IJARCCE

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101207

DEVELOPING AN EFFECTIVE MODEL FOR THE SEMANTIC SEGMENTATION OF REMOTE SENSING IMAGERY

Muazu Aminu Aliyu[1], Souley Boukari[2], Abdullahi Gamsha Madaki[3], Mustapha Lawal Abdulrahman[4]

[1,2,3,4] Department of Mathematical Science Abubakar Tafawa Balewa university Bauchi, Nigeria

Abstract This paper introduces an effective semantic segmentation of satellite imagery using 3D-Unet. We discussed previous work carried out and used deep learning to reproduce the work of Kemker et al. (2018) and other methods. We used performance metric to compare the performance of the proposed. This work underline that most RS and DL segmentation can be enhanced using DL models.

Keywords: Computer Vision, Deep Learning, Semantic Segmentation**,** 3D-Unet, Satellite Imagery.

1. INTRODUCTION

In order to make informed decisions concerning the environment, there are government ministries and agencies that are equipped to observe it. This enables us to make effective changes around us as appropriate or desirable. This process is referred to as earth observation, and has applications in disaster response, resource management and precision farming among others. Earth observation data is gathered by a range of techniques, and can be roughly categorized as remote sensing. It is where "the distance between the object and the sensor far surpasses the linear dimensions of the sensor Shivaprakash (2016). In the last decade, manual analyses of satellite imagery were feasible primarily because the volume of images available was quite low - but that is not the case now. Relevant information extraction from images thus becomes a problem with the high volume of data we deal with today. A major component of these problems is annotation (or labeling), wherein one identifies the structures and patterns visible in a satellite image. Over the years, research in the computer vision community has addressed this problem of automating the analysis of large-scale data in different ways. Machine learning techniques have proven to be strong candidates here, especially in the last few years Shivaprakash (2016). In this work, pixel-wise classification and semantic segmentation are synonymous. Semantic segmentation is the term more commonly used in computer vision and is becoming increasingly used in remote sensing. State-of-the-art semantic segmentation frameworks for RGB colors imagery are trained end-to-end and consist of convolution and segmentation sub-networks. Semantic segmentation (also named image classification in remote sensing) is the pixelwise classification of an image and is an important task for numerous applications of object recognition. With the rapid development of Remote Sensing (RS) technology, the RS imagery produced by high-resolution remote sensing satellites (such as IKONOS, SPOT-5, World View and Quick bird) have more abundant information to extract features and recognition ground object than the low-resolution remote sensing imagery. Many artificial objects that are difficult to be recognized in the past are now available to be detected Yuan et al. (2020). Semantic segmentation has been widely studied in computer vision (CV) and remote sensing, mainly using shallow features that were hand engineered by skilled people who have experience in the field and also often required domain-expertise Girshick et al (2014).

This also means that if the conditions change even slightly, a framework which works well in a given task may fail in another task and the whole feature extractor might have to be rewritten from scratch, which is very time-consuming and expensive. These disadvantages led researchers in the field looking for a more robust and effective approach Wu et al. (2019). At the time of writing, the state of the art in the automation of visual labeling tasks is seen in the deep learning research community, and that is where this thesis picks up at.

Semantic segmentation of remote sensing imagery has been employed in many applications and is a key research topic for decades. With the success of deep learning methods in the field of computer vision, researchers have made a great effort to transfer their superior performance to the field of remote sensing image analysis Yuan et al. (2020). In the past decade, deep learning methods have demonstrated much superior performance in many traditional computer vision applications including object classification, and semantic segmentation. Deep learning methods automatically derive features that are tailored for the targeted classification tasks, which makes such methods better choices for handling complicated scenarios. The great success in other domains excited the adoption and extension of deep learning methods for the problems in the field of remote sensing Yuan et al. (2020). Despite decades of efforts, literature review. The work reveals the following major challenges that require investigation and development of novel methods: (1) demand for pixel-level accuracy, (2) analysis of non-conventional data, and (3) lack of training examples. This opens up a significant room for further investigations to address the aforementioned challenges.

The existing models base on deep convolutional neural network for the semantic segmentation of satellite imagery has shown that synthetic imagery can be used to assist the training of end-to-end semantic segmentation frameworks when there is not enough annotated image data. The network initialization scheme has been shown to increase semantic segmentation performance when compared to traditional classifiers and unsupervised feature extraction techniques. However, there exist some more open issues as suggested by the author Kemker, et al. (2018) that requires future researchers to address such as, exploring deeper CNN architecture such as ResNet and U-Net models base on newer state-of-the-art convolution deep learning architectures and segmentation model by Including additional diverse classes to the synthetic data. Thus, this work hopes to solve semantic segmentation of RS imagery using 3D-UNET.

The remainder of this paper is organized as follows. Section 2 present the review of related work, section 3. Present the methodology, section 4 presents the results and finally section conclude the research findings proffer future work.

2. RELATED WORK

Pixel-wise classification and semantic segmentation are synonymous. Semantic segmentation is the term more commonly used in computer vision and is becoming increasingly used in remote sensing. State-of-the-art semantic segmentation frameworks for RGB imagery are trained end-to-end and consist of convolution and segmentation sub-networks. Recently, deep learning (DL) has become the fastest-growing trend in big data analysis and has been widely and successfully applied to various fields of computer application successfully including sequential data, processing of natural language, speech recognition and image classification Abdel-Hamid et al. (2012), because of its outstanding performance compared with that of traditional learning algorithms. Standing at the paradigm shift towards data-intensive science, machine learning techniques are becoming increasingly important. In particular, as a major breakthrough in the field, deep learning has proven as an extremely powerful tool in many fields.

