International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 27, No. 2, April, 2022, pp. 285-291
Journal homepage: http://iieta.org/journals/isi

A Novel Architecture Implementation Using Multi Scale Shared Residual Network from Remote Sensing Images for Extracting Water Bodies

Chatragadda Rajyalakshmi[1*], Koritepati Ram Mohan Rao[2], Ramisetty Rajeswara Rao[3]

[1] CSE Department, JNTU Kakinada, Kakinada 533003, India
[2] NRSC, ISRO, Balanager, Hyderabad 500037, India
[3] CSE Department, JNTUK, UCEV, Vizianagaram 535003, India

Corresponding Author Email: chrajyalakshmi84@gmail.com

https://doi.org/10.18280/isi.270212

Received: 28 January 2022
Accepted: 19 April 2022

Keywords:
multi-scale shared residual network (MSSResNet), convolutional block attention module (CBAM), satellite imagery

ABSTRACT

Accurate data acquisition plays a vital role in crucial time period. Now a days so many real situations are suffering with accurate data prediction during disaster period. Flood zone identification is interlinked with extraction of water bodies to estimate disaster situation. So many limitations observed in the traditional waterbody detection methods but it is very important for the real situations. This paper proposes all the limitations of previous methods can be overcome with support of semantic segmentation of deep convolution network of multi scale shared residual network (MSSResNet) with CBAM blocks using satellite imagery. Because of residual blocks inserting in to encoder circuit reducing network degradation problems. Introducing depth wise convolutions to increase the performance of feature extraction property by using CBAM blocks with expanded receptive fields.

1. INTRODUCTION

The use of satellite image data categorization, which attempts to assign semantic categories to remote sensing images based on pixel information. Remote sensing image being processed by deep learning has attracted a lot of interest and made major achievements. To our knowledge, however, there is still no thorough evaluation of recent advancements in deep learning for image classification of remote sensing images. Convolution neural networks are used in every Deep learning model auto encoder-based remote sensing image scene classification method, and generative adversarial network-based remote sensing image scene classification methods are a vital element of any research. In this class label to each image pixel to represent high-level semantic meaning of the image [1].

To perform pixel-wise classification, traditional machine learning model uses feature descriptors [2, 3].In the previous research some of the prominent methods used by the researchers those are FCN [4] represent fully convolution network, one more DenseNet [5] define dense network to process the data, and ResNet [6]all defined DCNN models have been applied, resulting in lot of improvements such as the reuse of features layer by layer mapping of residuals blocks into a deep model to give better segmentation with very good accuracy. Technological evolutionary part invents more suitable image processing algorithms like DeepLabV3 [7] and DeepLabV3+ [8] use spatial pyramid pooling module incorporating multiscale features processing encoder and the decoder networks. Recently so many accurate models embed with hierarchical feature representation methods are SharpMask [9], U-Net [10], and RefineNet [11] are including auto encoder and decoder networks for the better segmentationof satellite images. Alternative techniques, such as hierarchical feature representations [12, 13], multi-modal CNN [14], and fusion schemes [15-17], have been used to solve all practical concerns in recent applications. With the current LinkNet [18] methodologies, the adoption of more features is expanding year after year.

1.1 Literature survey with latest water detection methods

Currently, latest methodologies methods are deep learning-based water-body detection methods, which attract increasing attention. Many general deep networks [19-22] in the field of digital vision are conveniently use for waterbody delineation process, but they do not adequately consider the characteristics of the water bodies. Feng et al. proposed one method that includes the combination of deeper U-Net with super pixel-based conditional random field [23], but the drawback of this method is that it ignores multi-scale information and is thus insufficient in complex cases. Multi-scale feature extractor added to the FCN-based backbone [24]. multi-scale refinement network (MSR-Net) to take multi-scale information and designed an erasing-attention specific module to embed features during a multi-scale refinement strategy [25]. The motivation of proposed method is to process data in an accurate manner to analyze raster data in an efficient manner.

2. DATA PRE-PROCESSING AND DATA AUGMENTATION

In the Figure 1 shows that take n input raw image data and every image includes lot of information. Any type of image format accepted by the model. For the purpose of current research taken images from various cities to delineate water feature from the images to helpful for identification of flood zone areas in an effect way. Thus, merged images taken as input for convolution network and create a high-quality noise free data set. We also performed atmospheric correction on their mages using QGIS Software.

