Road Network Extraction Using Atrous Spatial Pyramid Pooling

Aravapalli Sri Chaitanya, Suvarna Vani Koneru, Praveen Kumar Kollu

Abstract: Road extraction from satellite images has several Applications such as geographic information system (GIS). Having an accurate and up-to-date road network database will facilitate transportation, disaster management and GPS navigation. Most active field of research for automatic extraction of road network involves semantic segmentation using convolutional neural network (CNN). Although they can produce accurate results, typically the models give up performance for accuracy and vice-versa. In this paper, we are proposing architecture for semantic segmentation of road networks using Atrous Spatial Pyramid Pooling (ASPP). The network contains residual blocks for extracting low level features. Atrous convolutions with different dilation rates are taken and spatial pyramid pooling is performed on these features for extracting the spatial information. The low level features from residual blocks are added to the multi scale context information to produce the final segmentation image. Our proposed model significantly reduces the number of parameters that are required to train the model. The proposed model was trained on the Massachusetts roads dataset and the results have shown that our model produces superior results than that of popular state-of-the art models.

Index Terms:. Convolutional neural network, Geographical information systems, Road network, Semantic segmentation.

I. INTRODUCTION

Having an up-to-date road network database can facilitate a large number of applications. Most of the times the database is far behind the current information. It is due to the fact that most of the road network extraction is done manually. As it is done at regular intervals it is not feasible to have a real time updates to the dataset. So there is a need for robust automatic extraction methods that have a real time performance and good accuracy. There are several approaches for automatic extraction. Most of them are based on image processing techniques or classification techniques or a combination of both. The main problems of these approaches are they are very specific to a dataset. As there are many variations of the roads and the angles of capturing from the satellite many methods struggle when the type of the road is changed or the image has variation from the angle it was captured on.

Recent advancements in the computer vision algorithms especially using the deep learning techniques have shown an unprecedented accuracy in image classification and segmentation. Most of the popular methods for road network extraction

are based on deep learning techniques. The classification is done at pixel level where each pixel is labelled as either road or non road. This pixel level classification produces a segmented image that contains the road network. Deep convolution neural network based models have shown near human accuracy and robust performance. One of the most popular approaches was fully convolutional networks [1] [10] [11] [14].

Initial application of deep neural networks for road network extraction was proposed by Mnih and Hinton [2]. Their work employs restricted Boltzmann machines for automatic segmentation of road areas. Cheng et al. [3] employed a cascaded end-to-end CNN architecture (CasNet) for road detection and centreline extraction. The model contains two cascaded networks, initial network detects the road pixels and latter network extracts the centerline. Zhang et al. [4] combined residual units with U-Net architecture called residual U-Net (ResUnet) for semantic road segmentation. It helps facilitate information propagation and can handle vanishing gradient problem.

Although deep neural networks can offer best results, they require large number of training parameters and significant amount of time is necessary to train them. Inspired from the DeepLab3+ [5] [12] [13]model and residual units [4] [5], we are proposing a semantic segmentation model that utilizes Atrous Spatial Pyramid Pooling (ASPP) for semantic road network segmentation. The proposed model uses residual units for extracting low level features which are given as input to the ASPP network that extracts multi-scale contextual information. The low level features are concatenated with the multi scale contextual information to detect smooth road network boundaries. A bilinear up-sampling is used to produce the final segmentation mask. The results have shown that our proposed model produces superior results than that of the state-of-the art models while requiring only half the parameters for training.

II. METHODOLOGY

A. Residual unit

 Deeper neural networks contain large number of hidden layers. When the model increases in depth it is hard to propagate the information throughout the network. This problem is termed as vanishing gradient problem. He et al. [6] proposed to use residual function mapping with reference to the layers. The network contains several residual units that facilitates information propagation. Each residual unit can be given as:

$$y = F(x, \{W_i\}) + x \qquad (1)$$

B. Atrous Convolution

Atrous convolutions are new convolution operations which perform convolutions with holes in the feature maps [12] [13]. Holes is the stride between the values in a single map. They are computationally efficient than the normal convolution [5]. The atrous convolution is given as:

$$y[i] = \sum_k x[i + r.k]w[k] \qquad (2)$$

Where w is the convolution filter and r is the stride between the filters.

C. Network Architecture

Training a semantic segmentation network requires large Deeper networks with large number of parameters. This can impact the performance of the model. Context information is also important for segmentation. Our proposed network contains residual units for extracting low level semantic features from the input vector. Atrous convolutions with multiple dilation rates are performed parallel on the filters from the residual blocks. Spatial Pyramid Pooling is performed to gather the context information. In our architecture we are considering 1,6,12 and 18 which were found to extract finer details from road structures. It is to be noted that dilation rate of 1 gives standard convolution.