Praveena and Singh (2015) proposed a hybrid clustering algorithm and feed-forward neural network classifier for land-cover mapping of trees, shade, building and road. The proposed technique performed better than all the existing algorithms taken for comparison. However, the results indicate that Moving KFCM is performed better than the existing algorithms for shade region classification. Additionally, an effective deep neural network was also proposed by Shuang et al. (2018) and compare the performance with SIFT, SURF, SAR-SIFT, PSO-SIFT, the experimental results shows that the applied transfer learning further improves the accuracy and reduces the training cost. However, the major limitation of this research is that there is no unified feature representation for different source images.

Jony et al. (2018) employs an ensemble classifier to detect water in satellite images for flood assessment and evaluate it against Mediaeval (2017), it was found that this approach is capable of producing good classification accuracy for a seen location when bands are used and an unseen location when NDWI is used. However, the major setback of the study was that the study achieved worse results on an unseen location.

More so, Wang et al, (2018) proposed a novel convolutional neural network (CNN) to classify cloud and snow on an object level. Specifically, a novel CNN structure capable of learning cloud and snow multiscale semantic features from high-resolution multispectral imagery is presented. In order to solve the shortcoming of "salt-and-pepper" in pixel level predictions, the author extend a simple linear iterative clustering algorithm for segmenting high-resolution multispectral images and generating super pixels. Results demonstrated that the new proposed method can with better precision separate the cloud and snow in the high-resolution image, and results are more accurate and robust compared to the other methods. However, the study fails to generalize the proposed convolutional neural network-based methods to another task in remote sensing fields such as urban water extraction and ship detection.

The study Kemker et al. (2018) demonstrated the utility of FCN architectures for the semantic segmentation of remote sensing MSI by proposing an end-to-end segmentation model, which uses a combination of convolution and pooling operations, is capable of learning global relationships between object classes more efficiently than traditional classification methods. The result showed that an end-to-end semantic segmentation framework provided superior classification performance on fourteen of the eighteen classes in RIT-18. However, the result can be improved via exploring deeper ResNet and U-Net models. Additionally, including additional diverse classes to the synthetic data should aid the development of more discriminative frameworks that yield superior performance.

In the past, remote sensing has proven to be an extremely valuable and effective tool for mapping

slums. The study in Wurm et al. (2019) aimed at analyzing transfer learning capabilities of FCNs to slum mapping in various satellite images. A model trained on very high-resolution optical satellite imagery from QuickBird is transferred to Sentinel-2 and TerraSAR-X data. While free-of-charge Sentinel-2 data is widely available, its comparably lower resolution makes slum mapping a challenging task. TerraSAR-X data on the other hand, has a higher resolution and is considered a powerful data source for intra-urban structure analysis. Due to the different image characteristics of SAR compared to optical data, however, transferring themodel could not improve the performance of semantic segmentation but we observe very high accuracies for mapped slums in the optical data to address the challenges in VHR image semantic segmentation performance. Mi and Chen (2020) proposed a Superpixel-enhanced Deep Neural Forest (SDNF). A fully differentiable forest is introduced to dominate the representation learning of deep convolutional layers in order to balance the classification ability and representation learning capability of DCNNs. Moreover, a Superpixel-enhanced

Region Module (SRM) is proposed to alleviate the classification noises and strengthens edges of ground objects. The efficiency of SDNF is evaluated on the ISPRS 2D labeling benchmark. Experimental results demonstrate that our method reaches a new state-of-the-art performance.

Similarly, Wu et al. (2019) proposed an FCN-based model is to implement pixel-wise classifications for remote sensing image in an end-to-end way, and an adaptive threshold algorithm is proposed to adjust the threshold of Jaccard index in each class. Experiments on DSTL dataset show that the proposed method produces accurate classifications in an end-to-end way. Results show that the adaptive threshold algorithm can increase the score of average Jaccard index from 0.614 to 0.636 and achieve better segmentation results. However, the major limitation of the research is training the model in a weak supervision way, to further enhance its applicability.

In 2020, You et al. (2020) propose a weed/crop segmentation network that provides better performance for precisely recognizing the weed with arbitrary shape in complex environment condition, and offers great support for autonomous robots to successfully reduce the density of weed. the deep neural network (DNN)-based segmentation model obtains persistent improvements by integrating four additional components by evaluating the network performance on two challenging Stuttgart and Bonn datasets. The state-of-the-art performance on the two datasets shows that each added component has notable potential to boost the segmentation accuracy. However, the research fails to exploit the domain knowledge in weed detection, and model the network to learn more correlated spatial information.

The application of drones has recently revolutionized the mapping of wetlands due to their high spatial resolution and the flexibility in capturing images. Hence, Bhatnagar et al. (2020)

proposed mapping of vegetation in wetlands using image segmentation. For this, ML and DL algorithms were compared by applying them to a set of drone images of Clara Bog, a raised bog located in the middle of Ireland. Overall, the accuracy of the DL was approximately 4% higher than the ML methods. Additionally, the DL method does not require any colour correction or the addition of extra textural features. However, DL requires a large amount of initial labelled training data (approximately 48 x 106 pixels).

So far, from this survey, it is noticed that nearly all the existing CNN architecture used for the semantic segmentation of multispectral images are based on 1D or 2D architectures. The 2D convolutional kernels are able to leverage context across the height and width of the slice to make predictions. However, because 2D CNNs take a single slice as input, they inherently fail to leverage context from adjacent slices. Voxel information from adjacent slices may be useful for the prediction of segmentation maps. 3D CNNs address this issue by using 3D convolutional kernels to make segmentation predictions for a volumetric patch of a images. The ability to leverage interslice context can lead to improved performance (Hamida, Benoit, Lambert, & Ben-Amar, 2016).