In this research proposes a novel architecture of deep semantic segmentation network with Residual Blocks (DSSNR) called the multi scale residual network with CBAM (MSRC) by considering the various characteristics of water bodies from multi resultionary images. With Novel Architecture increasing the sensitivity of different sizes and shapes of water bodies and retrieving detailed information about water boundaries. The main aim of the proposed novel algorithm is to facilitate monitoring water areas by analysing its dynamics nature at short intervals time period. The residual connection prevents the gradient from vanishing. In addition, the centre part of MSSRes Net includes a multi-scale dilated convolution (MSDC) module and a multi-kernel max pooling (MKMP) module. The MSDC module can effectively encode high-level semantic feature maps and gather varying degrees of context information by reducing the strength of the receptive field without changing the size of the feature map data. To enhance the model effectiveness and accuracy Data pre-processing and augmentation play a vital role to model the data for training.

2.1 Image splitting

In this research consider n number of raster images split in ton number of blocks to store the data in order to avoid large computational and memory requirement problems can be solved and parallel computation should be increasing. The data size can be shrunk to 512×512 pixels in RGB channels capacity is the other level channel for training is 256×256 pixels. If we take more input data, we can reduce over fitting problems during training stage. To train data more samples introduced overlapping split of image blocks and randomly select patches data for testing in the proposed model as illustrated in the Figure 1.

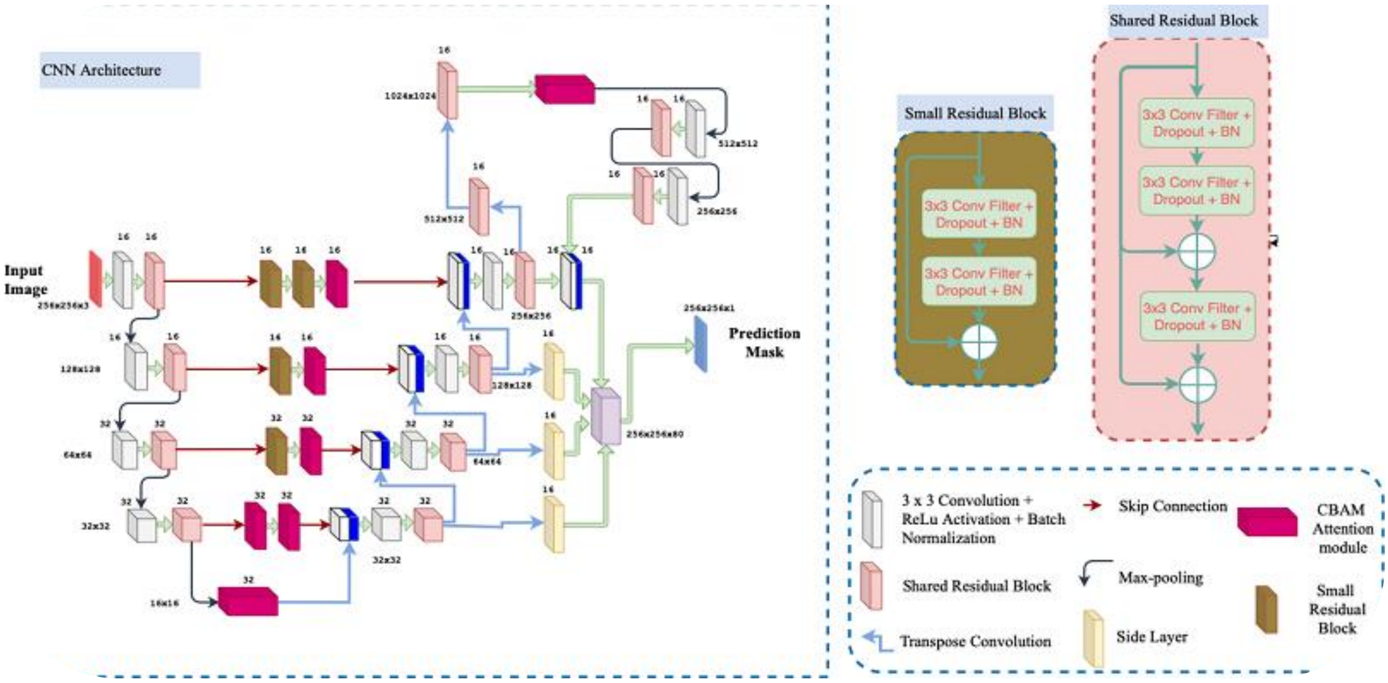


Figure 1. The structure of the proposed multi structure shared residual network (MSSResNet) with CBAM and SRB Modules

3. BUILDING MODEL ARCHITECTURE WITH PROPOSED ALGORITHM

In the Figure 1 accept Raster input image includes RGB channels from the n input images. All the images can be processed by respective convolution kernels in the multi-channel fusion network. Auto encoder and decoder used for pixel labelling for the feature maps of same size and grouping them as concatenated inputs to train the data. The proposed novel network includes the encoder network is performing down sampling operation due to that some information loss. The fusion channel decoder is an upgraded DeepLabV3+ [8] network that supports the CBAM and Resnet modules' fine-grained feature maps. The fusion channel decoder is an upgraded DeepLabV3+ [8] network that implements the fine-grained feature maps created by the CBAM and Resnet modules, and the ResNet-34 [6] model was based on ImageNet [26]. For the performance point of view each kernel height, width and number of convolutional layers is combined to generate output of feature maps.

