International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

Segmenting Roads from Aerial Images: A Deep Learning Approach Using Multi-Scale Analysis

Nadeem Akhtar[1], Manish Mandloi[2]
[1]Department of EXTC
SVKM's NMIMS MPSTME
Shirpur, Maharashtra, India
akhtar.nadeem002@nmims.edu.in
[2]Department of Information and Communication Technology
Pandit Deendayal Energy University
Gandhinagar, Gujarat, India
manish.mandloi@sot.pdpu.ac.in

Abstract—Road map generation requires frequent map updates due to the irregular infrastructural changes. Updating a manual road map is a lengthy process, whereas using aerial or remote sensing (RS) requires less time for the update. However, road extraction becomes more complex due to the similar texture appearance of building top roofs, shadows, and occlusion due to trees. The occluded roads appear as discontinuous road patch in segmented image of updated maps. In this paper, we propose a deep learning method that uses multi-scale analysis for road feature extraction. The dilated inception module (DI) in the up and down sampling paths of network extracts the local and global texture patterns of the road. Furthermore, we also utilize the pyramid pooling module (PP) which has average and max pooling to study the global contextual information under the shadow regions. In the proposed architecture, first, the road in the aerial images is segmented along with the tiny non-road segments. Next, the post processing, which exploits the geometrical shape features, is utilized for filtering the tiny non-road noises. The performance of proposed network is validated on using the publicly available Massachusetts road data by comparing with the other models available in literature.

Keywords- Road Segmentation, Pattern Recognition, Deep Learning, Remote Sensing, Neural Networks

I. INTRODUCTION

High-resolution Remote Sensing (RS) images serve as a primary resource for numerous geographic information applications, such as urban planning [1], economic assessment [2], land resource management [3], precision agriculture [4], and environmental protection [5]. With the recent technological advancements in the field of image processing and computer vision, RS images are being used in a wide variety of applications. This includes geographic information system, change detection, ecological research, land management, disaster monitoring and detection of natural and non-natural objects [6].

Roads play a crucial role in our modern lives. A high-definition road map is need of the hour for traffic monitoring, unmanned vehicles, urban planning and smart cities. The frequent changes in the civil infrastructure urges the regular map update which is crucial for various application, such as, route analysis and emergency response. Road map update can be obtained by performing manual labelling, driving car tracing and by using RS images [7]. Amongst these, manual labelling is costlier and time-consuming process and suffers from interpreters' discrepancy, while tracing the driving car may lead to the incomplete information about the road map [7]. On the other hand, road detection using RS images is more popular and economic as it can be updated within shorter interval of time. However, the presence of non-road objects, background complexity and occlusions make the road detection process a challenging task [8]. This motivates for further research on road detection problem using RS images

Recently, there has been an increased attention towards the use of Deep Learning (DL) methods for RS image processing [9]. The DL methods are capable of providing promising performance for road detection in terms of map update accuracy [10]. In the literature, researchers have used the convolutional neural network for road segmentation [11]. It includes different semantic segmentation architectures like U-net [12], Fully Convolutional Network (FCN) [13], SegNet [14] and DeepLab [15]. These architectures are an encoder and decoder-based architecture. The encoder uses the convolution and down sampling while decoder uses the convolution and up sampling. However, the differences in these methodologies lies in performing up, down sampling and fusion of features. Several methods based on these architectures are proposed in the literature for obtaining the enhanced performance in road detection. Basically, these methods have improved the performance on road detection by modifying the DL architectures [11], [16], [17], introducing modified layer connections [18], and proposing the modified loss function [19]. Zhou et al. [20] proposed the DlinkNet [20] by modifying the LinkNet [21]. The authors [20] adds a layer of dilated convolution with dilation rates in the bridge section of U-net. The varying dilation rates of dilated convolution enables the network to learn the multiscale road features. Gao et al. [22] proposed the multi feature pyramid network to enlarges the receptive field of feature points and extract the road features at different scale. Mendes et al. [23] integrated the network-in-network architecture and FCNs. This takes advantage of a large contextual window for fast road detection. Li et al. [24] proposed the hybrid convolutional network that integrates multiple subnetworks to extract roads at multiscale. The modified architecture for road segmentation in the literature are cascaded end-to-end encoder and decoder-based architecture [16], Joint-net [17] and Y-net [25]. Joint-net [17] use dense atrous

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

convolution block which maintains the feature propagation and achieves large receptive field. The Y-net [25] model includes both feature extraction and fusion modules, resulting in a mean region intersection over union of 67.75 percent. However, it fails in extracting narrow roads and requires more training time, furthers suffers with class imbalance for narrow road. The FCN [26] combines shallow fine-grained pooling layer features with those of the deep final score layer. On the other hand, the residual learning with dense connection is introduced for U-net in [27]. The residual learning in U-net [18] achieves betters results, but fails in detecting the shadowed roads of trees and parking lots. A spatial pyramid pooling is integrated with encoder and decoder architecture in [19] for enhancing the road features. The network proposed by [28] uses guided filters and residual units for road detection. However, the model fails in detecting the roads having similar spatial distribution and spectral values of non-road objects. All the aforementioned approaches segment the road in RS imagery, but fail in detecting the occluded road due to tree and building shadow [29], further the DL models faces the various issues like memory consumption, processing time and improving the results using light weight model. Despite of intensive efforts for road detection using RS images, the problem remains to be challenging in terms of required processing time, memory consumption and the prediction quality metric.