3. METHODOLOGY

In this research, pixel-wise classification and semantic segmentation are synonymous. Semantic segmentation is the term more commonly used in computer vision and is becoming increasingly used in remote sensing. State-of-the-art semantic segmentation frameworks for RGB imagery are trained end-to-end and consist of convolution and segmentation sub-networks. The goal of this study is to design and implement a remote sensing image segmentation model using Convolutional Neural Network (CNN) specifically the deeper U-Net architecture. In other to address the issues of the existing work, this research will include additional diverse classes to the synthetic data about each pixel using the Hamlin Beach State Park data set which is then use to train a U-Net convolutional neural network to perform semantic segmentation of a multispectral image with seven channels: three color channels, three near-infrared channels, and a mask. Additionally, the research will use the deep-learning-based semantic segmentation techniques to calculate the percentage vegetation cover in a region from a set of multispectral images. Finally, the performance of the proposed method will be evaluated against state-of-the-art approaches.

3.1 Convolutional Neural Network

This proposed work will employ 3D U-Net deep CNN architecture to perform semantic segmentation of remote sensing image by including additional diverse classes to the synthetic data. The general structure of a CNN is the combination of two components: The feature extractor in the first stage and the classifier as shown in Fig. 1.

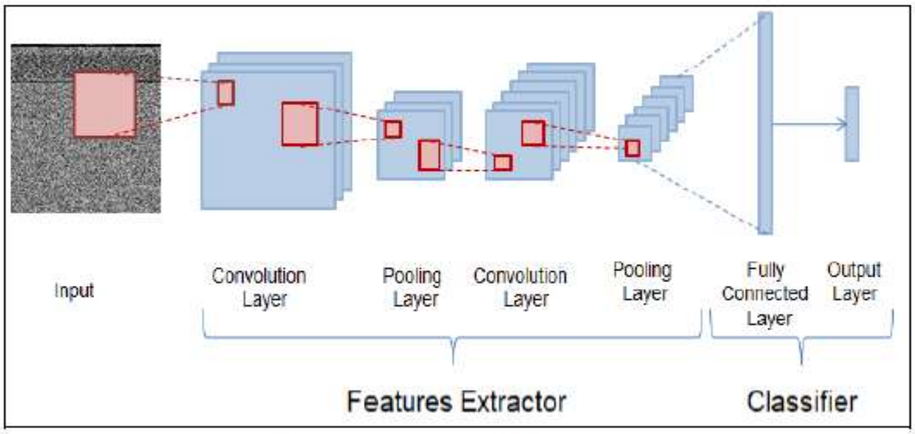

Figure 1. Convolutional Neural Network NN architecture. Kemker et al. (2018)

CNNs for segmentation can be categorized based on the dimension of convolutional kernel that is utilized. 2D CNNs use 2D convolutional kernels to predict the segmentation map for a single slice. Segmentation maps are predicted for a full volume by taking predictions one slice at a time. The 2D convolutional kernels are able to leverage context across the height and width of the slice to make predictions. However, because 2D CNNs take a single slice as input, they inherently fail to leverage context from adjacent slices. Voxel information from adjacent slices may be useful for the prediction of segmentation maps. 3D CNNs address this issue by using 3D convolutional kernels to make segmentation predictions for a volumetric patch of a images. The ability to leverage interslice context can lead to improved performance but comes with a computational cost as a result of the increased number of parameters used by these CNNs.

3.2 3D Convolutional Neural Network

This research proposed a deep network that learns to generate dense volumetric segmentations, but only requires some annotated 2D slices for training. This network can be used in two different ways as depicted in Fig. 2 the first application case just aims on densication of a sparsely annotated data set; the second learns from multiple sparsely annotated data sets to generalize to new data. Both cases are highly relevant. The network is based on the previous u-net architecture, which consists of a contracting encoder part to analyze the whole image and a successive expanding decoder part to produce a full-resolution segmentation. While the u-net is an entirely 2D architecture, the network proposed in this research takes 3D volumes as input and processes them with corresponding 3D operations, in particular, 3D convolutions, 3D max pooling, and 3D up-convolutional layers. Moreover, we avoid bottlenecks in the network architecture and use batch normalization for faster convergence. The architecture of proposed 3d U-Net deep CNN is presented in Figure 2.

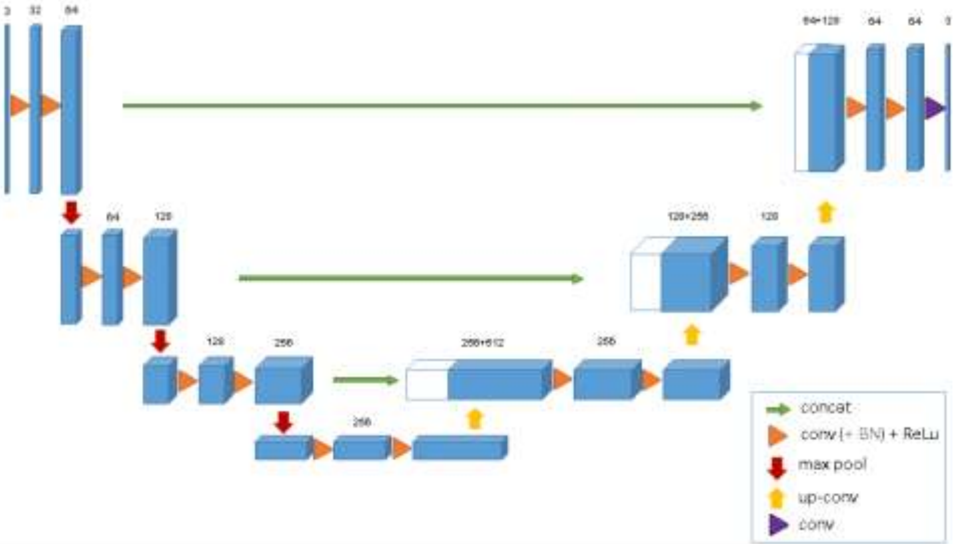

Figure 2. The 3D u-net architecture Cicek Et al. (2016)