3.1 Shared multi scale residual network

The key to enhancing the efficiency of water-body detections is to use multiscale structural information and merge it effectively, due to the varied sizes and shapes of water bodies in optical RS pictures. We suggest Multi Scale shared Residual Network (MSSResNet) to solve this problem based on this concept, as illustrated in Figure 1. This network is based on LinkNet [27], and it has a similar encoder-decoder structure. However, some spatial information that is important for images can be lost during the encoder's numerous down sampling operations U-Net [9] improves feature maps by using skip connections and combining low-level detail information from the encoder with high-level semantic information from the decoder. The proposed MSSResNet

Figure 2 not only retains more high-resolution detailed information in high-level feature maps, but also solves the problems of exploding gradient and vanishing gradient during the relatively deep network training process by combining low-level features with high-level features. U-Net [9] uses skip connections to improve feature maps by combining low-level detail information from the encoder with high-level semantic information from the decoder. By combining low-level data with high-level characteristics, the proposed MSSResNet Figure 2 addresses the issues of inflating gradient and fading gradient during the relatively deep network training process. The "Convolutional Block Attention Module" was included in addition to the Multi Scale Shared Residual Network (CBAM).

The exceptional accuracy of CNNs makes them ideal for image categorization and recognition. As a result of Yann LeCun's fascination with human visual perception and the ability to recognise things, this idea was floated in the late 1990s. ResNet's generalisation capacity is greatly enhanced by its identity mapping. ResNet has a variety of architectures. ResNet relies on what are known as "shortcut connections" or "identity connections" to hop between nodes. By traversing one or more layers, these links provide for easier access to each other. To overcome the problem of vanishing gradients in deep networks, these shortcut links were introduced.

In a residual block, layers are stacked so that the result of one layer is added to a higher layer in the stack. Following that, the non-linearity is put into action by combining it with the output from the matching layer inside the main path and applying it. With residual blocks, every layer feeds into next layer and immediately into the layers 2–3 hops away in a system with residual blocks the end. We will, however, concentrate on gaining an intuitive knowledge of why it was truly necessary, why it is so significant, and how comparable it appears to other cutting-edge systems.

The Convolution neural networks (CNNs) [10] is the most popular network for the computer vision based on powerful data processing parameter [28-31]. Every network configuration concentrate on three important parameters: 1. Depth 2. Width 3. Cardinality CNN use attention modules to assist it in learning and focusing on the most important data rather than unnecessary background information can be processed through "Convolutional Block Attention Module" in this paper. The channel and spatial axes are key dimensions to extract relevant features because convolution methods extract useful features by combining cross-channel and spatial information. CBAM is integrated into any CNN architectures for the feature refinement process. Residual Network [7] incorporate skip connections to optimize the issues of deep network. WideResNet [32], Inception-ResNet [33], and ResNeXt [34] all are designed with prerequisite knowledge of residual network. In this architecture proposed Residual network with more convolutional filters are incorporated along with CBAM network to reduce the depth of network. A simple 2D-convolutional layer, MLP, and sigmoid activation function at the conclusion of the attention module construct a mask of the input feature map. An encoder and decoder style attention module is used in the Residual Attention Network. Spatial resolution is the important parameter to extract water body that is lower the resolution poor area detection and higher the resolution very clear water areas can be detected.

3.1.1 Working principle of CBAM network

Based on an efficient design, we use spatial and channel-wise attention in CBAM in Figure 2, and empirically verify that employing both is preferable to using simple channel-wise attention [35]. The attention module was applied to channel and spatial dimensions in addition to the convolutional block attention module (CBAM). We'll look at how it works from the perspective of the channel and spatial attention modules. The previous 3D experience must be completed before accessing the CBAM. CNN's translation of the input data X was used to generate the feature map. $F_m \in R^{C\times H\times W}$ [2] (number of channels indicated by C, height defined by H, and width represented by W). Fm processes the channel capacity of the attention module and the spatial attention module as follows:

$$\begin{gathered} Fm' = Mc(Fm) \otimes Fm \\ Fm'' = Ms(Fm') \otimes Fm' \end{gathered} \tag{1}$$

The result of multiplying the channel attention map by the result of the feature map is denoted by Fm'. Fm'' denotes element-wise multiplication of the spatial attention map multiplied by F', or the final output denoted by ⊗. The attention sector helps in extracting the channels that hold essential information, as each channel in the feature map represents one individual detector. The channel attention module employed both a global average pooling layer and a global max pooling layer to complete the feature extraction process and reduce data loss.

