IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

CLASSIFICATION OF DIFFERENT DEEP-LEARNING TECHNIQUES FOR HIGH RESOLUTION SATELLITE IMAGERY OF DAMAGED BUILDING EXTRACTION

Ms. Tanvi Rajnikantbhai Navik
Research Scholar
Department of Computer Applications,
Sabarmati University ,
Ahmedabad,Gujarat, India
Email:
tanvinavik25890@gmail.Com

Dr. Mukta Agarwal
Associate Professor
Department of Computer Applications
Sabarmati University ,
Ahmedabad,Gujarat, India
Email:
mukta.agarwal@sabarmatiuniversity.edu.in

Abstract - With the improved availability of high-resolution satellite imagery, it is possible to detect detailed structures on our planet's surface. Expert knowledge, supervision, and fieldwork are required for traditional mapping approaches. This study examines the potential of machine learning methods such as artificial neural networks (ANN), support vector machines (SVM), and random forest (RF), as well as different deep-learning convolution neural networks (CNNs) for high-resolution satellite imagery using optical data from the satellite and topographic factors. To do multi-label categorization of Amazon satellite images, this uses a Convolutional Neural Network (CNN) model. It starts with a CNN model that obtains an F-score of 0.84. Then, using three deep CNN architectures that have recently performed well in the ImageNet Challenge, demonstrate that a ResNet-50 model can get a 0.91 F-score. The results reveal that without any additional post-processing steps, state-of-the-art methods outperform by 9.8% on the Intersection over Union metric. The DeepGlobe Building Extraction Challenge asks participants to extract all building polygons from satellite images. They use the segmentation algorithm Mask R-CNN and a method that combines Mask R-CNN with building boundary regularisation to construct polygons that are accurate and complete. This approach, however, provides better regularised polygons than Mask R-CNN, which is useful in many applications.

***Keywords*:** Semantic segmentation, Satellite Imagery, High Resolution Satellite Imagery, Deep Learning, Convolution Neural Networks, Mask R-CNN, Support Vector Machine

I. INTRODUCTION

Digital image processing is currently widely employed as a result of technological advancements. Satellites capture remote sensing images that are used in agriculture, defense, navigation, and other fields. Satellite image capturing techniques have vastly improved over time, and image size has correspondingly increased. It's tough to archive such images for future use. It takes a long time to process. Various image processing techniques can be used to alter some of the image's characteristics for improved clarity and storage.

As a result, applying processing techniques to satellite data is rather challenging. In addition, satellite data is captured over great distances and is influenced by the presence of undesired interferences, lowering image quality. This causes significant problems in subsequent processing phases and lowers the overall quality of the final image. The final image is extremely important for future analysis and decision-making. As a result, noise-distorted satellite images should be pre-processed before being subjected to subsequent image processing procedures. Remote sensing is now a constantly increasing technology for gathering satellite data and is critical for analyzing spatiotemporal changes on the earth's surface. This information is gathered by researchers and government agencies all around the world in order to study the earth's changes [1].

You must examine the incidence, features, and consequences of previous events, relate them to the current situation, and make projections for the future. Several researches have been conducted to this purpose, using a variety of knowledge-based methods, some of which also included machine learning (ML) techniques. Both the training and validation stages of machine learning methods necessitate such data sets. The classification and extraction of information from satellite images can be divided into two approaches: object-based and pixel-based. Although object-based image analysis is becoming more common, pixel-based approaches continue to be the most popular. Object-based and pixel-based image analysis have both been combined with various machine learning algorithms and applied to a variety of applications. In general, machine learning (ML) methods are thought to be beneficial for remote sensing applications, particularly image categorization and object detection. For high-resolution satellite imagery using optical data from the satellite and topographic factors, machine learning methods such as artificial neural networks (ANN), support vector machines (SVM), and random forest (RF), as well as different deep-learning convolution neural networks (CNNs), were used. In numerous image processing tasks in computer vision, deep-learning approaches and, in particular, convolutional neural networks

(CNNs) have obtained fairly good results. VHR image classification and segmentation, semantic segmentation, scene annotation utilizing CNNs have been used to recognize objects in a variety of high spatial resolution and aerial photos. Employing a training data set with labeled data is a need for using CNN algorithms. If there are enough training data, a supervised learning strategy like this usually produces good results. CNNs employ supervised machine learning to minimize a loss function by pushing learnable, i.e. adaptable, and feature extractor filters through clots of labeled data. The availability of larger training datasets, improved network topologies, and faster GPUs are all important factors in CNN performance [2].