In this paper, we propose to detect the road in aerial imagery using multiscale feature analysis with Dilated Inception module (DI) and Pyramid Pooling module (PP). The DI module uses dilated convolution to capture the object shape using different dilation rates. The proposed PP module aggregates more global contextual information with average and max pooling with varying sub-region to capture the shaded road features. The Basic Module (BM) along with DI and modified PP module segments the road and the variant roads in occluded regions of complex RS imagery. The isolated non-road regions in the binary image are cleaned by post processing technique which exploits the geometrical shape features of the segmented image. The proposed methods give more promising results in terms of road detection performance parameters in comparison with other methods present in the literature.

II. MATERIALS AND METHOD

The process flow of road segmentation from RS images are shown in Figure 1. The DL model is trained and tested on RS dataset using focal loss. The proposed model extracts the road feature that uses the multiscale analysis using zooming and shrinking the input feature map for segmentation. The binary images are post processed with geometrical shape analysis to clean the noises present due to similar spectral characteristics of roads.

A. Road Segmentation Model

The proposed architecture does not have any encoder and decoder, instead it has zooming and shrinking paths for road feature study. The proposed methodology uses the multi-scale analysis by zooming and shrinking the image by the factor of 2. The proposed architecture extracts the road feature by zooming and shrinking the input feature. The zooming process doubles while shrinking process halves the image size, moreover zooming and shrinking halves and doubles the number of feature maps. The number of features used during convolution, zooming and shrinking process are 64, 32 and 128. The number of feature variation is less as compared to other encoder decoder-based architectures such U-net, further it saves the lot of memory consumption and processing time. The Basic Model (BM) considered in this paper for road segmentation is inspired from Atli [30]. The architecture does not have encoder and decoder, instead it has zooming and shrinking paths. The detailed architecture of proposed model is shown in Table 1. Basically, it has two section, one section up samples the feature and other section down samples the features. The multiscale analysis is carried out by using the zooming and shrinking paths and its architecture for road segmentation is shown in Figure 2. The three colored block yellow, pink and orange represents the zooming, bridge and shrinking paths. The zooming of feature map is carried out using up sampling process using transposed convolution, while shrinking is achieved by using down sampling with convolution with stride of 2. The convolution with stride in down sampling process do not loses any information as compare to the pooling process. The convolution layer in the architecture is followed by batch normalization and ReLU layer. This shrinking and zooming with doubling and halving the feature size extracts road features for segmentation with short span of training.

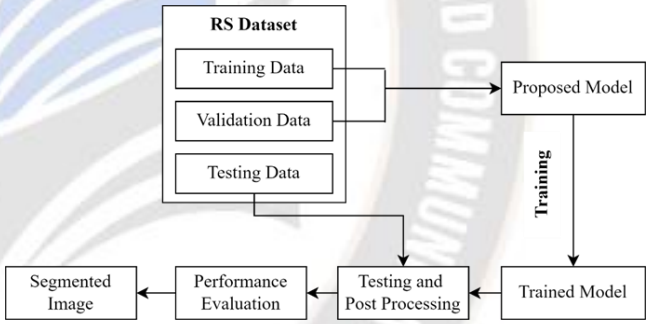


Figure 1: Road extraction process flow

The transposed convolution and convolution with stride are used for up and down-sampling. The up-sampling after convolution layer is followed by down-sampling in zooming path while converse for shrinking path. The model extracts thick and thin feature at different resolution scale, further these zooming and shrinking paths have residual skip connections.

B. Dilated Inception and Pyramid Pooling Module

The prediction accuracy in image processing is heavily influenced by the field of view, which refers to the amount of contextual information available for analysis. Insufficient field of view can result in suboptimal outcomes, especially when dealing with partial information about structures such as buildings or roads. This deficiency of data makes it hard to discriminate between objects with like textural features. However, dilated convolutions can be used to extract multi-scale contextual information while maintaining the same computational cost and increasing the field of view.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

DI Module shown comprises three parallel convolution paths with different dilation rates, allowing the units to achieve the larger receptive field without affecting the parameter size. The dilated convolution with different dilation rate is useful in shape analysis of roads under the occlusion. The outputs these three dilated convolutions with dilation rate of 4, 8, and 16 are fused using elementwise addition. DI module with varying dilation rate are captures the local and global texture patterns of roads present in the RS images. This aggregation of contextual information improves the overall feature discrimination capabilities of the network.

