference, maximum and minimum values, variance and the grey-level co-occurrence matrix (GLCM) textures. In addition, the average height from object-based

segmentation is also computed. Support vector machines (SVMs) were used as classifiers for all fusion schemes due to their capability and robustness for many classification problems. The three algorithms were evaluated using a 40-cm spatial resolution digital orthophoto and the corresponding LIDAR data of Odense, Denmark. Based on the experiments, the vector-stacking method with the Maximum–Minimum (Max–Min) feature, the reclassification method with the Max–Min feature and the post-processing approach have obtained promising results (94.7%, 95.0% and 94.6%, respectively), which are significantly higher than the spectral-only classification (82.5%).

Demir & Baltasvias, (2010) extracted man-made structures, especially buildings and secondly trees by combining information from aerial images and LIDAR data. They used four different approaches have been applied to exploit the information contained in the image and LIDAR data, extract different objects and finally buildings. The first method is based on DSM/DTM comparison in combination with NDVI (Normalised Difference Vegetation Index) analysis for building detection. For Vaihingen area, while there is no available DTM data, a morphological filtering approach (Zhang, Lin, & Ning, 2013) has been applied to detect off-terrain objects. The second approach is a supervised multispectral classification refined with height information from LIDAR data and image-based DSM. The third method uses voids in LIDAR DTM and NDVI classification. The last method is based on the analysis of the density of the raw DSM LIDAR data.

In each method, the basic idea was to get first preliminary results and improve them later using the results of the other methods. The methods have been tested on two dataset located at Zurich Airport, Switzerland, and Vaihingen region, Germany. The accuracy of the building detection process was evaluated by comparing the results with the reference data and computing the percentage of data correctly extracted and the percentage of reference data not extracted. The results from each method have been combined according to their error characteristics. Roof surfaces have been extracted and finally, the correctness of detection has been improved to 94% with remaining 7% omission error for Zurich airport, and 90% with remaining 17% omission error for the Vaihingen dataset.

Further processes will be applied for the quality assessment of the detected roof planes and then direct 3D edge matching will be done and detection of 3D inner and outlines using aerial images will be generated. First 2D line segments will be extracted using Harris corner and canny edge detectors with splitting edges in Harris corner points. Then 2D line matching will be performed to reconstruct 3D lines. After extraction of 3D lines, reconstruction of roofs will be completed by combination of 3D roof surfaces and 3D lines of the roofs. This combination has done with grouping of 3D lines according to their 3D surfaces.

3. DEEP LEARNING (DL)

Deep Learning (DL) is increasingly becoming an important field for Data Science, AI, Technology and our lives right now, and it deserves all of the attention is getting. Deep Learning is a subset of Machine Learning methodologies using artificial neural networks (ANN) inspired by the structure of neurons located in the human brain. DL came naturally to fit this missing spot to transform the way we do machine learning.

Deep Learning is an active field of research. Many researchers currently focus on finding the best models, topology of the networks, and the best ways to optimize their hyperparameters. The word deep refers to the presence of many layers in the artificial neural network. Recent years, tremendous progress has been made in applying Deep Learning techniques to computer vision. Their latest advancement in Deep Learning are blooming in computer science and more powerful and has been apply in wide range of application such as medical (Ronneberger et al., 2015), biomedical (McKinley Richard et al., 2016), audio (Hershey et al., 2017), face detection (Haoxiang & Lin, Zhe, 2015; Li, Haoxiang et. al, 2015) and Remote Sensing.

Traditional object detection methods are built on handcrafted features and shallow trainable architectures. Their performance easily stagnates by constructing complex ensembles which combine multiple low-level image features with high-level context from object detectors and scene classifiers. With the rapid development in Deep Learning, more powerful tools, which are able to learn semantic, high-level, deeper features, are introduced to address the problems existing in traditional architectures.

3.1 Deep Learning Architecture

With the advance in Deep Learning, researchers have developed novel deep architectures as an alternative to shallow architectures. Deep learning with deep neural networks is commonly used to learn high-level features hierarchically. Typical deep neural network architectures include stacked autoencoders (AE), deep belief networks (DBN), stacked denoising autoencoders (SDAE), and convolutional neural networks (CNN). Due to its local receptive fields, the CNN plays a dominant role for processing the visual-based issues. A primary use of the CNN is classification. The CNN can be employed to extract spatial information effectively via the local connections.