The structure of CBAM is shown in Figure 1 combines both Channel attention ad spatial attention modules.

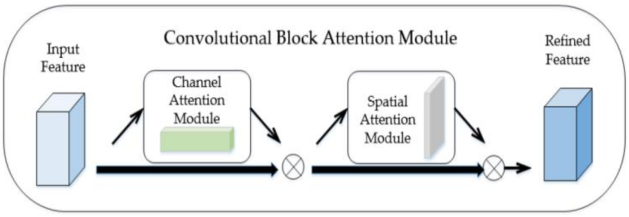


Figure 2. Basic block diagram for CBAM [32]

The active function for computing Mc's channel attention map is the sigmoid function $(Fm) \in R^{C\times 1\times 1}$, the following was the procedure:

$$\begin{gathered} Mc(Fm)=\sigma(MLP(Avgpool(Fm)) + \\ MLP(Maxpool(Fm))) \\ =\sigma(W1(W0(Fcavg))+W1(W0(Fc\ max))) \end{gathered} \tag{2}$$

A convolution technique known as "Max Pooling" is one in which the Kernel gets the most out of the area it convolves by extracting its maximum value. Essentially, Max Pooling tells the Convolutional Neural Network that it should only carry forward information with the greatest amplitude. The input layer is where the data that will be processed is sent. Using average pooling, each patch of a feature map is averaged. Means the feature map is down-sampled to the average value in each 2x2 area. The output layer is responsible for tasks like prediction and categorization. As many hidden layers as necessary can be used to perform computations in between input and output layers.

The spatial attention module, in contrast to the channel attention module, placed a larger emphasis on the regions of the feature map that had a better response. In order to highlight the regions containing crucial information, the global average pooling layer and the global max pooling layer were compressed into two 2D feature maps along the channel dimension for the feature maps generated in the channel attention module.

$$M_s(F_m)=\sigma(F^{7\times7}([AvgPool(F_m);\ MaxPool(F_m)])) \\ =\sigma(F^{7\times7}([F_s\ avg;\ F_s\ max])) \qquad (3)$$

The feature maps F s avg and F s max, which are squished into the channel dimension, depict the sigmoid function.

3.1.2 The multi-kernel max pooling (MKMP) module

Spatial pyramid pooling (SPP) [36] was originally used in natural picture classifications and object detections. We proposed the MKMP module, which was inspired by the SPP module. The proposed MKMP module offers context information for four different receptive field sizes, including 128×128, 64×64, 32×32, and 16×16 in SPP [36], which utilises a fixed-length representation and fully linked layers. In multiple kernel CNN [37], which employs an invariable kernel size and stride. The multikernel technique assumes that differentiating multiscale water bodies (such as rivers, ponds, and seas) using a single pooling kernel is difficult without a large increase in the number of neurons. It also takes use of more context information by extending receptive fields at the same time. Its four-level output also includes various-sized feature maps. In addition, after each level of max pooling, we employ a 1×1 convolution to minimize the dimensional weight, reducing the feature map's size to 1/n of the original dimension, channels in the original feature map represented by n. Bilinear interpolation samples the low dimensional feature map to get the same size. Finally, we combine the original feature map with the four-level outputs to form a total feature map. The MKMP module fuses local and global features at the feature map level, strengthening the final feature map's expression ability and improving water-body detection accuracy.

Using ID data, CNN can determine whether a object is a body or not. There are a total of two dimensions to this set of information. One dimension is the number of time-steps, while the other is a three-dimensional array of acceleration readings. The graphs that follow show how the kernel responds to accelerometer data in real time.

A convolution technique known as "Max Pooling" is one in which the Kernel gets the most out of the area it convolves by extracting its maximum value. Essentially, Max Pooling tells the Convolutional Neural Network that it should only carry forward information with the greatest amplitude. With the stride of 2, we can achieve maximum pooling on a 4*4 channel using a 2*2 kernel. As we are using a 2*2 Kernel for convolution. The channel has four values of 8,3,4,7 if we look at the first 2*2 set that the kernel is focusing on. That set contains "8" and "Max-Pooling" picks the highest value. As though the machine were a distributed system, a multikernel operating system views a multi-core machine as an interconnected network of independent cores. In this system, inter-process communication is implemented as message passing rather than shared memory.

4. EXPERIMENTS AND RESULTS

The experiment part will primarily focus on how to define water bodies from high-resolution photos using MSSResNet, as well as how training and validation may be accomplished using the suggested network that is depicted in the Figure 3.