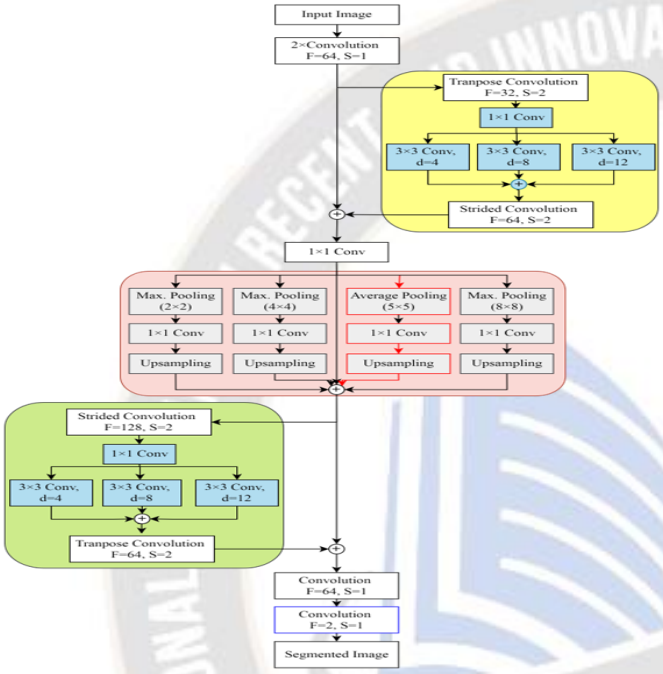


Figure 2: BM architecture PP and DI module for road segmentation. The symbol "+" represents elementwise addition of two branches. Strided convolution is applied for down sampling the input features.

The bridge component of Figure 2, is shown with pink color, is the PP module. The PP module generates a multi-scale representation of features by applying varying sub-regions of max and average pooling. The sub-regions of max pooling are 2×2, 4×4, and 8×8 while 5×5 for average pooling is employed. The max and average pooling down samples the input feature by picking the max and average value of feature in defined pooling window. Each pooling operation is followed by convolution and bilinear interpolation up sampling operation. The average pooling is effective for extracting the shaded road feature. These multiscale pooled feature with last extracted features is processed in shrinking paths of the model. The PP module enhances the performance, since it to capture both local and global features at deeper layers, that is effective in identifying the roads under the shaded regions.

TABLE I: PROPOSED ARCHITECTURE FOR ROAD SEGMENTATION. WHERE C IS NUMBER OF CLASSES, IN OUR CASE IT IS TWO

Module	Network Layer	Filter Size	Stride, dilation rate	Output Size
Convolution	Convolution	3×3×64	S=1, d=1	512×512×64
	Transpose Convolution	3×3×32	S=2, d=1	1024×1024×32
	Convolution	1×1×32	S=1, d=1	1024×1024×32
Dilated Inception Module	Convolution	3×3×32	S=1, d=4	1024×1024×32
	Convolution	3×3×32	S=1, d=8	1024×1024×32
	Convolution	3×3×32	S=1, d=12	1024×1024×32
Convolution	Convolution	3×3×64	S=2, d=1	512×512×64
	Convolution	3×3×64	S=1, d=1	512×512×64
Pyramid Pooling Module	Max Pool	2×2	S=2	255×255×64
	Max Pool	4×4	S=4	125×125×64
	Max Pool	8×8	S=8	57×57×64
	Average Pool	5×5	S=5	98×98×64
	Convolution	1×1×64	S=1, d=1	-
	Up sample Bilinear interpolation	-	-	512×512×64
Convolution	Convolution	3×3×128	S=2, d=1	256×256×128
	Convolution	1×1×128	S=1, d=1	256×256×128
Dilated Inception Module	Convolution	3×3×128	S=1, d=4	256×256×128
	Convolution	3×3×128	S=1, d=8	256×256×128
	Convolution	3×3×128	S=1, d=12	256×256×128
Convolution	Transpose Convolution	3×3×64	S=2, d=1	512×512×64
	Convolution	3×3×64	S=1, d=1	512×512×64
	Convolution	3×3×C	S=1, d=1	512×512×C
	Softmax	-	-	512×512×C

C. Model Training

The learning in DL models is an iterative process which minimizes the loss between the estimated and the actual value. Convergence of the model depends on the minimum value of loss achieved by the model. Initially model starts learning the road and background regions with few non-road regions detected as road, but as the learning process deepens, model able to distinguish between road and non-road having similar spectral signature of road perfectly. This process depends on many factors that includes model depth, layer connection and loss function used. The training process is of model depends on minimizing the loss function and updating the weight matrix. Since road appears as continuous structure, non-road regions having similar spectral characteristics appears isolated in binary image, hence we trained the model with 50 epochs with focal loss function. This partial training along with post processing gives most promising results as compare to other methods in this domain. The model is trained with focal loss since road

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

dataset has about 10 percent of road pixels and remaining is background. This creates a class imbalance in the dataset that makes the network to study more about the dominant classes. Therefore, it is very important to maintain the balance between the classes to avoid the degradation in the model performance. The loss functions play very crucial role in convergence of the network. It tries to minimize or optimize the gap between the predicted and actual value using DL algorithms. In every iteration it compares predicted value with the actual value and tries to optimize the gap between predicted and actual value by updating the layer weights. The focal loss handles the class imbalance by tuning the loss function parameter. The focal loss enables the model to learn hard examples and down weights the easy examples. The focal loss introduces an extra term in cross entropy loss expression to condense the impact of correct estimates and focus on incorrect examples. The focal loss is given by (1)

$$\text{Focal Loss} = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \quad (1)$$

In this $\gamma>0$, when $\gamma=1$ it, focal loss acts like cross entropy loss. It indicates effectiveness to focus on incorrect examples. The parameter α ranges between 0 to 1. This article uses focal loss with $\alpha=0.25$, and $\gamma=2$ for proposed compared methods.

