Sri Lanka Society of Transport and Logistics

Journal of South Asian Logistics and Transport
Volume 3 Issue 1 March 2023
JSALT 3.1.4: pp. 59-73
DOI: http://doi.org/10.4038/jsalt.v3i1.60

ROAD SEGMENTATION: EXPLOITING THE EFFICIENCY OF SKIP CONNECTIONS FOR EFFICIENT SEMANTIC SEGMENTATION

Madhumita D [a], Bharath H A [a *], Devendra V P [a] and Shivam B [b]

[a] RCG School of Infrastructure Design and Management, IIT Kharagpur, West Bengal, India

[b] Department of Electrical Engineering, IIT Kharagpur, West Bengal, India

* Correspondence should be addressed to *bharath@infra.iitkgp.ac.in*

ABSTRACT

Extraction of features is a unique approach that has several real-time applications. In remote sensing, road feature extraction includes recognising and extracting essential information about roads from high-resolution remote sensing photos. This data may include the shape, size, direction, and connection of roadways and their surroundings we can increase the accuracy and efficiency of road recognition algorithms and better comprehend the geographical distribution of roads in metropolitan settings. Nevertheless, traditional techniques of road recognition sometimes rely on manual labour or hand-crafted characteristics, which may be costly and time-consuming. Convolutional neural networks (CNNs) have been frequently employed for road detection due to their capacity to automatically learn characteristics from data. In this research, we provide Modified LinkNet for the extraction of road features from high-resolution remote-sensing images. The proposed model incorporates several variables resulting in superior performance compared to current cutting-edge road detection techniques. SpaceNet road detection dataset yields 0.81 mIoU, which is approximately 0.1, 0.11, and 0.03 mIoU greater than U-net, DeeplabV3+, and LinkNet, respectively (base architecture). In addition, the training of the model enables more efficient inference making our method applicable to real-time applications and are therefore a more effective and efficient solution for the extraction of road features from remote sensing images.

Keywords: *Road segmentation, Semantic segmentation, Convolution network, Machine learning*

1. INTRODUCTION

The extraction of roads in urban areas is one of the most sought-after functions of remote sensing imaging. Numerous industries - including the military, agriculture, environmental science, urban planning, unmanned vehicles, geospatial data integration, disaster recovery, and rail hailing - rely heavily on road information available in real-time.

Remote sensing images are captured by satellites with high precision [1]. Depending upon the location, remote sensing has different use cases. Urban remote sensing captures buildings, roads, trees, and water bodies. Road extraction using remote sensing images poses many challenges [2]. Due to vehicles, trees, and structures, the discontinuity or incompleteness of roadway imaging is one of the greatest obstacles. The occlusion or 'noise' created by shadows cast by buildings, trees, and cars is an additional significant obstacle in remote sensing imagery.

Moreover, conventional road extraction from remotely sensed imagery could be made more efficient and practical; present methods do not meet the demands for real-time processing [3-4]. Traditional methods are based on pixel-level information such as support vector machine, random forest, and maximum likelihood; because they are limited to the subject of colour phenomena, these methods use only the spectral information of images [5-6]. These methods use colour reflectance to classify images, which leads to a loss of information with regions of similar colour and backgrounds [7]. Over the past several years, a variety of methods for the extraction of roads have been proposed. However, road extraction can be broadly divided into, but not limited to, two categories: road area extraction, which aims to get complete road information (pixel-to-pixel), and road centreline extraction (single-pixel), which seeks to bring a strip of information from the road centreline. In addition, the road centreline can be derived from road area extraction using morphological thinning techniques; hence, the majority of research focuses on road area extraction [8].

Deep learning approaches employing Convolutional Neural Networks (CNNs) use data and processing resources to produce promising state-of-the-art results [9]. Despite the growing availability of data and computational power, there are constraints. To attain better results, we still require huge datasets, and the memory restrictions of GPUs (Graphics Processing Units) make deep learning approaches rather difficult [10, 11]. Traditional techniques use manually crafted feature extractors and pixel classification based on these characteristics. Handcrafted feature extractors must be meticulously built for each region. However, deep learning algorithms bypass the standard procedures and have significantly enhanced the results. Fully Connected Networks (FCN) were initially implemented in 2015 for

semantic segmentation. These FCNs, however, were not effective in preserving spatial information [7,12]. U-shaped network topologies were suggested as a solution to this issue. Studies revealed that shallow networks had limited learning capabilities and that computer vision tasks benefited from deeper networks; nonetheless, they were plagued with gradients that vanished over time. To address this challenge, it was proposed to train deeper networks using deep residual networks.