The multi scale filters from the atrous convolutions are concatenated along with the normal convolution with bilinear up-sampling. For the final segmentation mask the low-level semantic features are concatenated with the spatial features and final bilinear up-sampling is performed to get the desired spatial resolution. By leveraging Atrous convolutions and Spatial pyramid pooling considerably reduces the depth of the network while producing more context information. The proposed architecture is illustrated in Fig. 1.
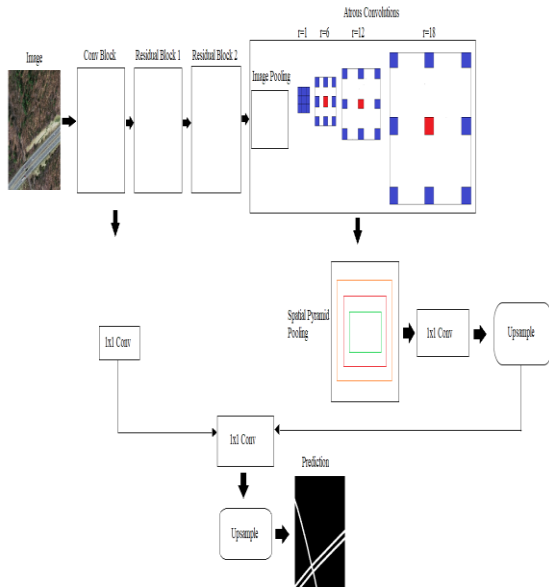

Fig. 1: Proposed network architecture

III. EXPERIMENTS

To evaluate our proposed architecture, the model was trained on Massachusetts roads dataset. Keras was used as deep learning framework for implementing our model [7].

A. Dataset

Mihn et al. [2] produced the Massachusetts roads dataset. It contains 1108 images as training set, 14 images as validation set and 49 images as test set. All the images are in .tif format with resolution of 1500x1500 that covers 1.2 meter per pixel.

B. Implementation details

For training our architecture we have used dice coefficient loss as loss function. Adam [8] was used as optimizer with an initial learning rate of 0.001 and was reduced by 0.0001 when the training loss is decreasing slowly. Due to memory limitations we have trained our architecture on an image size of 250x250. Around 17K images are generated from the original training images and were fed in to the network. The batch size was fixed at 10 and it took around 60 epochs to achieve convergence.

C. Evaluation Metrics

We have evaluated our model using Intersection over Union (IoU) metric. IoU metric measures the overlap between the target mask and the predicted mask. It is given as:

$$IoU = \frac{TM \cap PM}{TM \cup PM} \qquad (3)$$

D. Results

We have evaluated our proposed architecture on test data of Massachusetts roads dataset. The total number of parameters in our proposed model is compared against other popular semantic segmentation models U-Net and ResUnet. The comparisons are reported in Table I. It can be observed that our model takes least number of parameters than the other models. We have also evaluated IoU metric for our predicted masks which are reported in the Table II. Our model gives an average IoU of 0.68 on the test set. Some examples of predictions of our network was given in Fig.2 It can be seen that our model was able to detect roads from occlusions especially from trees. It effectively detected shallow and parallel roads.

TABLE I: Comparisons of training parameters for different road extraction methods on massachusetts road dataset

Model	Parameters(in Million)
U-Net [9]	30.6M
ResUnet [4]	7.8M
Proposed Architecture	3.2M

TABLE II: Comparisons of IoU metric for different road extraction methods on massachusetts road dataset