D. Post Processing

The segmented image has road as well as falsely detected road objects. This falsely detected road appears as noise and are having similar texture of road. These isolated non-roads are cleaned by using geometrical shape analysis, since the road appears as elongated object with larger lengths and shorter widths. The objects in binary image are analyzed based on three geometrical parameters as: area (S), Complex Rate (CR) and Fullness Ratio (FR) [31]. The segmented objects FR identifies the elongated objects in the segmented image; hence FR helps in identifying the isolated non-road objects. The complexity of shape defined by (2) using perimeter (P) and S while FR is by (3).

$$CR=P^2/S \quad (2)$$

$$FR=S/S_{MER} \quad (3)$$

The square and circle shape have CR of 12.6 and 16 [31] respectively. The SMER in (3) is the area of minimum bounding rectangle (see Figure 3 (b)). The threshold in (3) for area (S_T), CR (C_T) and FR (F_T) are selected after number of trials to be 150, 20 and 0.7 for optimal results

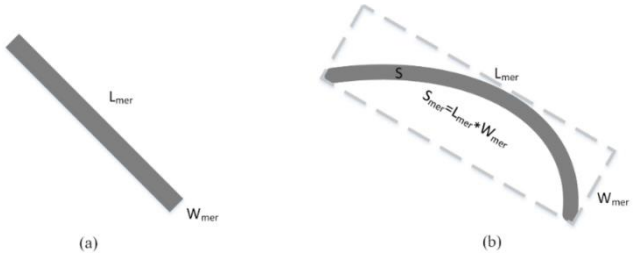


Figure 3: Bounding rectangle for (a) Straight; and (b) curved-road [31]

$$(S>S_T)\ \&\&\ (C>C_T)\ \&\&\ (F<F_T) \quad (4)$$

The tiny and isolated object in the binary image that do not satisfy (4) are discarded as non-road patches. The post processing improves the overlap area between predicted and labelled image by cleaning the isolated non-road patches

E. Dataset and Implementation Details

For experiment on road segmentation, we have used Massachusetts road dataset [32]. The Massachusetts dataset has total 1171 aerial images of the Massachusetts state (U. S.). The dataset has image size of 1500×1500 pixels which contains different roads structures of urban, suburban, and rural region. The dataset has only two classes as road and background. We have extracted the patches of 512×512 and randomly chosen 335 patches as dataset. The road appears narrow in Massachusetts dataset. The model is trained with ratio of 90 and 10 for training and testing dataset. The validation data is about 25 percent out of 90 percent training dataset. All models in this experiment are trained with an Adam optimizer and batch size of 6. The progressive learning rate is used for training the model. The initial learning rate for training is 0.001 which reduces to 0.0001 after 22 epochs and 0.00001 after 44 epochs. The model is trained on core i7, 32GB RAM and RTX3070 GPU with dedicated 8 GB internal memory.

F. Evaluation Metrics

For road segmentation correctness, completeness, and quality metrics are used for performance evaluation denoted by (5), (6) and (7) respectively. Correctness of road represents the correctly classified road pixels out of total labelled road pixels. Completeness is the correctly identified road pixels from total pixels. Quality metric of road detection represents the intersection over union of road segmentation. Dice is the harmonic mean of completeness and correctness, is given by (8). The high value of dice parameter indicates the high performance [33].

$$\text{Correctness} = (TP / (TP + FP)) \times 100 \quad (5)$$

$$\text{Completeness} = (TP / (TP + FN)) \times 100 \quad (6)$$

$$\text{Quality} = (TP / (TP + FP + FN)) \times 100 \quad (7)$$

$$\text{Dice} = (2 \times \text{Correctness} \times \text{Completeness}) / (\text{Correctness} + \text{Completeness})) \times 100 \quad (8)$$

III. RESULTS AND DISCUSSION

The experimental study of road segmentation is carried on Massachusetts road dataset with proposed architecture followed by post processing. The proposed method is compared and listed in Table 3 in same environment with post processing. It is also compared with other methods published statics on same dataset.

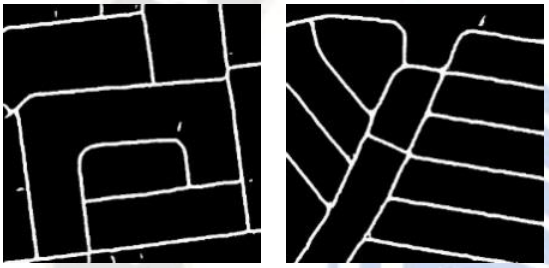
International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

A. Effect of Post Processing

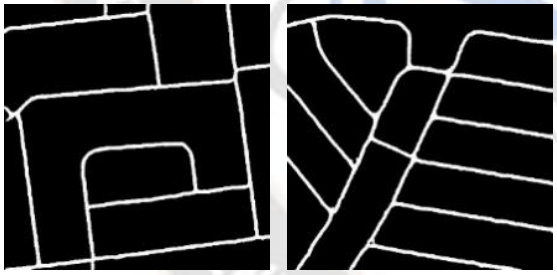
The post processing aims to filter isolated noise from road in binary image. This noise appears as an isolated tiny object from the road, which makes it easy to filter using geometrical properties. The post processed images are shown in Figure 4 (b). Post-processing was used for all models considered in the simulations. The post processed simulations results are shown in Figure 5. The quantitative effects of post-processing improve the road segmentation are presented in Table 2. The values presented here in Table 2 are the mean of 32 testing images.