As a solution for the challenges presented in the previous sections, we introduce a new three-dimensional based architecture that is dedicated to hyperspectral images and tackles most of the DL for RS aspects of difficulty. This research

DOI: 10.17148/IJARCCE.2021.101207

proposes to use a new 3D CNN architecture that, unlike the previously mentioned approaches, simultaneously processes the spatial and spectral components with real 3D convolutions giving better investments of the few samples available with fewer trainable parameters. This proposal decomposes the problem as the processing of a series of volumetric representations of the image. Therefore, each pixel is associated to an n×n spatial neighborhood and a number of f spectral bands. As a result, each pixel is treated as a n×n×f volume. The main concept behind this architecture is to combine the traditional CNN network with a twist of applying 3D convolution operations instead of using 1D convolution operators that only inspect the spectral content of the data. An overview of the 3D architecture is presented in Figure 2.

Different blocks of CNN layers are stacked on top of each other in order to ensure deep efficient representations of the image. Firstly, a 3D convolution-based set of layers is introduced in order to cope with the three-dimensional input voxels. Each and every one of these layers encompass a number of volumetric kernels that simultaneously execute convolutions on the width, height and depth axis of the input. Such 3D convolutions stack is followed by a set of 1 × 1 convolution (1D) layers that discards the spatial neighborhood and a series of Fully Connected layers. Basically, the proposed architecture considers 3D voxels as input data and first generates 3D feature maps that are gradually reduced into 1D feature vectors all along the layers. This procedure is ensured by the choice of specific configurations of the convolution filter strides and paddings.

3.3 Proposed Approach

Figure 2 illustrates the network architecture. Like the standard u-net, the 3D U-Net has an analysis and a synthesis path each with four resolution steps. In fact, each pixel is taken into account as a 3D n×n×f voxel in the research contex. In the analysis path, each layer contains two 3x3x3 convolutions each followed by a rectified linear unit (ReLu), and then a 2x2 x2 max pooling with strides of two in each dimension. In the synthesis path, each layer consists of an up convolution of 2x2x2 by strides of two in each dimension, followed by two 3x3x3 convolutions each followed by a ReLu. Shortcut connections from layers of equal resolution in the analysis path provide the essential high-resolution features to the synthesis path. In the last layer a 1x1x1 convolution reduces the number of output channels to the number of labels which is 7 in our case. The architecture has 19069955 parameters in total. we will avoid bottlenecks by doubling the number of channels already before max pooling. We also adopt this scheme in the synthesis path. The input to the network is a N x N x F voxel tile of the image with 7 channels. Our output in the final layer is N x N x F voxels in x, y, and z directions respectively. With a voxel size of 1:76x1:76x2:04xm3, the approximate receptive field becomes 155x155x180xm^3 for each voxel in the predicted segmentation. Thus, each output voxel has access to enough context to learn efficiently.

We also introduce batch normalization (\BN") before each ReLU. each batch is normalized during training with its mean and standard deviation and global statistics are updated using these values. This is followed by a layer to learn scale and bias explicitly. At test time, normalization is done via these computed global statistics and the learned scale and bias. However, we have a batch size of one and few samples. In such applications, using the current statistics also at test time works the best. The important part of the architecture, which allows us to train on sparse annotations, is the weighted SoftMax loss function. Setting the weights of unlabeled pixels to zero makes it possible to learn from only the labelled ones and, hence, to generalize to the whole volume.

3.4 Network Training

In order to train the network, first we use a random patch extraction datastore to feed the training data to the network. This datastore extracts multiple corresponding random patches from an image datastore and pixel label datastore that contain ground truth images and pixel label data. Patching is a common technique to prevent running out of memory for large images and to effectively increase the amount of available training data. We will train the network using stochastic gradient descent with momentum (SGDM) optimization by Specifying the hyperparameter settings for SGDM by using the training Options (Deep Learning Toolbox) function.

Training a deep network is time-consuming. Hence, we will accelerate the training by specifying a high learning rate. However, this can cause the gradients of the network to explode or grow uncontrollably, preventing the network from training successfully. We will keep the gradients in a meaningful range, and then enabling gradient clipping by specifying 'Gradient Threshold' as 0.05, and specify Gradient Threshold Method to use the L2-norm of the gradients.

To create the 3D U-Net Layer, this research will use a variation of the 3D U-Net network. In U-Net, the initial series of convolutional layers are interspersed with max pooling layers, successively decreasing the resolution of the input image. These layers are followed by a series of convolutional layers interspersed with up sampling operators, successively increasing the resolution of the input image. The name U-Net comes from the fact that the network can be drawn with a symmetric shape like the letter U. This research will modify the U-Net to use zero-padding in the convolutions, so that the input and the output to the convolutions have the same size. Hence, we will create a U-Net with a few preselected hyperparameters.

3.5 Datasets

This research will use a high-resolution multispectral data set to train the network. The image set was captured using a drone over the Hamlin Beach State Park, NY. The data contains labeled training, validation, and test sets, with 18 object class labels. The size of the data file is 3.0 GB.

3.6 Choice of Metric

This research will adopt accuracy to evaluate the performance of the proposed model: this performance metric deals with the correct prediction made by the model and this metric can be expressed as:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \qquad (1)$$

4. IMPLEMENTATION OF THE SYSTEM

The experiment was conducted on MATLAB 2021a running a Microsoft Window 10 with 64-bit Operating System, 8GB (RAM) and Intel(R) Core (TM) i7-4000M @ 2.4 GHz. To implement the proposed model, this research uses a high-resolution multispectral data set to train the network. The image set was captured using a drone over the Hamlin Beach State Park, NY. The data contains labeled training, validation, and test sets, with 18 object class labels as shown in Fig. 3. The size of the data file is ~3.0 GB.