Model	IoUmetric
U-Net [9]	0.63
ResUnet [4]	0.65
Proposed Architecture	0.68

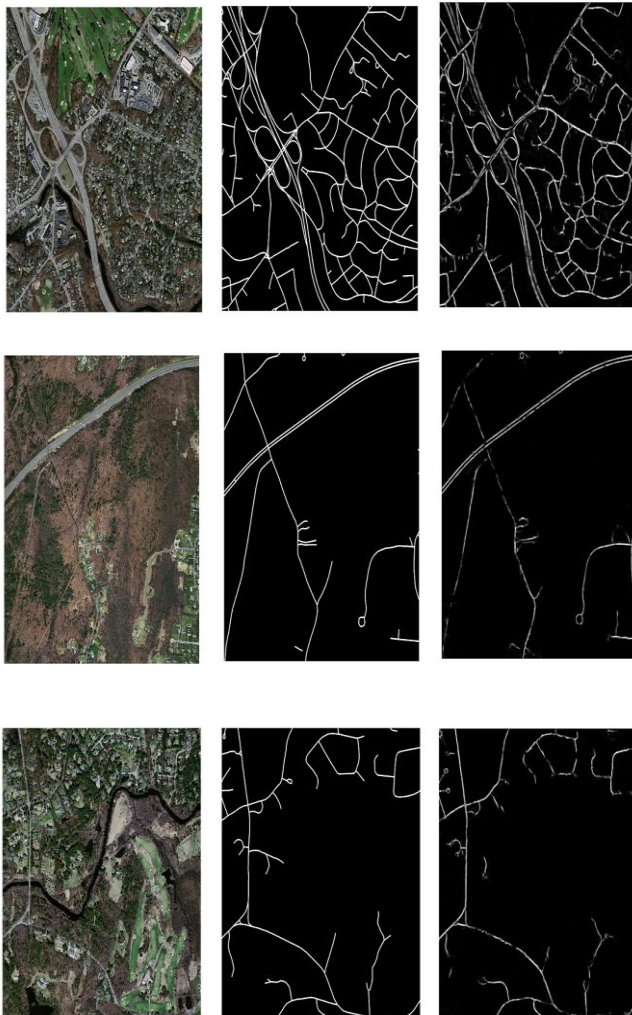
Fig. 2: Example predictions on the test set (a) Original Image (b) Ground truth (c) Predicted mask

IV. CONCLUSION

Road Network extraction using deep neural networks requires a large number of parameters. Most models fail to detect finer spatial information which is important for image segmentation. In this letter, we are proposing architecture for extracting road network from high resolution satellite images. Our proposed architecture has less number of parameters compared to other popular models in road extraction. The Atrous spatial pyramid pooling (ASPP) facilitates capturing context information and also helps controlling the field-of-view. The proposed architecture has shown superior results in extracting the road network. The ultimate purpose that the knowledge obtained from the proposed Atrous spatial Pyramid pooling was designed for road network extraction.

REFERENCES

1. J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks forsemantic segmentation," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 3431-3440.
2. V. Mnih and G. E. Hinton, "Learning to detect roads in high-resolution aerial images," in Proc. ECCV, 2010, pp. 210–223.
3. G. Cheng, Y. Wang, S. Xu, H. Wang, S. Xiang, and C. Pan, "Automatic road detection and centerline extraction via cascaded end-to-end convolutionalneural network," IEEE Trans. Geosci. Remote Sens., vol. 55,no. 6, pp. 3322–3337, Jun. 2017.
4. Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. "Road extraction by deep residual u-net." In IEEE Geoscience and Remote Sensing Letters.IEEE, 2018.
5. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.,"Encoderdecoder with atrous separable convolution for semantic image segmentation." arXiv preprint arXiv 1802.02611 (2018).
6. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in CVPR, 2016, pp. 770–778.
7. F. Chollet et al. (2015). Keras. [Online]. Available: https://github.com/fchollet/keras
8. Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
9. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. MICCAI, 2015, pp. 234–241.
10. Favyen Bastani, Songtao He, Sofiane Abbar, Mohammad Alizadeh, Hari Balakrishnan, Sanjay Chawla, Sam Madden, David DeWitt, "Road-Tracer: Automatic Extraction of Road Networks from Aerial Images", arXiv:1802.03680, 2018.
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.
12. L. C. Chen G. Papandreou I. Kokkinos K. Murphy A. ;L. Yuille "Semantic image segmentation with deep convolutional nets and fully connected crfs" Computer Science vol. 4 pp. 357-361 2014.
13. L. C. Chen G. Papandreou I. Kokkinos K. Murphy A. L. Yuille DeepLab: Semantic Image Segmentation with Deep Convolutional Nets Atrousn Convolution and Fully Connected CRFs 2016.
14. G. M´attyus, W. Luo and R. Urtasun, "DeepRoadMapper: Extracting Road Topology from Aerial Images," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 3458-3466, 2017.

AUTHORS PROFILE

Aravapalli Sri Chaitanaya is studying M.tech in Department of Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College. Areas of interest are, deep learning, Machine learning, bioinformatics and cyber security.

Dr. Suvarna Vani Koneru is working as professor in Department Science and Engineering ,Velagapudi Siddhartha Engineering College and 20+ experiences in Teaching .Her areas of interest are Bio-informatics, Machine learning, Deep learning.

Praveen Kumar Kollu is working as Sr. Assistant professor in Department of Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College and He is having 15+ experiences in Teaching. His areas of interest are Network intrusion detection, deep learning, Machine learning, bioinformatics and cyber security.