TABLE II: IMPACT OF POST PROCESSING ON TEST IMAGES.

	Completeness	**Correctness**	**Quality**	**Dice**
Without Postprocessing	85.50	87.80	76.23	86.67
Post processed	**85.51**	**88.07**	**76.65**	**86.74**



(a) Road segmented images with isolated noises.



(b) Cleaning of noises using post processing

Figure 4: Effect of post processing on segmented image (a) Segmented and (b) Post processed image

B. Results on Massachusetts Road Dataset

The Massachusetts dataset was used to simulate the proposed model, and the results of the simulation are compared to other methods in Figure 5 and Table 3. These numeric values in Table 3 represent the mean of 32 testing images. The highest and second highest values in Table 3 are highlighted in bold and italics. SegNet has the highest completeness, while the proposed method has the highest correctness, dice, and quality.

The quality metric indicates the intersection between labelled and predicted image. Table 3 shows that SegNet and FCN have comparable quality metrics, but FCN has the highest memory consumption. ResU-net and U-net lag behind by 6% and 11%, respectively. DenseResSegnet (DR Segnet) has the second highest quality and dice metric.

TABLE III: ANALYSIS OF PROPOSED AND OTHER METHODS FOR ROAD DETECTION MASSACHUSETTS DATASET

Network	**Completeness**	**Correctness**	**Quality**	**Dice**
SegNet [14]	**87.15**	*84.65*	73.47	84.65
U-net [12]	75.33	82.93	65.09	78.65
ResU-net [18]	82.77	81.31	69.44	81.90
FCN [13]	85.46	84.11	73.60	73.60
DRSegnet [1]	82.94	83.17	*74.41*	*85.30*
DlinkNet [20]	68.48	77.11	55.85	71.31
Proposed	*85.51*	**88.07**	**76.65**	**86.74**

Figure 5 shows the qualitative analysis of proposed method. Figure 5(a) shows an image with a completely occluded road. U-Net and ResU-Net fail to bridge this gap, as shown in Figure 5(a) for U-Net and ResU-Net. Figure 5(a) shows that FCN fails to identify a small segment of the road, which is indicated by the red box. Figure 5(b) shows that a few parts of the road are shadowed by a tree. The proposed method, SegNet, and DR Segnet all maintain the continuity of the road network. In Figure 5(d), the partially shaded road is not segmented correctly by SegNet, ResU-net, and U-net, but the proposed method partially maintains the continuity of the road. In Figure 5(e), the track road is identified as a non-road region by the proposed method and DlinkNet [20], while SegNet, DR Segnet, U-net, ResU-net, and FCN identify it as a road. The road junction is the point where more than two roads are meeting, this different appearance than the roads due to road circle etc. The proposed method is able to identify the road junctions as well while other methods fail in detection the roads junction points. The proposed method is compared to other methods in Table 4, and it has a higher quality value. The methods in Table 4, such as attention-based, densely connected, and dense U-net architecture, are deeper networks that require more processing time and memory consumption. Our method has a smaller number of feature variations, which makes it faster and more memory-efficient.

C. Ablation Study on Road Detection

An ablation study in deep learning is a technique used to investigate the contribution of individual components of a deep learning model to its overall performance. The term ablation is derived from the medical procedure of removing a part of an organ or tissue to study its function.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