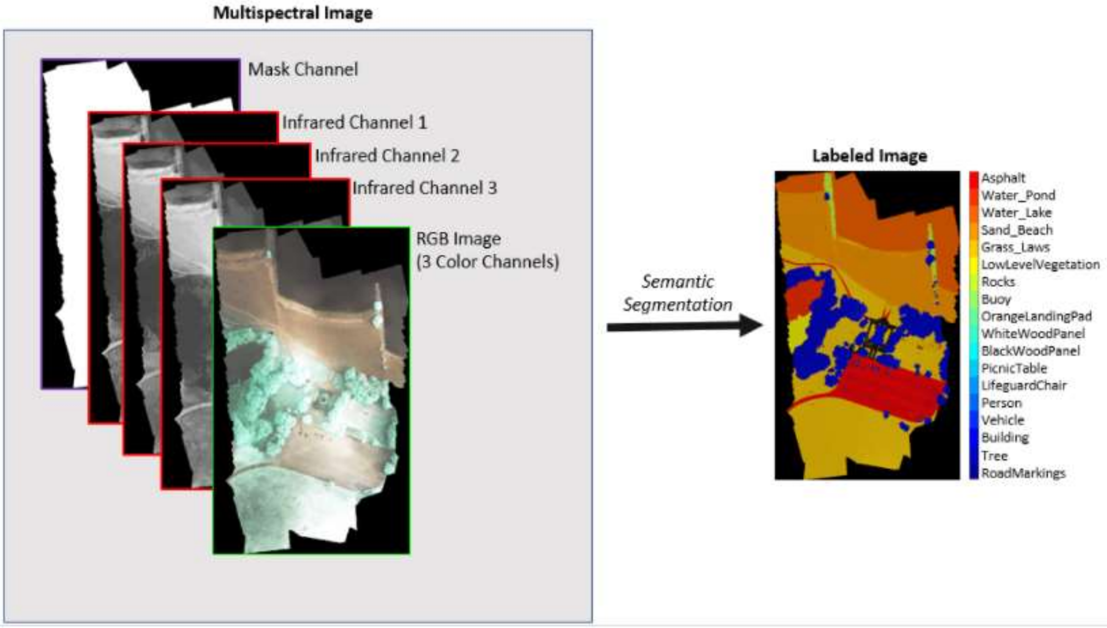

Figure 3: Multispectral images with 18 object class labels.

The multispectral image data is arranged as num Channels-by-width-by-height arrays. To reshape the data so that the channels are in the third dimension. The RGB color channels are the 3rd, 2nd and 1st image channels shown in Figure 4 displaying the color component of the training, validation, and test images as a montage.

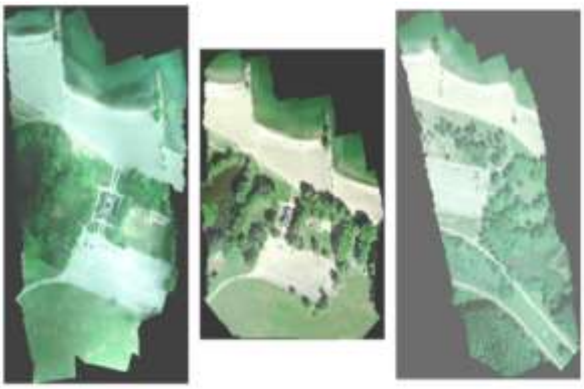

Figure 4: RGB component of training image (left), validation image (center) and test image (right)

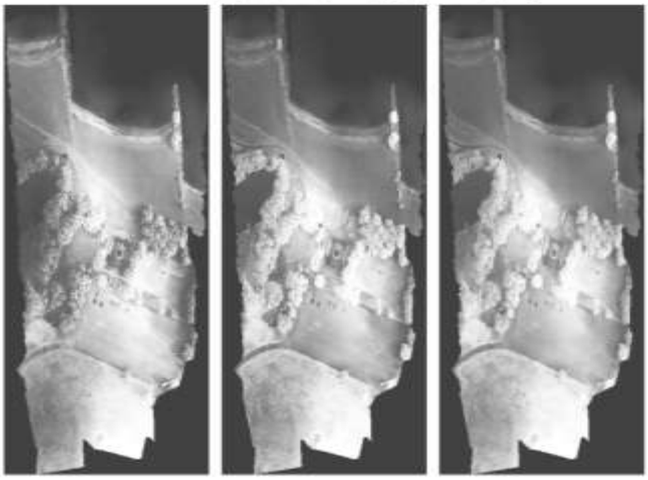

Figure 5: IR channel 1 (left), 2 (center) and test image (right)

Figure 6: Mask of training image (left), validation image (center) and test image (right)

The labeled images contain the ground truth data for the segmentation, with each pixel assigned to one of the 18 classes. Table 1 shows the 18 classes and their IDs.

Table 1. Image Classes and IDs for the rits18 datasets.

IDs	Class Name
0.	Other Class/Image Border
1.	Road Markings
2.	Tree
3.	Building
4.	Vehicle (Car, Truck, or Bus)
5.	Person
6.	Lifeguard Chair
7.	Picnic Table
8.	Black Wood Panel
9.	White Wood Panel
10.	Orange Landing Pad
11.	Water Buoy
12.	Rocks
13.	Other Vegetation
14.	Grass
15.	Sand

16.	Water Lake
17.	Water Pond
18.	Asphalt (Parking Lot/Walkway)

Figure 4.5 depict the training labels for the 18 classes.