4.1 Training procedure

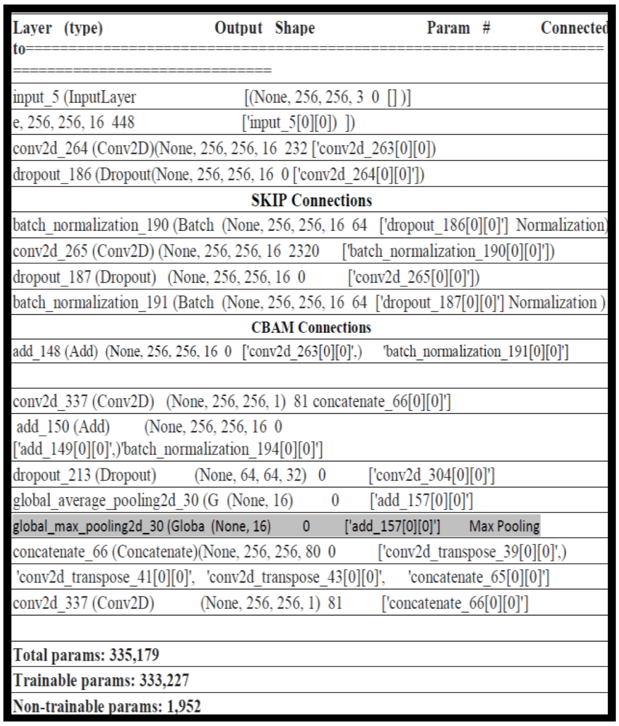
Finally, we train the network by determining which network parameters w minimize the cost function in Eq. (2). For each training sample x, the derivatives of the cost function J(w, y, t) with respect to various parameters w are recursively computed via backpropagation [38].

Proposed shared multi-scale residual network with CBAM and SRM modules processing results are shown in Figure 4. A trained network is evaluated using two measures. As stated in Eq. (4), the F1 score is the average of the precision and recall factors, with recall equaling the fraction of correctly expected positive (water area) pixels and precision equaling the proportion of correctly labeled pixels. Finally, the percentage of pixels that are correctly identified across the board is referred to as accuracy.

$$F1=2*[(precision*recall)/(precision+recall)] \qquad (4)$$

In the result section the comparable network depths considered, our proposed architecture regularly outperforms the U-Net reference. The data also illustrates that encoder depth has a significant impact on overall system performance, and that increasing encoder depth improves overall system performance [39].

Figure 4 demonstrates a precise approach for determining the location of water. A couple bodies of water weren't covered in the previous study since the photographs weren't processed correctly. This research made advantage of multi-scale shared standardized residual networks featuring good attenuation modules. It is possible to train a greater amount of data in a shorter period of time since more skip connections could be made in less time. All photos were processed in the proposed study using n raster images as input. Our suggested design consistently outperforms the U-Net benchmark at comparable network depths, as shown in Figure 5. Increasing encoder depth enhances overall system performance, according to the results of this study.



```
Layer   (type)                    Output   Shape                   Param   #        Connected
to=================================================================
=============================
input_5 (InputLayer                    [(None, 256, 256, 3  0  [] )]
e, 256, 256, 16  448                   ['input_5[0][0])  ])
conv2d_264 (Conv2D)(None, 256, 256, 16  232 ['conv2d_263[0][0])
dropout_186 (Dropout(None, 256, 256, 16  0 ['conv2d_264[0][0]'])
                              SKIP Connections
batch_normalization_190 (Batch  (None, 256, 256, 16  64   ['dropout_186[0][0]']  Normalization)
conv2d_265 (Conv2D) (None, 256, 256, 16  2320     ['batch_normalization_190[0][0]'])
dropout_187 (Dropout)   (None, 256, 256, 16  0          ['conv2d_265[0][0]'])
batch_normalization_191 (Batch  (None, 256, 256, 16  64  ['dropout_187[0][0]'] Normalization )
                              CBAM Connections
add_148 (Add)  (None, 256, 256, 16  0   ['conv2d_263[0][0]',)     'batch_normalization_191[0][0]']

conv2d_337 (Conv2D)   (None, 256, 256, 1)  81 concatenate_66[0][0]']
 add_150 (Add)        (None, 256, 256, 16  0
['add_149[0][0]',)'batch_normalization_194[0][0]']
dropout_213 (Dropout)        (None, 64, 64, 32)   0          ['conv2d_304[0][0]']
global_average_pooling2d_30 (G  (None, 16)         0       ['add_157[0][0]']
global_max_pooling2d_30 (Globa  (None, 16)        0        ['add_157[0][0]']       Max Pooling
concatenate_66 (Concatenate)(None, 256, 256, 80  0          ['conv2d_transpose_39[0][0]',)
 'conv2d_transpose_41[0][0]',   'conv2d_transpose_43[0][0]',     'concatenate_65[0][0]']
conv2d_337 (Conv2D)          (None, 256, 256, 1)  81         ['concatenate_66[0][0]']

Total params: 335,179
Trainable params: 333,227
Non-trainable params: 1,952
```