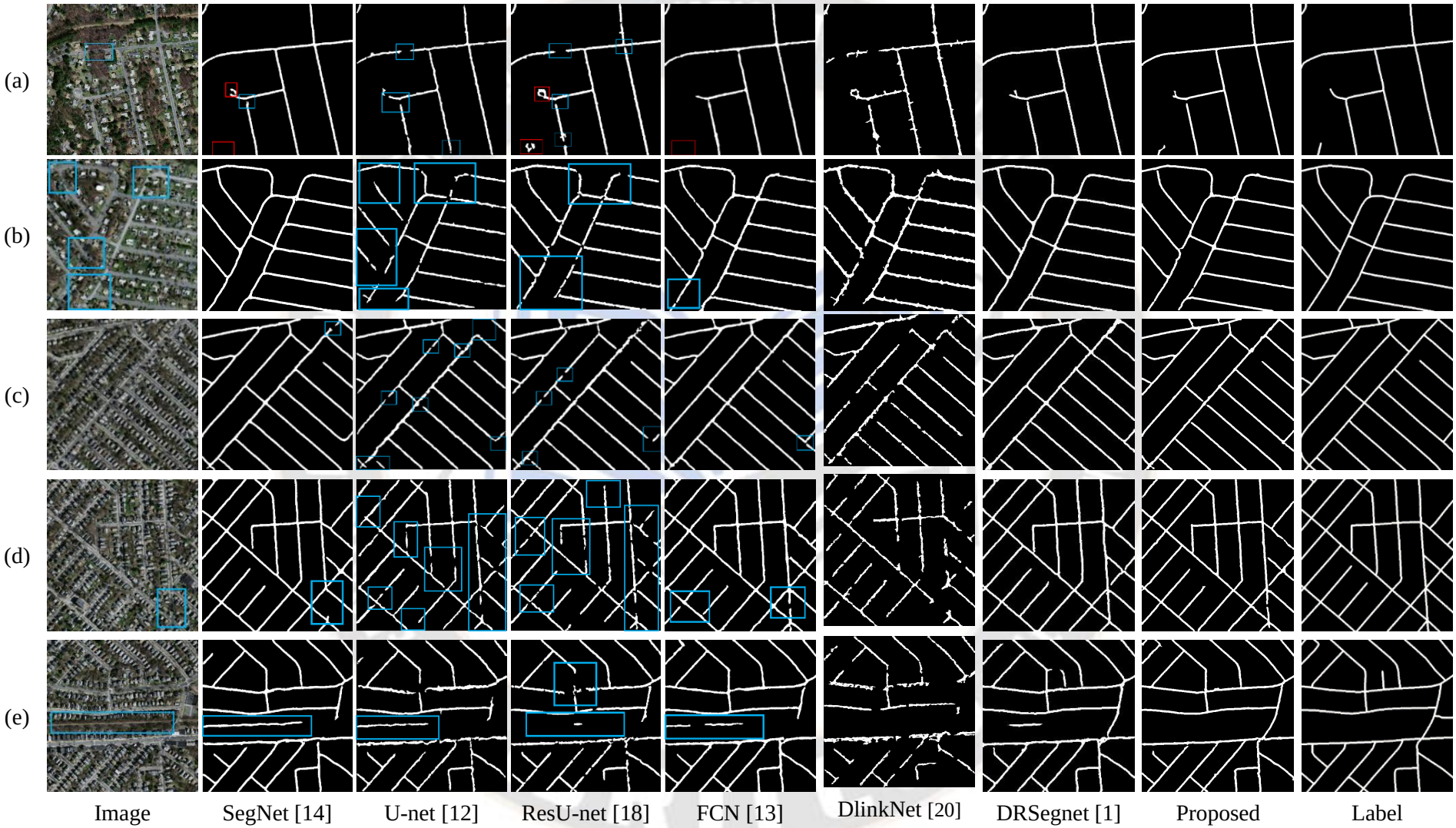


Figure 5: Qualitative analysis of proposed algorithm on Massachusetts dataset. The blue boxes represent the broken road segments. The first column represents the image identification viz. (a) to (e).

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

In the context of deep learning, an ablation study involves removing one or more components of a model and observing how this affects the model's performance. Ablation studies can be a valuable tool for understanding and improving deep learning models. They can help researchers to identify the key components of a model and to develop more efficient and effective models. In this ablation is carried with DI and PP module. **Error! Reference source not found.** shows the i nfluence of each component in proposed model on Massachusetts dataset.

TABLE IV: ABLATION STUDY OF THE PROPOSED METHOD WITH ADDED COMPONENTS IN NETWORK ON MASSACHUSETTS DATASET.

Network	Completeness	Correctness	Quality	Dice
BM	78.21	*88.25*	70.79	82.85
BM + PSP	81.42	**89.49**	*74.32*	*85.21*
BM + DI	82.62	87.96	74.17	85.11
Proposed	**85.41**	88.00	**76.52**	**86.66**

Table 4 indicates that the proposed model improves the completeness by 7%, the Dice coefficient by 4%, and the quality by 7% over the BM for the Massachusetts dataset. Figure 6 shows the improvement in dice and quality scores when the DI and PP modules are added to the BM. The quality of detection is poor BM and further decreases with the Pyramid Scene Parsing (PSP) Network [34]. The DI module mainly improves the quality and dice scores, while the proposed PP module further improves the results of BM and DI.

The proposed method performs multiscale analysis of road features with limited training epochs. It is a free-encoder and decoder model that up samples and down samples images by doubling the number of features for shrunken images and halving the number of features for zoomed images. The number of features in each layer is limited to 32, 64, and 128, which reduces memory requirements. This makes the proposed method less memory-intensive than other methods. BM has poor continuity detection capabilities (see Figure 6). The DI module improves the field of view without increasing the size of the weight matrix. It helps in recognizing global and local texture patterns of roads, which is beneficial for maintaining road continuity.

The PP module extracts global contextual information by down sampling the information through pooling operations. The combination of average and max pooling helps to distinguish road textures under shadow conditions. The PP module extracts road features at varying scales, which helps to identify the global texture of roads and maintain road continuity. Additionally, it helps to identify roads with building and parking lots that have the same texture as roads. The proposed network converges faster than other networks with a shorter number of epochs. Simulation results show that SegNet has better completeness, i.e., road continuity, but suffers from FP pixels around roads. U-Net and ResU-Net can improve their performance with increasing processing time and number of epochs. A bridging algorithm in post-processing that can bridge the gap between two broken road segments can further improve the performance.