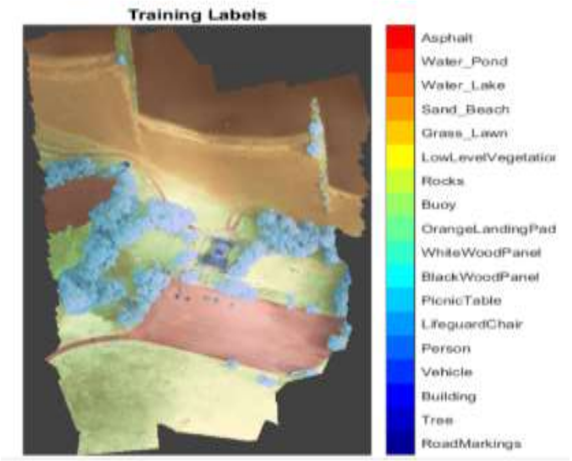

Figure 7: Training labels for Rits18 datasets.

In order to create random patch extraction datastore for Training, we use a random patch extraction datastore to feed the training data to the network. This datastore extracts multiple corresponding random patches from an image datastore and pixel label datastore that contain ground truth images and pixel label data. Patching is a common technique to prevent running out of memory for large images and to effectively increase the amount of available training data.

To create the U-Net network layers, this research uses a variation of the U-Net network. In U-Net, the initial series of convolutional layers are interspersed with max pooling layers, successively decreasing the resolution of the input image. These layers are followed by a series of convolutional layers interspersed with up sampling operators, successively increasing the resolution of the input image. The name U-Net comes from the fact that the network can be drawn with a symmetric shape like the letter U. We train the network using stochastic gradient descent with momentum (SGDM) optimization. We Specify the hyperparameter settings for SGDM as depicted in Table 3. Training a deep network is time-consuming. Hence, we accelerate the training by specifying a high learning rate. However, this can cause the gradients of the network to explode or grow uncontrollably, preventing the network from training successfully. To keep the gradients in a meaningful range we enable gradient clipping. To quantify segmentation accuracy, we create a pixel Label Datastore for the segmentation results and the ground truth labels. The final goal of this research is to calculate the extent of vegetation cover in the multispectral image by dividing the number of vegetation pixels by the number of valid pixels. The settings for hyper parameters and training option is depicted in Table. 2.

Table. 2 Parameter settings

Parameters	Settings
Initial Learning Rate	0.05
Max Epochs	150
Mini batch Size	16
l2reg	0.0001
Momentum	0.9
Learn Rate Schedule	piecewise
Shuffle	every-epoch
Gradient Threshold Method	l2norm
Gradient Threshold	0.05
Verbose Frequency	20

Hence, we can now use the developed 3D-U-Net to semantically segment the multispectral image. The prediction results in terms of accuracy are presented in the preceding section.

4.1 Result Presentation and Discussion

To perform the forward pass on the trained network and performs segmentation on image patches using the semantics function. The typical sample of the segmented image is presented in Figure 4.6.

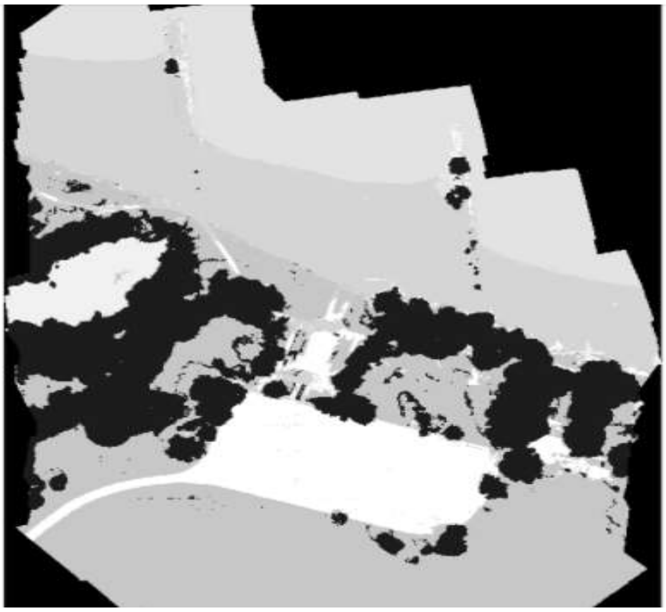

Figure 8: Segmented Image.

The segmented image and ground truth labels are saved as PNG files. These will be used to compute accuracy metrics by overlaying the segmented image on the histogram-equalized RGB validation image as shown in Figure 9.