3.1.1 Autoencoders (AE)

The emerging development in Deep Learning of different architectures, which can be supervised or unsupervised have been designed and implemented for feature extraction and classification. An autoencoder as a neural network is an unsupervised learning algorithm that applies backpropagation, setting the target values to be equal to the inputs (Er̄molaev, 1998) achieved great success in generating abstract features of high dimensional data. An AE is composed of an encoder and a decoder. The main principle of AE as it name “Auto” presents that this method is unsupervised and “encoder” means it learns another representation of data (Meng et al., 2017). They can learn deep features automatically layer after layer, thus eliminating the need for manually designed features. Unsupervised deep neural architectures such SAE, SDAE and DBN have shown much better results than the previous methods regarding classification accuracy and have been apply of many researcher in Deep Learning for Remote Sensing.

3.1.2 Convolutional Neural Networks (CNN)

CNN is a supervised Deep Learning based method that has been widely used in analysis and processing of visual imagery outperforming existing Deep Learning methods in detection and classification tasks. A typical CNN architecture comprises of an input layer, a layer comprising of the various output classes (output layer), and in between them there are multiple hidden

layers. The hidden layers comprise of convolution layers, own sampling or pooling layers, normalization layers, and fully connected (FC) layers. The final FC layer is linked to the output layer. During the training process, a CNN can automatically learn features from input data layer after layer, where high-level features are learnt from lower level features. These features are detected by filters in the output of the previous layer.

CNNs can directly train a deep network without pretraining, achieving better Remote Sensing scene classification accuracies. The deep structure of CNNs allows the model to learn highly abstract feature detectors and to map the input features into representations that can clearly boost the performance of the subsequent classifiers. Deep learning models have been attractive for their high performance in learning hierarchical features from high-dimensional unlabelled data. By learning multilevel feature representations, Deep Learning models have been proved to be an effective tool for rapid object-oriented classification problems. Figure 1 shows the example of neural network architecture which the following structure is constructed with two (2) Neurons in the input layer, eight (8) Neurons in the 1st hidden layer, eight (8) Neurons in the 2nd hidden layer and one (1) Neuron in the output layer.

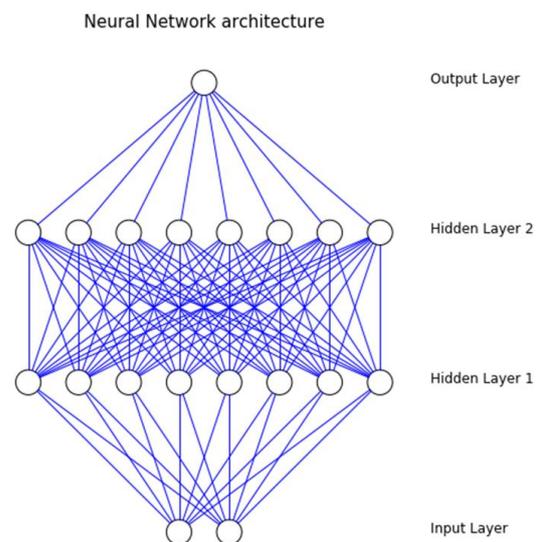


Figure 1. Example of neural network architecture

3.2 Building Extraction Using Deep Learning

Object recognition or object extraction in Remote Sensing is a core task. The used of Deep Learning method in Remote Sensing image has been preferred to extract the object for many purposes. Saito et al., (2016) applied the convolutional neural networks to extract multiple objects from aerial image. Sherrah (2016), applied fully convolutional neural networks for dense semantic labelling of high-resolution aerial image. The approach is applied to the problem of labelling high-resolution aerial imagery where fine boundaries are importance.

Generally, when using deep neural networks, we just normalize images and apply transformations to artificially increase the dataset, like mirror and small rotations. This happens because the convolutional layers of the neural network will gradually become feature extractors that could out-perform most of manual pre-processing.

CNN technique is implemented due to being not only efficient in learning highly discriminative image features such as buildings, but also partially robust to incomplete and poorly registered target maps. Zou et al., (2015) introduce Deep Learning-based feature extraction for hyperspectral data classification by applying autoencoder (AE), which is one of the deep architecture-based models, to learn deep features of hyperspectral data in an unsupervised manner. The author exploit single-layer AE and multi-layer stacked AE (SAE) to learn shallow and deep features of hyperspectral data, respectively and propose a new way of extracting spatial dominated information with dealing the joint spectral–spatial information for classification framework.