2. LITERATURE REVIEW

Feature extraction is vital to several applications in the fields of transportation planning, urban planning, and disaster response [13]. Accurate road detection is required in each of these applications because it can provide useful information about the geographical distribution of roads in urban and rural locations. Convolutional neural networks (CNNs) have been frequently employed for road detection in recent years due to their capacity to automatically learn features from data. Yang [14] proposed spatially improved and densely connected Unet (SDUNet), a refinement to the encoder-decoder architecture of UNet. Using the geometric topology of the road network in conjunction with the densely connected encoder block and spatial intensifier (DULR module), SDUNet was able to discover the road topology in all four cardinal directions. However, this causes an increase in the number of architectural components and a reduction in time efficiency. Lu et al. [15] proposed Spatial Information Inference Structure, an RNN-based information Processing unit that may be incorporated into any conventional deep learning framework for segmentation. The proposed technology might resolve occlusions in satellite photos and maintain road network continuity.

Hong et al. [16] employed the notion of vector field learning in road extraction to construct three separate learnable vectors to characterise the road centerline, road surface area, and road margins. The model was constructed to mine all of the road network's deep features and used dense atrous convolution and spatial pyramid pooling. The suggested Road-RCF model was unable to extract road features when the road was obscured by shadows cast by nearby objects. Due to this, the model was unable to extract the road width efficiently and precisely. Additionally, the issue of mixed pixels exists. Li et al. [17] proposed a unique patch-wise semantic segmentation algorithm for remote sensing applications that handles large-scale image sizes. Due to many patches, feature information becomes distorted, preventing the model from identifying roadways in complicated urban environments. In the extraction of roadways from a complex urban landscape, the integration of border and contextual information plays a significant role.

Zhou et al., [18] explored the image pre-processing technique with histogram equalisation resulting in better contrast enhancement by nearly one per cent. Due to limited resources, the authors combined the prediction output with the original image. In addition, the output was combined with a post-processing module to enhance connectivity outcomes. Kestur et al. [19] compared the results of the proposed U-shaped Fully Connected Network (UFCN) with standard well, known state of art algorithms and concluded that the UFCN was faster and more stable for real-time data acquisition operations. Abolfazl [8] A RoadVecNet architecture including Squeeze-Excitation and Dense Dilated Spatial Pyramid Pooling modules was presented for concurrent road segmentation and vectorization. A focused loss weighted by the median frequency balancing loss function was utilised to overcome extremely unbalanced datasets with few positive values. The results of the experiment reveal that the model was frail and ineffective when covered by shadows. There were discontinuities in road fragments throughout extensive and continuous obstacle zones.

In a related article, Abolfazl [9] introduced SC-RoadDeepNet, a road extraction deep learning-based network that improves road network connection. The model does not account for the road network's centreline and is confined to the perimeter of semantic roads. Lu et al. [20] used an improved bilateral filtering method that considers edge characteristics, an upgraded marker watershed based on LAB colour space, and to solve the difficulty of recovering urban impermeable surfaces from high-resolution remote sensing images, fuzzy c-means clustering is utilised. In an infinite loop, Li et al. [21] paired the module for deep learning with the module for knowledge-guided ontological reasoning. The first layer enriches the model with low- and medium-level feature data, hence enhancing classification performance. The latter offers domain expertise at a high level to the first layer and steers misclassifications from the first layer. However, the model does not take contextual information into account which is very important in such studies.

According to the survey conducted by Song et al. [22], the present uses of convolutional neural networks (CNNs) in remote sensing (RS)-based image classification are equal to computer vision classification problems. Typical and prevalent applications involve scene classification, object detection, and object segmentation. Therefore, more emphasis might be placed on applications such as occlusions, super-resolution reconstruction, and multi-label retrieval of RS images.