Figure 3. Shared multi scale residual network results

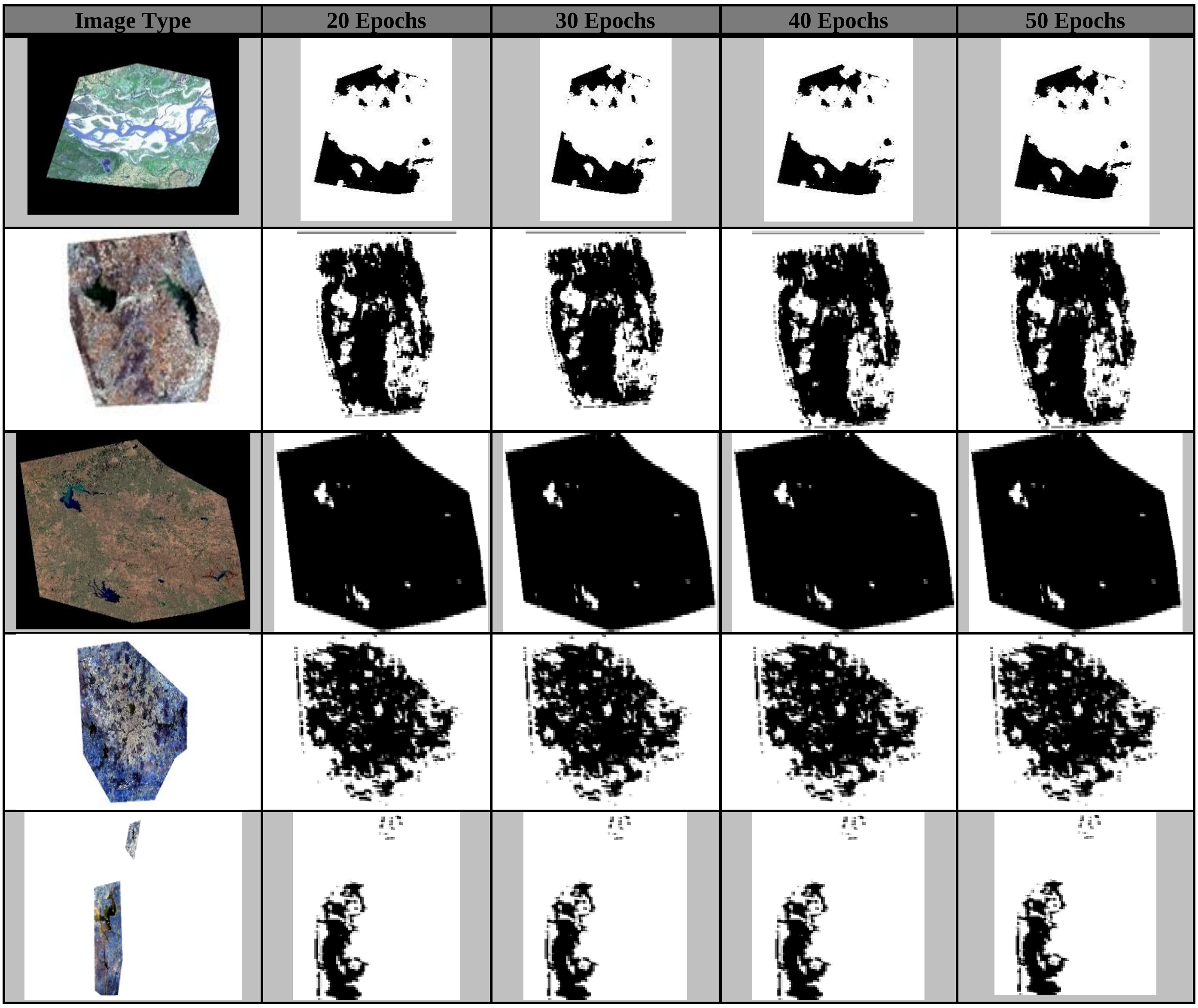
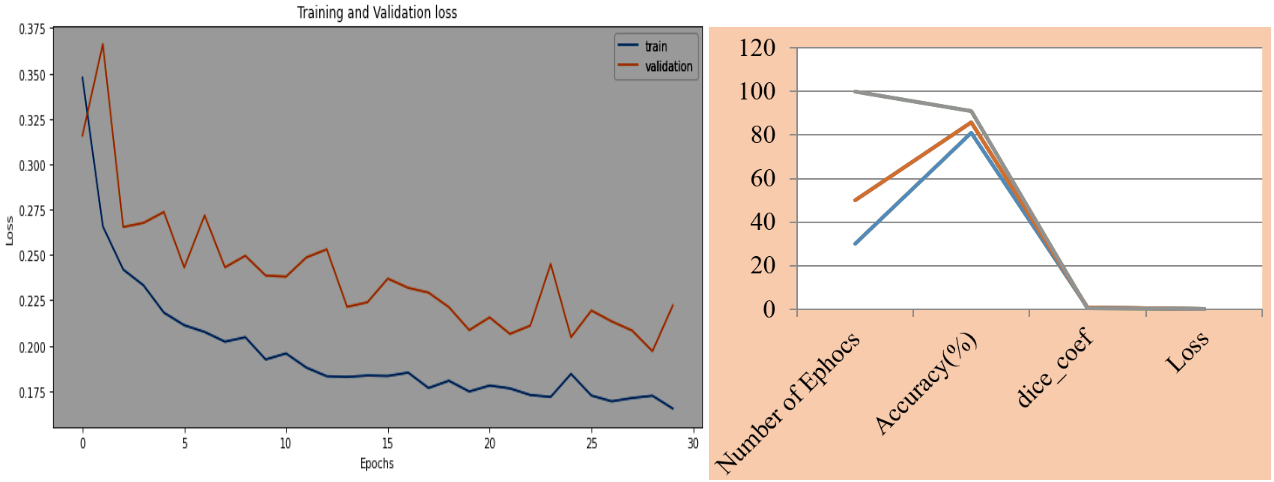