The maps showed the locations, sizes, and shapes of cities, they were unable to identify small towns and villages in remote rural areas, which necessitate careful consideration in development strategies such as public health, energy development, and transportation infrastructure development. In remote rural locations, the presence of buildings is a good indicator of human activity. Several studies used image recognition techniques such as machine learning and deep learning to develop automatic methods of building detection from high-resolution satellite images. Because building detection using high-resolution satellite images has traditionally relied on human visual interpretation, automated methods are novel for the development of global human settlement mapping, allowing coverage to be extended to towns and villages in rural areas that would otherwise be missed by moderate-resolution satellite data applications. Because machine learning and deep learning approaches typically need a lot of computer resources, the applications were thought to take a long time to cover the entire globe. Massive parallel processing on a high-performance computing system, on the other hand, might dramatically reduce the time necessary for global processing. On a high-performance computing system, we propose an experimental development of an automated method for recognizing buildings in high-resolution satellite images [3].

In remote rural locations, the presence of buildings is a good indicator of human activity. The automatic recovery of a class for each pixel in Very High Resolution (VHR) geospatial images is an issue that most state-of-the-art technologies address using supervised machine learning algorithms. Pre-computed hand-crafted characteristics are input into such techniques, which frequently necessitate specialist expertise. During a training stage, features are constructed to capture and mimic the properties of each class. When categorizing mono-temporal VHR images with superior textural content and intra-class variability, both aspects are improved: feature extraction and training set design both have a significant impact on classification performance. This becomes considerably more challenging at large scales because classes exhibit strong spectrum changes and some classes have comparable properties. Furthermore, a wide geographic scale may result in changes in the collection of classes to evaluate, such as from urban to rural to mountains. Deep Convolutional Neural Networks (DCNNs) have the advantage of being able to generalize information well. This comes at a great cost in terms of training data, as each class's properties and examples must be pinpointed. DCNNs extract their own features from the training dataset in an end-to-end way during the training stage, whereas traditional classifiers of the past decades required hand-crafted features [4].

Because the polygons generated by instance segmentation have irregular shapes that differ significantly from real building footprint borders, they can't be used directly in many cartographic and engineering applications. Building extraction from huge satellite images is still a difficult problem to solve. People have been urged to present automated solutions for extracting buildings from satellite images as part of the DeepGlobe Building Extraction Challenge. The DeepGlobe Building Extraction Challenge asks you to locate all building polygons in the satellite images you've been given. Using an existing instance segmentation approach based on Mask R-CNN, we can construct polygons. Many cartographic and technical applications find it challenging to apply such unevenly shaped polygons directly. As a result, we convert the regularized polygons created by Mask R-CNN [5].

II. PROPOSED SYSTEM

The High resolution satellite images are large and come from a long distance, and they are impacted by noise and other environmental conditions. As a result, before researchers can use them for analysis, they must be processed. For satellite image processing, a range of mathematical approaches and algorithms have been presented. Artificial neural networks (ANN), support vector machines (SVM), and random forest (RF) are among the machine learning methods explored, as are deep-learning approaches such as convolution neural networks (CNNs), deep convolutional neural networks (DCNN), and the segmentation algorithm Mask R-CNN.

i. Artificial Neural Network (ANN)

The ANNs are designed to replicate human brain function, and they can solve difficult nonlinear problems by detecting patterns. A multilayer perceptron (MLP) architecture was employed in this study, and an ANN technique was trained with the back propagation algorithm (BPA), which is the most used approach for training the ANN method. The number of hidden layer units in any MLP is determined by the problem's complexity. With a hidden layer of 30 neurons, the network was feed-forwarded with the identical training input data set in our example. The BPA chooses the initial weights at random. The difference between the output values and the predicted values is then calculated for all observations. Every terrain unit is compared, and forward signals and back-propagating mistakes are fed until the mean-square error stabilizes at an acceptable low level. As a result, the backward process updates all weightings that were randomly picked at the start of each cycle to minimized inaccuracy. [6].

ii. Support Vector Machines (SVM)