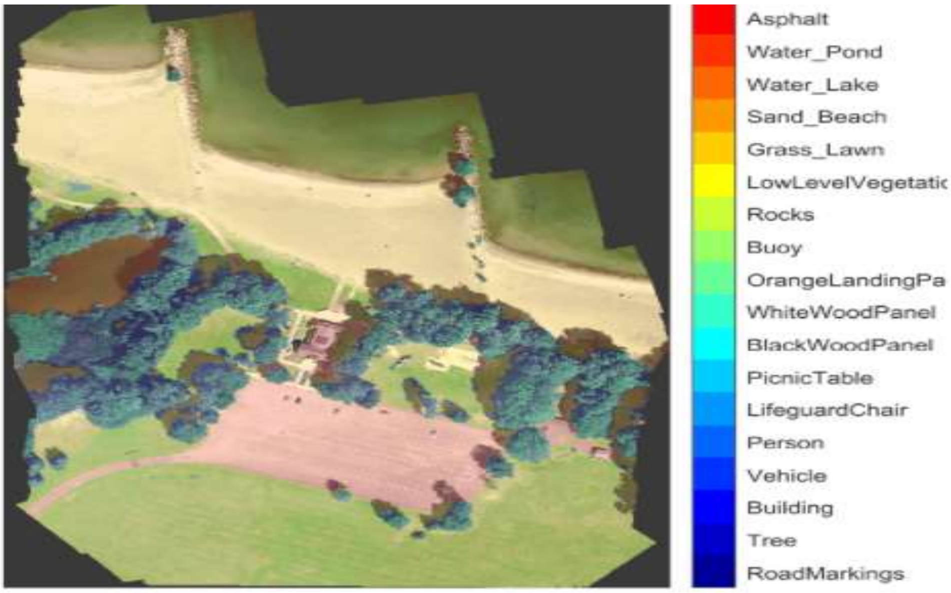

Figure 9: label Validation Image

After overlaying the segmented image, Measure the global accuracy of the semantic segmentation as shown in Table 3. The final goal of this research is to calculate the extent of vegetation cover in the multispectral image by dividing the number of vegetation pixels by the number of valid pixels. For this experiment, the total vegetation cover from the segmented image is 51.72%

We evaluate the performance of the proposed model against the benchmark classification framework on RIT-18 datasets. Table 3 depicts the mean-class accuracy (AA) on the RIT-18 test set compare with other existing studies.

Table 3. Performance comparison base on mean-class accuracy against Kemker et al. (2018).

Model	Mean Accuracy (%)
Proposed	90.698
MLP	30.4
KNN	27.7

SVM	29.6
Sharp Mask	57.3
Refine-Net	59.8

From table 4. It is quite obvious that the deep learning algorithms outperformance the conventional approach in terms of accuracy. This result is further depicted in Figure 10 in a graphical representation for clear understanding on how the difference in performance trend continues.

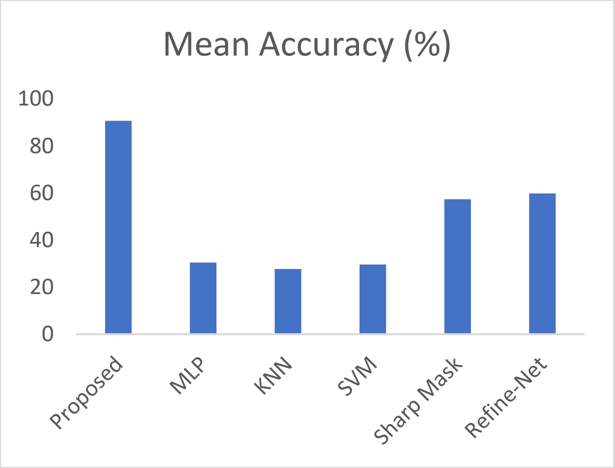

Figure 10: Performance comparison base on mean class accuracy on the RIT-18 test set

From Fig. 10 it is observed that the proposed model achieved the best performance compare to the other approaches. For mean class accuracy metric, the higher the value, the better the better the segmentation performance of the model, in this case, it can be observed that the proposed model achieved 90.698% accuracy which is the best score when compare to existing approach. However, it is also notice that, the sharp mask and Refine-Net proposed by Kemker et al. (2018). where superior than the other machine learning algorithms. The Refine -Net deep learning algorithm achieved accuracy of 59.8% follow by the sharp mask which achieved accuracy of 57.3%. the worst performance was observed in the case of KNN with 27% accuracy, SVM with 29.9% accuracy and MLP with 30.4%. this experiment further demonstrates the robustness of deep learning algorithms over conventional machine learning algorithms. This performance improvement can be attributed to the features of deep learning algorithms such as higher layers of abstraction that makes them suitable in modelling complex image processing task as stated earlier in the literature.

5. CONCLUSION

This research has shown that using a new 3D CNN architecture that is dedicated to multispectral images can tackles most of the difficulty in DL for RS aspects, by employing newer end-to-end DCNN segmentation frameworks via 3D U-Net convolutional neural network to perform semantic segmentation of a multispectral image with seven channels. It is belief that these techniques can aid the development of more discriminative frameworks that yield superior performance. Unlike previous studies, the research also calculates the vegetation cover of the segmented image. Hence, the results in this study show that this improved model offer more to image processing in the context of remote sensing and satellite image enhancement.

One of the major limitations of this research is that, this research focuses on mean segmentation accuracy as the key evaluation metric of the models, there is need for further studies to measure the computing time of each model as an index for evaluating the quality of the models.

REFERENCES

[1]. Abdel-Hamid, O., Mohamed, A.-r., Jiang, H., & Penn, G. (2012). *Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition.* Paper presented at the 2012 IEEE international conference on Acoustics, speech and signal processing (ICASSP).

[2]. Bhatnagar, S., Gill, L., & Ghosh, B. (2020). Drone image segmentation using machine and deep learning for mapping raised bog vegetation communities. *Remote Sensing, 12*(16), 2602.

[3]. Bre, F., Gimenez, J. M., & Fachinotti, V. D. (2018). Prediction of wind pressure coefficients on building surfaces using artificial neural networks. *Energy and Buildings, 158*, 1429-1441.

[4]. Eleyan, A. (2012). *Breast cancer classification using moments.* Paper presented at the 2012 20th Signal Processing and Communications Applications Conference (SIU).

[5]. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). *Rich feature hierarchies for accurate object detection and semantic segmentation.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

[6]. Goodfellow, M., & Fiedler, H.-P. (2010). A guide to successful bioprospecting: informed by actinobacterial systematics. *Antonie Van Leeuwenhoek, 98*(2), 119-142.

DOI: 10.17148/IJARCCE.2021.101207

[7]. Hamida, A. B., Benoit, A., Lambert, P., & Ben-Amar, C. (2016). *Deep learning approach for remote sensing image analysis.* Paper presented at the Big Data from Space (BiDS'16).

[8]. Han, W., Feng, R., Wang, L., & Cheng, Y. (2018). A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 145*, 23-43.