Figure 4. Water extraction images with various Epochs



(a) (b)

Figure 5. (a): Training and validation loss; (b): Parameters association

5. CONCLUSIONS

To verify our method, numerous raster images are used to extract water-body segmentation dataset (images with RGB bands) and the GID dataset (images with NIR-RGB bands) were used to analyse our method and compare it with existing approaches. The results are comparable and greatly improved with previous algorithms and also helpful for flood zone prediction with accurate results.

REFERENCES

[1] Yuan, K., Zhuang, X., Schaefer, G., Feng, J., Guan, L., Fang, H. (2021). Deep-learning-based multispectral satellite image segmentation for water body detection. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14: 7422-7434. https://doi.org/10.1109/JSTARS.2021.3098678

[2] Guo, J., Zhou, H., Zhu, C. (2013). Cascaded classification of high resolution remote sensing images using multiple contexts. Information Sciences, 221: 84-97. https://doi.org/10.1016/j.ins.2012.09.024

[3] Wang, M., Wan, Y., Ye, Z., Lai, X. (2017). Remote sensing image classification based on the optimal support vector machine and modified binary coded ant colony optimization algorithm. Information Sciences, 402: 50-68. https://doi.org/10.1016/j.ins.2017.03.027

[4] Li, L., Yan, Z., Shen, Q., Cheng, G., Gao, L., Zhang, B. (2019). Water body extraction from very high spatial resolution remote sensing data based on fully convolutional networks. Remote Sensing, 11(10): 1162. https://doi.org/10.3390/rs11101162

[5] Iandola, F., Moskewicz, M., Karayev, S., Girshick, R., Darrell, T., Keutzer, K. (2014). Densenet: Implementing efficient convnet descriptor pyramids. arXiv preprint arXiv:1404.1869. https://arxiv.org/abs/1404.1869

[6] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778.

[7] Chen, L.C., Papandreou, G., Schroff, F., Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587. https://arxiv.org/abs/1706.05587.

[8] Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), pp. 801-818.

[9] Pinheiro, P.O., Lin, T.Y., Collobert, R., Dollár, P. (2016). Learning to refine object segments. In European Conference on Computer Vision, pp. 75-91. https://doi.org/10.1007/978-3-319-46448-0_5

[10] Ronneberger, O., Fischer, P., Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[11] Lin, G., Milan, A., Shen, C., Reid, I. (2017). Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1925-1934.