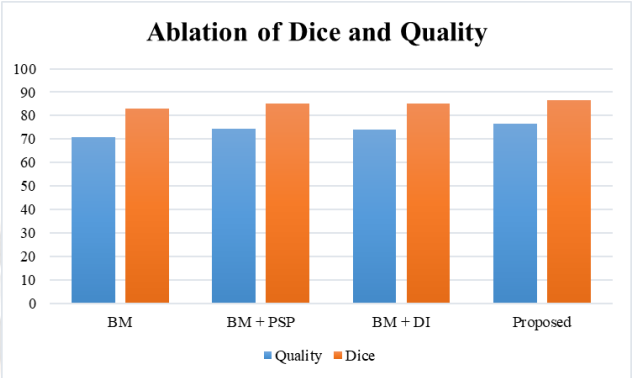


Figure 6: Ablation of Dice and Quality metric of road detection

IV. CONLUSION

This article proposes road detection that uses multiscale feature analysis for global and local texture learning for narrow and wide road dataset. A dilate inception module was used in zooming and shrinking paths of the network to increase the field of view. The increased field of view analysis the local and global texture of road under varying conditions. The deeper contextual information is analyzed by the proposed PP module for object shape analysis. The combination of average and max pooling maintains the continuity of road under the shaded regions also. The PP module extracts the road features with varying scale, able to identify the global texture of roads and helps in identifying roads with building and parking lots having same texture of the roads. The zooming and shrinking image capture the local and global textures of the road objects which is further enhanced by the DI and PP module in the proposed architecture. The post processing is employed to clean the isolated noises present in the binary image. The post processing cleans the noise using geometrical shape analysis. The quantitative results are obtained to show the generalization ability of proposed network in detecting both narrow as well as wide road networks in large-scale region within short span. A bridging algorithm in post processing that can bridge the gap between the two broken segments of the road can further improves the performance.

Conflict of Interest: The authors don't have any conflict of interest among them.

REFERENCES

[1]N. Akhtar and M. Mandloi, "DenseResSegnet: A Dense Residual Segnet for Road Detection Using Remote Sensing Images," 2023 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS), Hyderabad, India, 2023, pp. 1-4, doi: 10.1109/MIGARS57353.2023.10064603.

[2]M. Song, B. Li, P. Wei, Z. Shao, J. Wang, and J. Huang. "DMF-CL: Dense Multi-scale Feature Contrastive Learning for Semantic Segmentation of Remote-Sensing Images." In Chinese Conference on Pattern Recognition and Computer Vision (PRCV), pp. 152-164. Cham: Springer Nature Switzerland, 2022. doi: 10.1007/978-3-031-18916-6_13

[3]X. Tong, G. Song Xia, Q. Lu, H. Shen, S. Li, S. You, and L. Zhang. "Land-cover classification with high-resolution remote sensing images using transferable deep models." Remote

Sensing of Environment, vol. 237, pp. 111322, 2020. https://doi.org/10.1016/j.rse.2019.111322.

[4]M. Weiss, F. Jacob, and G. Duveiller. "Remote sensing for agricultural applications: A meta-review." Remote sensing of environment vol. 236, pp. 111402, 2020. doi: 10.1016/j.rse.2019.111402.

[5]S. Subudhi, R. N. Patro, P. K. Biswal and F. Dell'Acqua, "A Survey on Superpixel Segmentation as a Preprocessing Step in Hyperspectral Image Analysis," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 5015-5035, 2021, doi: 10.1109/JSTARS.2021.3076005.

[6]R. Lian, W. Wang, N. Mustafa and L. Huang, "Road Extraction Methods in High-Resolution Remote Sensing Images: A Comprehensive Review," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 5489-5507, 2020, https://doi.org/10.1109/JSTARS.2020.3023549.

[7]Y. Zhang, G. Xia, J. Wang and D. Lha, "A Multiple Feature Fully Convolutional Network for Road Extraction from High-Resolution Remote Sensing Image Over Mountainous Areas," in IEEE Geoscience and Remote Sensing Letters, vol. 16, no. 10, pp. 1600-1604, Oct. 2019, https://doi.org/10.1109/LGRS.2019.2905350.

[8]J. Wang, J. Song, M. Chen, et al., "Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine," International Journal of Remote Sensing, vol. 36, no. 12, pp. 3144–3169, 2015.

[9]N. Akhtar, N. S. Choubey, and U. Ragavendran, "Investigation of non-natural information from remote sensing images: A case study approach," Computational Intelligence and Sustainable Systems: Intelligence and Sustainable Computing, pp. 165–199, 2019.

[10] W. Zhang and B. Hu, "Forest roads extraction through a convolution neural network aided method," International Journal of Remote Sensing, vol. 42, no. 7, pp. 2706–2721, 2021.

[11] Y. Wei, Z. Wang and M. Xu, "Road Structure Refined CNN for Road Extraction in Aerial Image," in IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 5, pp. 709-713, May 2017, https://doi.org/10.1109/LGRS.2017.2672734.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, 234–241, Springer (2015).