Machine learning with SVM is one of the most often used methods, and the output is utilized to produce a classification map. The radial basis function was employed as the classification's basis function. However, the SVM does not guarantee correct findings, and there are certain classification errors. The most extensively used method for object-based categorization is the supervised method. Another extensively used way is to model the segmentation problem using Support Vector Machines (SVM). The SVM is a machine learning method that uses non-linear transformers to translate the problem's dataset into a higher-dimensional space, where an ideal hyperplane is produced for separating the dataset characteristics. When the dividing margins between the defined classes are at their maximum, the ideal hyperplane will be discovered. The support vectors are the maximum separation margins. In a variety of fields, the SVM method has been employed for data classification and regression analysis. Wherever there is a requirement to find patterns or categorized items into certain classes, the SVM algorithm can be employed. Each data sample in the data scatter diagram is represented as a point in the n-dimensional space in this method. One of the graph's point coordinate parameters is determined by the value of each data feature. The SVM classification base is a linear classification of data, and a more trustworthy line is used in linear segmentation of data [7].

iii. Random Forest (RF)

Another popular categorization system is Random Forest (RF). Both supervised and unsupervised classification techniques are available. Multiple decision trees are used in the RF approach. It's been employed in a variety of remote sensing projects. In comparison to other decision trees, the RF approach is less sensitive to the over-fitting problem caused by complicated datasets because it leverages the training data set to generate several deeper decision trees. Each RF decision tree predicts an output, which is weighted according to the value determined from the votes received. In the final classification, the majority voting on an output and a degree of convergence in fitting results. In the final classification, the majority voting on an output and a degree of convergence in fitting results, RF classifications produce good results when used to classify satellite images. As a result, RF is one of the most effective non-parametric ensemble learning methods in image analysis, and it was chosen as a machine learning method for landslide detection in our study [8].

iv. Convolution Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a type of neural network that uses CNNs are image detection deep learning algorithms that use kernel convolutions to extract and learn crucial information from images. It's used for feature detection as well as classification, ensuring that when a fully trained network is given an image to identify, it reacts correctly. We must first create a bounding box around an item or instance within a photograph in order to detect it. Deep learning algorithm CNN is a cutting-edge image processing technology with the ability to self-learn. The architecture of the CNN model assists in the detection of hidden patterns or features. The CNN model has the advantage of assisting in the detection of insects on crops. CNN's multi-layer feed-forward neural networks can extract an image's effective feature representations, allowing them to recognize visual laws in the image without the use of sophisticated human-designed rules. CNNs have a special architecture, with convolutional and pooling layers in each of the so-called hidden layers, with the convolutional layer being the basic building component of any CNN. The original input image is convolved with a set of trainable kernels that scan the whole input image, yielding a collection of feature maps [9]. The pooling layer is frequently computed right after a convolutional layer and is used to down/sub-sample the convolutional layer's output in order to obtain a condensed set of feature maps. The most frequent and widely used pooling layer is max-pooling, which allows you to maintain only the feature maps' maximal values. The main operation of any CNN is believed to be max-pooling. It considerably reduces the spatial size of feature maps and, as a result, the computation volume for the following layers to be processed. The following equation can be used to summarize the key processes done in any CNN [10]:

$$O^{l} = P(\sigma(O^{l-1} * W^{l} + b^{l})) \qquad (1)$$

where O l-1 is the output feature map from the previous layer of the lth layer, W l and b l are the layer's weights and biases, respectively, that convolve the O l-1 by the linear convolution *, and (.) is the non-linearity function outside the convolutional layer. Following these steps, a pooling operation, denoted by P in Equation 1, is frequently performed (1).

v. Deep Convolutional Neural Networks (DCNN)

A method for picture segmentation that uses deep convolutional networks is widely used in applications where buildings must be separated from the backdrop. This method employs supervised classification and makes use of a large training set. The method entails quantification, grouping, and calculating the likelihood using the fewest possible clusters. For semantic segmentation in high-resolution satellite images, a Deep Convolution Neural Network (DCNN) is proposed. First, a unique model is created by adding boundary detection to a SEGNET encoder in this manner. It's a convolutional network and an encoder combined. However, the model's primary flaw is its large size, which is a source of concern for researchers who use it. Furthermore, the extracted borders are rather hazy. The use of deep feature learning to classify high-resolution satellite photos is proposed. The image is first separated into several scales, with images from each scale being used to train the DCNN. Spatial pyramid pooling can help speed up the training process (SPP). The SPP nets can handle a variety of image scales while maintaining the same weight settings. Changing the parameters in a fully connected model can also speed up the training process in each SPP net. Multiple kernel learning is used to optimize the weight during the classification technique. They created DCNN models for semantic segmentation of high-resolution aerial photographs, which clearly represent and extract the borders between semantically distinct regions. Including class boundaries improves several DCNN architectures empirically, and it was the single most substantial performance gain in our final model, which obtains great results on the ISPRS Vaihingen and Potsdam benchmarks. DCNNs are the most advanced technique for semantic segmentation; nonetheless, they have reached a point of saturation, and additional advances in segmentation quality will most likely be limited, time-consuming, and problem-specific [11].