[12] Chan, L., Hosseini, M.S., Plataniotis, K.N. (2021). A comprehensive analysis of weakly-supervised semantic segmentation in different image domains. International Journal of Computer Vision, 129(2): 361-384. https://doi.org/10.1007/s11263-020-01373-4

[13] Tao, A., Sapra, K., Catanzaro, B. (2020). Hierarchical multi-scale attention for semantic segmentation. arXiv preprint arXiv:2005.10821. https://arxiv.org/abs/2005.10821

[14] Peng, C., Li, Y., Jiao, L., Chen, Y., Shang, R. (2019). Densely based multi-scale and multi-modal fully convolutional networks for high-resolution remote-sensing image semantic segmentation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(8): 2612-2626. https://doi.org/10.1109/JSTARS.2019.2906387

[15] Yu, B., Yang, L., Chen, F. (2018). Semantic segmentation for high spatial resolution remote sensing images based on convolution neural network and pyramid pooling module. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(9): 3252-3261. https://doi.org/10.1109/JSTARS.2018.2860989

[16] Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J. (2017). Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881-2890.

[17] 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. https://doi.org/10.1016/j.isprsjprs.2018.04.014

[18] Chaurasia, A., Culurciello, E. (2017). Linknet: Exploiting encoder representations for efficient semantic segmentation. In 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, pp. 1-4. https://doi.org/10.1109/VCIP.2017.8305148

[19] Long, J., Shelhamer, E., Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431-3440.

[20] Chaurasia, A., Culurciello, E. (2017). Linknet: Exploiting encoder representations for efficient semantic segmentation. In 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, pp. 1-4. https://doi.org/10.1109/VCIP.2017.8305148

[21] Sun, K., Zhao, Y., Jiang, B., et al. (2019). High-resolution representations for labeling pixels and regions. arXiv preprint arXiv:1904.04514. https://arxiv.org/abs/1904.04514.

[22] Ronneberger, O., Fischer, P., Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[23] Feng, W., Sui, H., Huang, W., Xu, C., An, K. (2018). Water body extraction from very high-resolution remote sensing imagery using deep U-Net and a superpixel-based conditional random field model. IEEE Geoscience and Remote Sensing Letters, 16(4): 618-622. https://doi.org/10.1109/LGRS.2018.2879492

[24] Guo, H., He, G., Jiang, W., Yin, R., Yan, L., Leng, W. (2020). A multi-scale water extraction convolutional neural network (MWEN) method for GaoFen-1 remote sensing images. ISPRS International Journal of Geo-Information, 9(4): 189. https://doi.org/10.3390/ijgi9040189

[25] Duan, L., Hu, X. (2019). Multiscale refinement network for water-body segmentation in high-resolution satellite

imagery. IEEE Geoscience and Remote Sensing Letters, 17(4): 686-690. https://doi.org/10.1109/LGRS.2019.2926412

[26] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Li, F.F. (2009). Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, pp. 248-255. https://doi.org/10.1109/CVPR.2009.5206848

[27] Chaurasia, A., Culurciello, E. (2017). Linknet: Exploiting encoder representations for efficient semantic segmentation. In 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, pp. 1-4. https://doi.org/10.1109/VCIP.2017.8305148

[28] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778. https://doi.org/10.1109/CVPR.2016.90

[29] Angel, M.C.M., Humberto, D.B., Alfredo, T.M. (2021). Optimization of vehicle flow times in a single crossing system through the development of a multi-agent platform. Ingénierie des Systèmes d'Information, 26(4): 387-392. https://doi.org/10.18280/isi.260406

[30] Krizhevsky, A., Hinton, G. (2009). Learning multiple layers of features from tiny images.

[31] Lin, T.Y., Maire, M., Belongie, S., et al. (2014). Microsoft coco: Common objects in context. In European Conference on Computer Vision, pp. 740-755. https://doi.org/10.1007/978-3-319-10602-1_48

[32] Zagoruyko, S., Komodakis, N. (2016). Wide residual networks. arXivpreprint arXiv:1605.07146.

[33] Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K. (2017). Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1492-1500. https://doi.org/10.1109/CVPR.2017.634

[34] Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In Thirty-first AAAI Conference on Artificial Intelligence, 31(1). Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/11231.

[35] Hu, J., Shen, L., Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132-7141. https://doi.org/10.1109/CVPR.2018.00745

[36] He, K., Zhang, X., Ren, S., Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(9): 1904-1916. https://doi.org/10.1109/TPAMI.2015.2389824

[37] Shi, H., Wang, H., Jin, Y., Zhao, L., Liu, C. (2019). Automated heartbeat classification based on convolutional neural network with multiple kernel sizes. In 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService), Newark, CA, USA, pp. 311-315. https://doi.org/10.1109/BigDataService.2019.00055

[38] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11): 2278-2324. https://doi.org/10.1109/5.726791

[39] Li, M., Wu, P., Wang, B., Park, H., Yang, H., Wu, Y. (2021). A deep learning method of water body extraction from high resolution remote sensing images with multisensors. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14: 3120-3132. https://doi.org/10.1109/JSTARS.2021.3060769