Xie et al. [23] presented a multi-modal framework based on deep learning for the identification of urban functional areas. The design employs an attention method to extract the crucial characteristics of multimodal data, hence enhancing the linkage. The findings of the prediction, however, indicate that the addition of multimodal data,

such as building outline data, affects classification accuracy. Wang et al. [24] developed a UNet-like Transformer for the effective semantic segmentation of urban scene photos with remote sensing imageries. A global-local transformer block and a feature refinement head were utilised to optimise the context extraction of urban features. Through benchmark trials on multiple datasets, they demonstrated the usefulness and efficiency of their technique, obtaining state-of-the-art performance on the Vaihingen dataset. Nevertheless, the new local context layer lacks spatial consistency, and the global context layer lacks location knowledge of semantic properties. Liu et al. [25] proposed a method for concurrently extracting and detecting lane-marking characteristics in a variety of road conditions. To identify lane markers, the approach employs an adaptive hat filter and a connected graph structure. The experimental results suggest that the strategy outperforms previous methods on the KIST and Caltech datasets, although at the cost of broken networks. CNN's results, however, were ineffective in reliably detecting and extracting roadways in complicated urban environments with circular tracts and roundabouts.

Consequently, the expected results continue to contain flawed fragments and patches. In this study, we offer a redesigned architecture that effectively learns the image's feature mapping while preserving the image's structural integrity. The design is influenced by Unet and LinkNet, although the network is more robust than previous systems. The results of experiments indicate that the suggested road extraction network could precisely optimise segmentation outcomes. Compared to previous experimental methods, this prediction was superior in terms of precision and connectedness.

3. METHODOLOGY

3.1. DATASETS

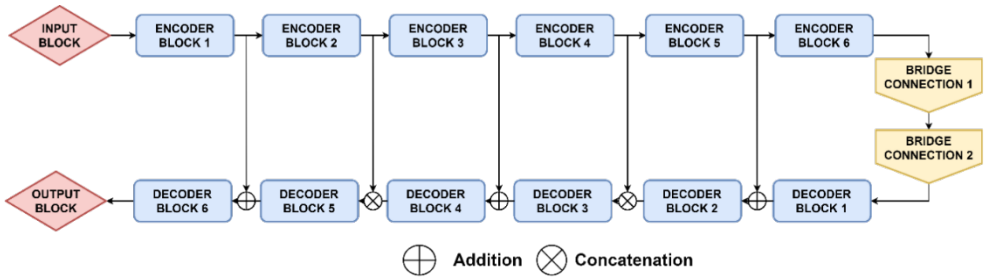
We have used the SpaceNet 3 (2018): Road Network Detection data set [26] for the scope of this paper. The SpaceNet data collection is a benchmark dataset that is publicly accessible. The data set covers 800 km of roads in four different cities: Las Vegas (United States, North America), Shanghai (China, Asia), Paris (France, Europe), and Khartoum (Sudan, Africa). Each region has 400 m x 400 m (1300 pixels x 1300 pixels) tiles of images. Each area of interest (AOI) has a multi-spectral (MUL) image, pansharpened (PAN) image, multi-spectral fused pansharpened (MUL-PanSharpen) image, and 3-band pansharpened (MUL-PanSharpen) image. The data set included training and testing images of heterogenous and complex urban scenarios. However, the testing data set does not have labels.

Hence, only the training data set has been used for experimentation purposes. The total number of images with corresponding masks considered for each region is different - 989 for Las Vegas, 310 for Paris, 919 for Shanghai, and 283 for Khartoum. Apart from the number of images for each region, It is observed that the images' quality is also different within the same area. Khartoum had 153 corner images with mainly black regions and high brightness compared to the other images.

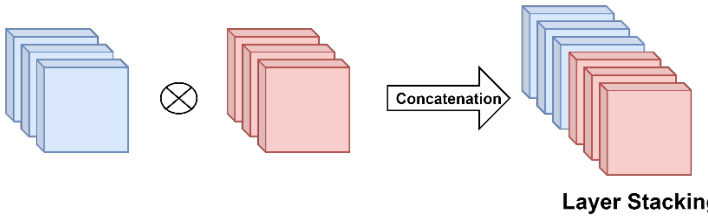
Similarly, Shanghai had 166 corner images, Paris had 78 corner images, and Las Vegas has 95 corner images. The data set labels contain the centreline of roads. The mask generated from each label is 2m in width. For our experimentation, 878 images of Las Vegas have been split randomly into 80% training images and 20% validation images.