vi. Mask R-CNN

The Mask R-CNN is an extension of the Faster R-CNN that adds a network branch for predicting segmentation masks on each Region of Interest to the original Faster R-CNN (RoI). Given the Faster R-CNN framework, which allows for a large range of configurable architecture designs, Mask R-CNN is straightforward to implement and train. In terms of network topology and hyper-parameter tweaking, R-CNN is basic. We generate polygon points of each unique building after running Mask R-CNN. These polygon points usually have irregular forms and match to building boundaries. The majority of building footprints, however, have regularized borders. Furthermore, many cartographic and technical programs find it challenging to directly use such unevenly shaped polygons. As a result, we use a building boundary regularization method that we adapted from previous work. The regularized polygons are then created from the polygons generated by Mask R-CNN. Furthermore, rather of training a single universal building extractor, we create four building extractors, each taught to handle a distinct city. The Mask R-CNN is made up of two parts: A convolutional backbone architecture that extracts features throughout a complete image, and (ii) a network head that performs classification, bounding box identification, and mask prediction on each RoI separately. Due to the localization of Mask R-pixel-wise CNN's labeling, the initial polygons created by Mask R-CNN have uneven and noisy outlines. We modify earlier work to be applicable in an image domain for implicitly regularizing noisy building boundaries in an iterative approach to convert the initial polygons into regularized ones [12].

III. RESULTS AND DISCUSSION

It is now feasible to spot complex structures on our planet's surface thanks to the increased availability of high-resolution satellite imagery. The increased performance of deep convolutional neural networks (DCNN) has recently been established in the literature for numerous categorization assessments of Very High Spatial Resolution (VHR) geospatial imagery.

Table 1: Statistical analysis of Building Extraction Results.

Method	Count	Minimun	Maximum	Sum	Mean	SD
CNN22,5	286	0.175	136.85	524.32	1.89	9.75
CNN32,8	335	0.205	174.4	508,31	2.16	6.43
CNN12,5	247	0.175	77.16	555,63	1.54	5.37
CNN16,5	281	0.175	96.32	480,52	1.15	4.97
ANN5	489	0.175	117.95	991,98	1.05	4.64
ANN8	546	0.175	153.95	1125,75	0.99	4.5
D-CNN32,5	268	0.195	58.87	467,52	1.27	4.28
D-CNN32,8	277	0.2	76	509,83	1.84	6.45
D-CNN48,5	306	0.22	65.6	589,93	1.49	4.39
SVM5	421	0.175	322.16	754,85	2.24	14.31
SVM8	514	0.175	352.09	798,72	1.69	15.77
RF5	333	0.175	117.9	565,93	1.47	7.29
RF8	459	0.18	282.02	568,11	1.27	11.62

The described ML and CNN algorithms were utilized in the study in two training zones and tested for another zone, utilizing all of the parameters mentioned. To account for geometric inconsistencies between the fieldwork samples and the satellite data, we deleted those indicated landslide items that were smaller than 70 pixels from all experiments. The best thresholds were utilized for the CNN approaches. Table 1 shows the lowest, maximum, total, mean, and standard deviation detection maps as a result of statistical analysis.

In this section, we outline some accuracy assessment methods that are common and widely used in the remote sensing and computer vision domains, and which were used to evaluate the effectiveness and performance of the applied methods of MLs and CNNs by analyzing the conformity between the landslide inventory dataset and the products of the applied methods by analyzing the conformity between the landslide inventory dataset and the products of the applied methods. The total number of retrieved polygons Building areas, landslide area sizes (minimum and maximum), total identified areas, mean and standard deviation values As a result, any uncertainty in the distribution, position, and boundaries of the regions designated as landslides in the inventory data set has an impact on the accuracy assessment processes' outcomes [13].