[9]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

[10]. Huang, F., Yu, Y., & Feng, T. (2019). Automatic Building Change Image Quality Assessment in High Resolution Remote Sensing Based on Deep Learning. *Journal of Visual Communication and Image Representation*, 102585.

[11]. Jony, R. I., Woodley, A., Raj, A., & Perrin, D. (2018). *Ensemble Classification Technique for Water Detection in Satellite Images.* Paper presented at the 2018 Digital Image Computing: Techniques and Applications (DICTA).

[12]. Kemker, R., Salvaggio, C., & Kanan, C. (2017). High-resolution multispectral dataset for semantic segmentation. *arXiv preprint arXiv:1703.01918*.

[13]. Kemker, R., Salvaggio, C., & Kanan, C. (2018). Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS journal of photogrammetry and remote sensing, 145*, 60-77.

[14]. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems, 25*, 1097-1105.

[15]. LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks, 3361*(10), 1995.

[16]. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278-2324.

[17]. Mi, L., & Chen, Z. (2020). Superpixel-enhanced deep neural forest for remote sensing image semantic segmentation. *ISPRS journal of photogrammetry and remote sensing, 159*, 140-152.

[18]. Neagoe, V.-E., & Neghina, C.-E. (2018). *An Artificial Bee Colony Approach for Classification of Remote Sensing Imagery.* Paper presented at the 2018 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI).

[19]. Nogueira, K., Penatti, O. A., & dos Santos, J. A. (2017). Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognition, 61*, 539-556.

[20]. Ouyang, W., Wang, X., Zeng, X., Qiu, S., Luo, P., Tian, Y., . . . Loy, C.-C. (2015). *Deepid-net: Deformable deep convolutional neural networks for object detection.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

[21]. Peng, M., Wang, C., Chen, T., & Liu, G. (2016). Nirfacenet: A convolutional neural network for near-infrared face identification. *Information, 7*(4), 61.

[22]. Praveena, S., & Singh, S. (2015). *Hybrid clusteing algorithm and Neural Network classifier for satellite image classification.* Paper presented at the 2015 International Conference on Industrial Instrumentation and Control (ICIC).

[23]. Pritt, M., & Chern, G. (2017). *Satellite image classification with deep learning.* Paper presented at the 2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR).

[24]. Radhika, K., & Varadarajan, S. (2017). *Satellite image classification of different resolution images using cluster ensemble techniques.* Paper presented at the 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET).

[25]. Romero, A., Gatta, C., & Camps-Valls, G. (2015). Unsupervised deep feature extraction for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing, 54*(3), 1349-1362.

[26]. Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-net: Convolutional networks for biomedical image segmentation.* Paper presented at the International Conference on Medical image computing and computer-assisted intervention.

[27]. Shen, R., Huang, A., Li, B., & Guo, J. (2019). Construction of a drought monitoring model using deep learning based on multi-source remote sensing data. *International Journal of Applied Earth Observation and Geoinformation, 79*, 48-57.

[28]. Shivaprakash, M. (2016). *Semantic segmentation of satellite images using deep learning.* Master's thesis (Czech Technical University in Prague & Luleå University of ….

[29]. Shuang, W., Guo, Y., Quan, D., Liang, X., Ning, M., & Jiao, L. (2018). A deep learning framework for remote sensing image registration. *ISPRS Journal of Photogrammetry and Remote Sensing, 145*(1), 148-164.

[30]. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[31]. Slavkovikj, V., Verstockt, S., De Neve, W., Van Hoecke, S., & Van de Walle, R. (2015). *Hyperspectral image classification with convolutional neural networks.* Paper presented at the Proceedings of the 23rd ACM international conference on Multimedia.

[32]. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., . . . Rabinovich, A. (2015). *Going deeper with convolutions.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

[33]. Tao, C., Pan, H., Li, Y., & Zou, Z. (2015). Unsupervised spectral–spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification. *IEEE Geoscience and Remote Sensing Letters, 12*(12), 2438-2442.

[34]. Wang, L., Chen, Y., Tang, L., Fan, R., & Yao, Y. (2018). Object-based convolutional neural networks for cloud and snow detection in high-resolution multispectral imagers. *Water, 10*(11), 1666.

[35]. Wu, Z., Gao, Y., Li, L., Xue, J., & Li, Y. (2019). Semantic segmentation of high-resolution remote sensing images using fully convolutional network with adaptive threshold. *Connection Science, 31*(2), 169-184.

[36]. Wurm, M., Stark, T., Zhu, X. X., Weigand, M., & Taubenböck, H. (2019). Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS journal of photogrammetry and remote sensing, 150*, 59-69.

[37]. Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into imaging, 9*(4), 611-629.

[38]. Yang, Z., Mu, X.-d., & Zhao, F.-a. (2018). Scene classification of remote sensing image based on deep network grading transferring. *Optik, 168*, 127-133.

[39]. You, J., Liu, W., & Lee, J. (2020). A DNN-based semantic segmentation for detecting weed and crop. *Computers and Electronics in Agriculture, 178*, 105750.

[40]. Yuan, X., Shi, J., & Gu, L. (2020). A Review of Deep Learning Methods for Semantic Segmentation of Remote Sensing Imagery. *Expert Systems with Applications*, 114417.

[41]. Zhang, H., Li, Y., Xue, X., Jiang, Y., & Shen, Q. (2018). Deep learning for remote sensing image classification: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8*(6), e1264.

[42]. Zhao, R., Yan, R., Wang, J., & Mao, K. (2017). Learning to monitor machine health with convolutional bi-directional LSTM networks. *Sensors, 17*(2), 273.

[43]. Cicek El at. (2016). 3D- learning Volumetric Segemetantion from Sparse Annotation. Researchgate Article.