Romero et al., (2016) introduces the use of single-layer and deep convolutional networks into AE for data analysis to analyse the influence of depth and pooling of such networks on a wide variety of Remote Sensing images of different spatial and spectral resolutions, from MS and HS images, to very high geometrical resolution problems. The results shows the trained networks are very effective at encoding spatio spectral information of the images. Y. Chen, et al., (2014) adopt the AE for hyperspectral data classification methods and SAEs deep architecture. It is shown that AE extracted features are useful for classification.

Deep CNN have been extensively used for object detection. Which became the state-of-the art for image recognition problems, it became possible to automatically detect buildings in Remote Sensing data (Bittner et al., 2017). CNN technique is implemented due to being not only efficient in learning highly discriminative image features such as buildings, but also partially robust to incomplete and poorly registered target maps.

Automatic building footprint extraction become more challenges despite plenty of the algorithm have been develop using traditional computer vision to solve the problem in adequate to extract the building. Deep Learning shows the tremendous success to solve the problem for automatic building generalization. Sester et al., (2018) used existing network and proved that Deep Learning gives advantage of the building generalization task which not provided one classification for a given image patch but lead the classification of each pixel. The fact that many training data sets are available from given map series. Duarte et al., (2018) extract building damage with CNN by using three multi-resolution CNN feature fusion approach.

Hamed et al., (2018) proposed Deep Learning approach for building detection using LIDAR-orthophoto fusion. The method utilized object - based analysis to create objects, a feature - level fusion, an autoencoder-based dimensionality reduction to transform low -level features into compressed features, and a CNN to transform compressed features into high - level features, which were used to classify objects into buildings and background. The proposed architecture was optimized for the grid search method, and its sensitivity to hyperparameters was done with the best hyperparameters of the model are 128 filters in the CNN model, the Adamax optimizer, 10 units in the fully connected layer of the CNN model, a batch size of 8, and a dropout of 0.2. The study was compare with the support vector machine (SVM) show that the proposed model with or without dimensionality reduction outperforms the SVM models in the working area. The result shows that the use of an autoencoder in DL models can improve the accuracy of building recognition in fused LIDAR–orthophoto data.

Nogueira et al., (2017) analysed of three possible strategies for exploiting the power of existing convolutional neural networks (ConvNets or CNNs) in different scenarios from the ones they were trained: full training, fine tuning, and using ConvNets as feature extractors and shown the result using the features from the fine-tuned ConvNet with linear SVM obtains the best results. Yang et al., (2018) proposed the combination signed-distance labels with SegNet, the preferred CNN architecture identified by extensive evaluations, to advance building extraction results to instance level and demonstrate the usefulness of fusing additional near IR information into the building extraction framework.

Deep Learning, in short, is going much beyond machine learning and its algorithms that are either supervised or unsupervised. In Deep Learning it uses many layers of nonlinear processing units for feature extraction and transformation. Learning is based on multiple levels of features or representation in each layer with the layers forming a hierarchy of low-level to high-level features where traditional machine learning focuses on feature engineering, Deep Learning focuses on end-to-end learning based on raw features. Deep Learning has proven its performance in Remote Sensing field including in classification task and object recognition and have a good potential for future tool especially in building extraction task in urban area.

4. CONCLUSION

Nowadays, there is an emerging interest worldwide on the integration between the SAR and other remote sensors for various surveying and mapping applications. Resulting from SAR capabilities to generate wide-range DEM, the synergy between SAR and multiple sensors has brought in the potential of synoptic viewing and repetitive coverage that is expected to be capable of drastically reducing costs, timelines and improving the accuracy of topographic maps. It is also expected that both life and value of the initial investment in developing a topographic map via this approach can be extended – given the rapid advancement in Remote Sensing and image processing technologies to update maps on regular basis - thus changing the paradigm of traditional mapping in the past.

Building extraction has been an active research area in Remote Sensing field. Even though there have been many achievement, there are still challenges that need to be addressed particularly for urban mapping and planning purposes. The emergence of Deep Learning will solve this limitation in which the extraction of building with fusion of multisource data can provide data duplications and useful information faster and more accurate compared to conventional methods. Recent literatures abroad highlighted Deep Learning as a new exciting tool in building extraction with advantages in dealing with occlusion, scale transformation and other limitation. Object detection using Deep Learning is growingly becoming a research hotspot in coming years and has the potential to be used for mapping applications in Malaysia.