This study demonstrates the importance of choosing the right methodologies and parameters. It's not as straightforward as comparing ML approaches to CNNs, for example. As popular scientific publications and magazines may suggest, there are a variety of ways to create a CNN, and a CNN does not always outperform alternative approaches. Two separate training data sets, various numbers of layers, and different CNN depths were used in this work. It began by concentrating on spectral data. The results from this five-layer training data set were more accurate than the eight-layer training data set, which included three topographic layers as well. This was unexpected, and it's unclear whether topographical data could be used to increase overall categorization accuracy in other research regions with different settings. Although topographical information reduced overall accuracy, it was extremely useful in discriminating between settlement regions and landslide sites, which have similar spectral behavior. Because most landslides occur on steep slopes that are unlikely to be settlement sites, the slope layer proved very valuable in this case. As a result, the majority of the eight-layer results identified settlement areas from landslide zones. Because there were just a few and small settlement areas, the disadvantage of having these misclassifications associated with using eight layers was more severe in our case study than the benefit of having improved settlement detection. The impact of varied input window sizes on CNNs was investigated. Between the two training datasets, any employed CNN input window size provided equivalent results, however different CNN input window sizes resulted in varying accuracies.

Because the data has the most training instances of these classes, the baseline model is quite excellent at predicting primary and clear labels. The baseline performs relatively well on the majority of uncommon atmospheric and ground features, but nearly all of the rare features receive a 0.0 F2 score. This is the baseline model's major flaw: by failing to recognize rare labels, it is unable to distinguish the rare but crucial human incursion that we are most concerned about. To be genuinely effective, a multi-label classifier seeking to reduce deforestation should be able to distinguish mining, logging, and slash-and-burn activity. In most of these crucial areas, the ResNet model improves. The accuracy metric that was used to choose the optimum weights simply compares 1s and 0s to see how many are correct. Because it labels any given feature with a simple rounding, all or most photos will never have rare features for the accuracy calculation. As a result, our optimizer has a hard time identifying weights that will increase the accuracy above 0.96. We can account for this with our variable thresholds when making predictions, but if the model reaches a limit on optimization improvement, we won't be able to find the remaining 4% of accuracy [14].

One of the most difficult problems in computer vision is semantic segmentation. Convolutional neural networks, such as fully convolutional networks and encoder-decoder-based designs, have been effectively applied to this problem and have marginally outperformed standard computer vision algorithms. A comprehensive review of FCN-based semantic segmentation designs can be found in. Down-sampling with pooling layers is one of the primary issues when using CNNs for semantic segmentation tasks. Convolutional kernels' range of view is increased as a result, but high-frequency details in the image are lost. They employ multi-task learning to improve building footprint segmentation forecasts. The goal is to combine the semantic term with an additional term that integrates the segmentation mask's boundary information into a single loss function. The network is trained on two independent tasks to achieve this unified representation of semantic and border information. The purpose of our multi-task strategy is to use geometric aspects in network training in addition to semantic information regarding class labels. Despite the fact that there are a variety of geometric attributes that can be recovered, such as shape and edge information, we focus on the distance between pixels and constructing borders. On the Inria

Aerial Image Labeling Dataset, the proposed methodology outperforms existing methods by 9.8 %, demonstrating the usefulness of this work [15].

The F1 score is calculated using the DG-BEC, which is a harmonic average of precision and recall that combines precision and recall accuracy.

Table 2: F1 Scores of building extraction results

Method	**Las Vegas**	**Paris**	**Shanghai**	**Khartoum**	**Total F1 Score**
Mask R-CNN	0.881	**0.760**	**0.646**	**0.578**	**0.717**
XD_XD	**0.885**	0.745	0.597	0.544	0.683
Nofto	0.787	0.584	0.520	0.424	0.579
DG-BEC	0.879	0.753	0.642	0.568	0.713
Wleite	0.829	0.679	0.581	0.483	0.643

This method clearly results in a more accurate portrayal of building footprints with more consistent bounds. The accuracy and completeness of building extraction, as well as the effect of building boundary regularization on building extraction, were used to arrive at these conclusions. Because both SpaceNet Challenge and DG-BEC employ the same data, assessment metric, and evaluation code, SpaceNet Challenge was chosen as a baseline result. Table 2 compares the F1 scores of the Mask R-CNN and DG-BEC methods. The MASK R-CNN approach has a little higher F1-score than the DG-BEC method; nonetheless, the two scores from both methods are nearly identical. DG-BEC approaches, on the other hand, reduced the number of polygon points by 86%, greatly simplifying the created polygons. Many apps may directly import these polygons.