[13] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 4, pp. 640–651, 2016.

[14] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder decoder architecture for image segmentation," IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 12, pp. 2481–2495, 2017.

[15] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," in Proceedings of the European Conference on Computer Vision (ECCV), V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., Munich, Germany, Sep. 8-14, 2018, pp. 801-818. https://doi.org/10.1007/978-3-030-01231-1_49.

[16] G. Cheng, Y. Wang, S. Xu, H. Wang, S. Xiang and C. Pan, "Automatic Road Detection and Centerline Extraction via Cascaded End-to-End Convolutional Neural Network," in IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 6, pp. 3322-3337, June 2017, https://doi.org/10.1109/TGRS.2017.2669341.

[17] Z. Zhang and Y. Wang, "Jointnet: A common neural network for road and building extraction," Remote Sensing, vol. 11, no. 6, p. 696, Mar. 2019. https://doi.org/10.3390/rs11060696.

[18] Z. Zhang, Q. Liu and Y. Wang, "Road Extraction by Deep Residual U-Net," in IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 5, pp. 749-753, May 2018, https://doi.org/10.1109/LGRS.2018.2802944.

[19] H. He, D. Yang, S. Wang, et al., "Road extraction by using atrous spatial pyramid pooling integrated encoder-decoder network and structural similarity loss," Remote Sensing, vol. 11, no. 9, pp. 1015, Apr. 2019. https://doi.org/10.3390/rs11091015.

[20] L. Zhou, C. Zhang and M. Wu, "D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 2018, pp. 192-1924, doi: 10.1109/CVPRW.2018.00034.

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

[22] X. Gao et al., "An End-to-End Neural Network for Road Extraction from Remote Sensing Imagery by Multiple Feature Pyramid Network," in IEEE Access, vol. 6, pp. 39401-39414, 2018, doi: 10.1109/ACCESS.2018.2856088.

[23] C. C. T. Mendes, V. Frémont and D. F. Wolf, "Exploiting fully convolutional neural networks for fast road detection," 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 2016, pp. 3174-3179, doi: 10.1109/ICRA.2016.7487486.

[24] Y. Li, L. Guo, J. Rao, L. Xu and S. Jin, "Road Segmentation Based on Hybrid Convolutional Network for High-Resolution Visible Remote Sensing Image," in IEEE Geoscience and Remote Sensing Letters, vol. 16, no. 4, pp. 613-617, April 2019, doi: 10.1109/LGRS.2018.2878771.

[25] Y. Li, L. Xu, J. Rao, L. Guo, Z. Yan, and S. Jin, "A Y-Net deep learning method for road segmentation using high-resolution visible remote sensing images," Remote Sensing Letters, vol. 10, no. 4, pp. 381-390, Apr. 2019. https://doi.org/10.1080/2150704X.2018.1556498.

[26] Z. Zhong, J. Li, W. Cui and H. Jiang, "Fully convolutional networks for building and road extraction: Preliminary results," 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 2016, pp. 1591-1594, https://doi.org/10.1109/IGARSS.2016.7729406.

[27] X. Xin, Y. Jiang, X. Zhang, Z. Zhang, and W. Fang, "Road extraction of high-resolution remote sensing images derived from DenseUNet," Remote Sensing, vol. 11, no. 21, p. 2499, Oct. 2019. https://doi.org/10.3390/rs11212499.

[28] Y. Xu, Y. Feng, Z. Xie, A. Hu and X. Zhang, "A Research on Extracting Road Network from High Resolution Remote Sensing Imagery," 2018 26th International Conference on Geoinformatics, Kunming, China, 2018, pp. 1-4, https://doi.org/10.1109/GEOINFORMATICS.2018.8557042

[29] S. Wang, X. Mu, D. Yang, H. He, and P. Zhao, "Road extraction from remote sensing images using the inner convolution integrated encoder-decoder network and directional conditional random fields," Remote Sensing, vol. 13, no. 3, p. 465, Jan. 2021. https://doi.org/10.3390/rs13030465.

[30] I. Atli and O. S. Gedik, "Sine-Net: A fully convolutional deep learning architecture for retinal blood vessel segmentation," Engineering Science and Technology, an International Journal, vol. 24, no. 2, pp. 271-283, Feb. 2021. https://doi.org/10.1016/j.jestch.2020.09.013.

[31] L. Chen, Q. Zhu, X. Xie, H. Hu, and H. Zeng, "Road extraction from VHR remote-sensing imagery via object segmentation constrained by Gabor features," ISPRS International Journal of Geo-Information, vol. 7, no. 9, p. 362, Sept. 2018. https://doi.org/10.3390/ijgi7090362.

[32] V. Mnih, "Machine learning for aerial image labeling," PhD thesis, University of Toronto, Toronto, ON, Canada, 2013.

[33] A. Tharwat, "Classification assessment methods," Applied Computing and Informatics, vol. 17, no. 1, pp. 168-192, Jan. 2021. https://doi.org/10.1016/j.aci.2020.11.001.

[34] H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia, "Pyramid Scene Parsing Network," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 6230-6239, https://doi.org/10.1109/CVPR.2017.660.