REFERENCES

- Abd Mubin, N., Nadarajoo, E., Shafri, H., & Hamedianfar, A. (2019). *Young and mature oil palm tree detection and counting using convolutional neural network deep learning method. International Journal of Remote Sensing.* <https://doi.org/10.1080/01431161.2019.1569282>
- Awrangjeb, M., Ravanbakhsh, M., & Fraser, C. S. (2010). Automatic detection of residential buildings using LIDAR data and multispectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(5), 457–467. <https://doi.org/https://doi.org/10.1016/j.isprsjprs.2010.06.001>
- Awrangjeb, M., Zhang, C., & Fraser, C. S. (2013). Automatic extraction of building roofs using LIDAR data and multispectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing.* <https://doi.org/10.1016/j.isprsjprs.2013.05.006>
- Bellakaout, A., Cherkaoui, M., Ettarid, M., & Touzani, A. (2016). Automatic 3D Extraction of Buildings, Vegetation and Roads from LIDAR Data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLI(XXIII ISPRS Congress)*, 173–180. <https://doi.org/10.5194/isprsarchives-XLI-B3-173-2016>
- Bittner, K., Cui, S., & Reinartz, P. (2017). Building extraction from remote sensing data using fully convolutional networks. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives.* <https://doi.org/10.5194/isprs-archives-XLII-1-W1-481-2017>
- Canaz, S., Karsli, F., Guneroglu, A., & Dihkan, M. (2015). Automatic boundary extraction of inland water bodies using LiDAR data. *Ocean & Coastal Management*, 118, 158–166. <https://doi.org/https://doi.org/10.1016/j.ocecoaman.2015.07.02>
- Chen, Y., Lin, Z., Zhao, X., Wang, G., & Gu, Y. (2014). Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.* <https://doi.org/10.1109/JSTARS.2014.2329330>
- Daneshtalab, S., & Rastiveis, H. (2017). Decision Level Fusion of Orthophoto and Lidar Data Using Confusion Matrix Information for Land Cover Classification. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-4/W4*, 59–64.
- Demir, N., & Baltasvias, E. (2010). Combination of Image and Lidar Data for Building and Tree Extraction. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.*
- Duarte, D., Nex, F., Kerle, N., & Vosselman, G. (2018). Multi-Resolution Feature Fusion for Image Classification of Building Damages with Convolutional Neural Networks. *Remote Sensing.* <https://doi.org/10.3390/rs10101636>
- Ermolaev, A. M. (1998). Atomic states in the relativistic high-frequency approximation of Kristić-Mittleman. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 31(3), 1–19. <https://doi.org/10.1088/0953-4075/31/3/001>
- Hamed Nahhas, F., Shafri, H., Ibrahim, M., Pradhan, B., & Mansor, S. (2018). *Deep Learning Approach for Building Detection Using LiDAR-Orthophoto Fusion. Journal of Sensors* (Vol. 2018). <https://doi.org/10.1155/2018/7212307>
- Hamedianfar, A., Shafri, H. Z. M., Mansor, S., & Ahmad, N. (2014). Improving detailed rule-based feature extraction of urban areas from WorldView-2 image and lidar data. *International Journal of Remote Sensing.* <https://doi.org/10.1080/01431161.2013.879350>
- Haoxiang, L., & Lin, Zhe. (2015). A Convolutional Neural Network Approach for Face Identification. *IEEE Conference on Computer Vision and Pattern Recognition.* <https://doi.org/10.1109/CVPR.2015.7299170>
- Hershey, S., Chaudhuri, S., Ellis, D. P. W., Gemmeke, J. F., Jansen, A., Moore, R. C., ... Wilson, K. (2017). CNN architectures for large-scale audio classification. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings.* <https://doi.org/10.1109/ICASSP.2017.7952132>
- Huang, R., Yang, B., Liang, F., Dai, W., Li, J., Tian, M., & Xu, W. (2018). A top-down strategy for buildings extraction from complex urban scenes using airborne LiDAR point clouds. *Infrared Physics & Technology*, 92, 203–218. <https://doi.org/https://doi.org/10.1016/j.infrared.2018.05.021>
- Huang, X., Zhang, L., & Gong, W. (2011). Information fusion of aerial images and LIDAR data in urban areas: Vector-stacking, re-classification and post-processing approaches. *International Journal of Remote Sensing.* <https://doi.org/10.1080/01431160903439882>
- Karsli, F., Dihkan, M., Acar, H., & Ozturk, A. (2016). Automatic building extraction from very high-resolution image and LiDAR data with SVM algorithm. *Arabian Journal of Geosciences* (Vol. 9). <https://doi.org/10.1007/s12517-016-2664-7>
- Khodadadzadeh, M., Li, J., Prasad, S., & Plaza, A. (2015). Fusion of Hyperspectral and LiDAR Remote Sensing Data Using Multiple Feature Learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (Vol. 8). <https://doi.org/10.1109/JSTARS.2015.2432037>
- Li, Haoxiang and Lin, Zhe and Shen, Xiaohui and Brandt, Jonathan and Hua, G. (2015). A Convolutional Neural Network Approach for Face Detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* <https://doi.org/10.1109/CVPR.2015.7299170>
- McKinley Richard, Gundersen, T., Wagner, F., Chan, A., Wiest, R., & Reyes, M. (2016). Nbla-net: a deep sag-like convolutional architecture for biomedical image segmentation: application to white-matter lesion segmentation in multiple sclerosis.
- Meesuk, V., Vojinovic, Z., Mynett, A. E., & Abdullah, A. F. (2015). Urban flood modelling combining top-view LiDAR data with ground-view SfM observations. *Advances in Water Resources.* <https://doi.org/10.1016/j.advwatres.2014.11.008>
- Meng, Q., Catchpoole, D., Skillicom, D., & Kennedy, P. J. (2017). Relational autoencoder for feature extraction. In *Proceedings of the International Joint Conference on Neural*