CONCLUSION

In this research, we looked at and compared various machine learning and deep learning methods and analyses the findings of multiple studies conducted by different researchers. In our experiments, we discovered that CNNs only outperform ANN, SVM, and RF in ideal conditions. CNNs with only spectral information and a 16-pixel input window size produced the best results in this example. For the recognition of complicated picture patterns and semantic classifications, CNN may be the most efficient. We also discovered that for the same CNN structure, higher input window sizes resulted in significantly lower accuracies. The use of deeper CNN layer architectures can be beneficial. We may infer that CNNs have a lot of potential, but findings from crowdsourcing campaigns in computer vision should not be used to make decisions. Other deep-learning object detection approaches require less human supervision than previous methods and may be quickly adapted to other regions by retraining the model with new training data

On multi-label categorization of satellite images, deep CNN models created for the ImageNet Challenge paired with task-specific refinement layers offer good results. The accuracy metric that was used to choose the optimum weights simply compares 1s and 0s to see how many are correct. The optimizer has a hard time finding weights that will increase the accuracy above 0.96. We can account for this with our variable thresholds when making predictions, but if the model reaches a limit on optimization improvement, we won't be able to find the remaining 4% of accuracy.

Then we look at an approach for extracting buildings that combines Mask R-CNN with building boundary regularization. The suggested approach and Mask R-CNN obtained nearly identical F1 scores, which is the DG-BEC evaluation metric. However, unlike Mask R-CNN, which produces irregularly shaped polygons, our approach yields regularized polygons that can be used in a variety of cartographic and engineering applications.

REFERENCE

[1] Asokan, Anju; J. Anitha, “Machine Learning based Image Processing Techniques for Satellite Image Analysis -A Survey” [IEEE 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), pp. 119–124, February 2019.

[2] O. Ghorbanzadeh, T. Blaschke, K. Gholamnia, S. R. Meena, D. Tiede, & J. Aryal, “Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection” Remote Sensing, 11(2), 196, pp. 1-21, January 2019.

[3] H. Miyazaki, K. Kuwata, W. Ohira, Z. Guo, X. Shao, Y. Xu, & R. Shibasaki, “Development of an automated system for building detection from high-resolution satellite images” 4th International Workshop on Earth Observation and Remote Sensing Applications (EORSA) IEEE, pp. 245-249, July 2016.

[4] T. Postadjian, A. Le Bris, H. Sahbi, & C. Malle, “Domain adaptation for large scale classification of very high resolution satellite images with deep convolutional neural networks” In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 3623-3626, July 2018.

[5] K. Zhao, J. Kang, J. Jung, & G. Sohn, “Building extraction from satellite images using mask R-CNN with building boundary regularization” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 247-251, 2018.

[6] S. Park, C. Choi, B. Kim, & J. Kim, "Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea" Environmental earth sciences, 68(5), pp. 1443-1464, August 2012.

[7] M. Hajeb, S. Karimzadeh, & M. Matsuoka, "SAR and LIDAR datasets for building damage evaluation based on support vector machine and random forest algorithms—A case study of Kumamoto earthquake, Japan" Applied Sciences, 10(24), 8932, pp. 1-18, December 2020.

[8] B. Melville, A. Lucieer, & J. Aryal, "Object-based random forest classification of Landsat ETM+ and WorldView-2 satellite imagery for mapping lowland native grassland communities in Tasmania, Australia" International journal of applied earth observation and geoinformation, 66, pp. 46-55, 2018.

[9] D. Strigl, K. Kofler, & S. Podlipnig, "Performance and scalability of GPU-based convolutional neural networks" In 2010 18th Euromicro Conference on Parallel, Distributed and Network-based Processing, IEEE, pp. 317-324, Frbruary 2018.

[10] C. Zhang, I. Sargent, X. Pan, H. Li, A. Gardiner, J. Hare, & P. M. Atkinson, "An object-based convolutional neural network (OCNN) for urban land use classification" Remote sensing of environment, 216, pp. 57-70, June 2018.

[11] D. Marmanis, K. Schindler, J. D. Wegner, S. Galliani, M. Datcu, & U. Stilla, "Classification with an edge: Improving semantic image segmentation with boundary detection" ISPRS Journal of Photogrammetry and Remote Sensing, 135, pp. 158-172, 2018.

[12] K. He, G. Gkioxari, P. Dollár, & R. Girshick, "Mask r-cnn" In Proceedings of the IEEE international conference on computer vision, pp. 2961-2969, 2017.

[13] M. R. Mezaal, B. Pradhan, & H. M Rizeei, "Improving landslide detection from airborne laser scanning data using optimized Dempster–Shafer" Remote Sensing, 10(7), 1029, pp. 1-26, June 2018.

[14] D. Gardner, & D. Nichols, "Multi-label classification of satellite images with deep learning", pp. 1-7, 2017.

[15] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, & J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation" arXiv preprint arXiv:1704.06857, pp. 1-23, 2017.