Networks. <https://doi.org/10.1109/IJCNN.2017.7965877>

Michaelsen, E., Soergel, U., & Thoennessen, U. (2005). Potential of building extraction from multi-aspect high resolution amplitude SAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. XXXVI(Part 3/W24), 149–154.

Michaelsen, E., Stilla, U., Soergel, U., & Doktorski, L. (2010). Extraction of building polygons from SAR images: Grouping and decision-level in the GESTALT system. *Pattern Recognition Letters*, 31(10), 1071–1076. <https://doi.org/https://doi.org/10.1016/j.patrec.2009.10.004>

Nogueira, K., Penatti, O. A. B., & dos Santos, J. A. (2017). Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognition*, 61. <https://doi.org/10.1016/j.patcog.2016.07.001>

Plaza, J., Plaza, A., Gamba, P., & Trianni, G. (2007). Efficient multi-band texture analysis for remotely sensed data interpretation in urban areas. In *2007 Urban Remote Sensing Joint Event, URS*. <https://doi.org/10.1109/URS.2007.371780>

Romero, A., Gatta, C., & Camps-Valls, G. (2016). Unsupervised deep feature extraction for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*. <https://doi.org/10.1109/TGRS.2015.2478379>

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-24574-4_28

Saito, S., Yamashita, T., & Aoki, Y. (2016). Multiple Object Extraction from Aerial Imagery with Convolutional Neural Networks. *Journal of Imaging Science and Technology*. <https://doi.org/10.2352/J.ImagingSci.Technol.2016.60.1.010402>

Sester, M., Feng, Y., & Thiemann, F. (2018). *Building Generalization Using Deep Learning*. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (Vol. XLII-4). <https://doi.org/10.5194/isprs-archives-XLII-4-565-2018>

Shaharum, N. S., Shafri, H., A. W. A. K. Ghani, W., Samsatli, S., Yusuf, B., Al-Habshi, M., & Prince, H. M. (2018). *Image Classification For Mapping Oil Palm Distribution Via Support Vector Machine Using Scikit-Learn Module*. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (Vol. XLII-4/W9). <https://doi.org/10.5194/isprs-archives-XLII-4-W9-133-2018>

Sherrah, J. (2016). *Fully Convolutional Networks for Dense Semantic Labelling of High-Resolution Aerial Imagery*.

Wegner, J. D., Soergel, U., & Thiele, A. (2009). Building extraction in urban scenes from highresolution InSAR data and optical imagery. In *2009 Joint Urban Remote Sensing Event*. <https://doi.org/10.1109/URS.2009.5137481>

Yang, H. L., Yuan, J., Lunga, D., Laverdiere, M., Rose, A., & Bhaduri, B. (2018). Building Extraction at Scale Using Convolutional Neural Network: Mapping of the United States.

IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(8). <https://doi.org/10.1109/JSTARS.2018.2835377>

Zhang, J., Lin, X., & Ning, X. (2013). SVM-Based classification of segmented airborne LiDAR point clouds in urban areas. *Remote Sensing*. <https://doi.org/10.3390/rs5083749>

Zou, Q., Ni, L., Zhang, T., & Wang, Q. (2015). Deep Learning Based Feature Selection for Remote Sensing Scene Classification. *IEEE Geoscience and Remote Sensing Letters*, 12(11), 2321–2325. <https://doi.org/10.1109/LGRS.2015.2475299>

Revised August 2019