3.2. MODEL ARCHITECTURE

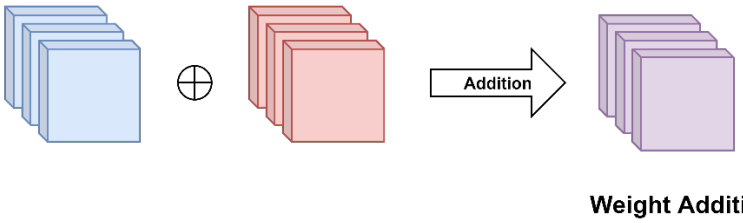
Figure 1 shows the model architecture of our Modified LinkNet for road extraction. UNet and LinkNet inspire the proposed architecture. The modified LinkNet includes two bridge connections linking the model's encoder and decoder blocks. The model has more trainable parameters compared to the base architecture due to the increased number of encoder and decoder blocks, which contributes to the model's accurate learning.



a. Model architecture of Modified LinkNet



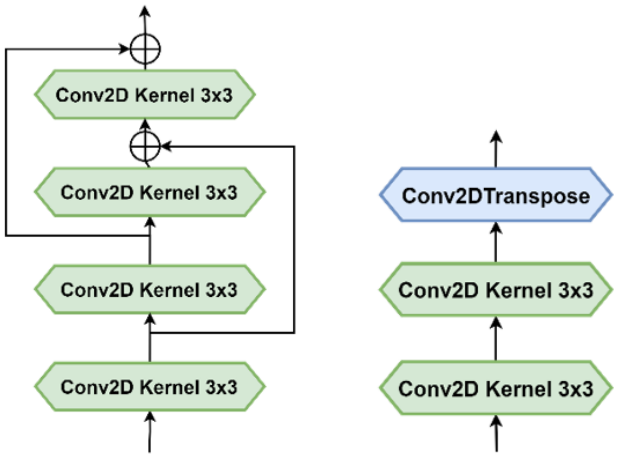
b. Model structure of concatenation layer used in Modified LinkNet



c. Model structure of additional layer used in Modified LinkNet

Figure 1: The model architecture of the proposed Modified LinkNet

The input block undergoes immediate downsampling twice in a row, resulting in a reduced image and faster model processing with efficient prediction results. We have introduced skip connections between the encoder blocks that prevent gradient exploding. It also allows the prior layer to automatically skip to the next in case of gradient exploding. The added shortcut connections enable the model to transfer the spatial information to each decoder. Hence, the model can effectively recognise the fine-grained feature details.



a. Encoder-block design b. Decoder-block design

Figure 2: Structure of encoder and decoder blocks of Modified LinkNet

We use a similar structure to LinkNet; however, it differs from the original LinkNet in terms of the filter number used in different layers. The encoder block shown in fig 2(a) has filter numbers 64, 64, 128, 128, 256, and 512. At the same time, the decoder shown in fig 2(b)has filter numbers 128, 64, 32, 32, 16, and 16. Furthermore, we have introduced two bridge connections between the encoder and decoder layer to retain

the feature loss in the intermediate connection. The bridge connection preserves the sudden loss in filter number and makes the smooth flow of information from the encoder to the decoder layer.

Since U-Net and LinkNet inspire our model, both have concatenation and addition layers. Therefore, we have alternately introduced addition and concatenation layers shown in fig 1 (b) and (c) in the network architecture. This allows us to extract more features because the additional; layer provides weight addition, which increases the likelihood of accurate prediction. On the other hand, concatenation stacks together all of the feature maps from the current layer and any layers that came before it. This technique also makes the model efficient in predicting quick results and reduces the training time.

3.3. IMPLEMENTATION DETAILS

Using the Keras framework, this model has been trained on a computing cluster with Ubuntu 18.04.6 LTS, 128GB VRAM 3 and 192GB RAM. Adam (Adaptive Moment Estimation) was chosen as the optimiser with a default learning rate of 0.0001 to train our model because of its excellent converging power. Batch size is taken as eight per batch, and step size is set as a variable which means the model will automatically adjust step size. Call-back has been provided with the condition of val_loss and patience of four which means for consecutive four times, if val_loss doesn't decrease, then model training will automatically get stopped to avoid overfitting. The normal kernel initialiser has been utilised, which is based on the standard deviation and probabilistic values of the images [27]. This initialiser analyses the image set and sets the kernel values so we can get a fair amount of accuracy at the start.

3.4. EVALUATION METRICS

Four metrics were used to evaluate the performance of road extraction methods: these were mIoU, F1score, precision, and recall. Although there are numerous criteria for evaluating the efficacy of an experiment, including graph-based methods for centreline extraction. However, because the focus of this study is limited to pixel segmentation, we evaluated the performance of projected outcomes using pixel-based metrics. mIoU computes intersection over union for each pixel class (two in this study for road and non-road) before calculating the average over classes. The F1 score measures the precision of a model. The harmonic mean of recall and precision is used to calculate it. Precision is the proportion of a model's correctly predicted positive outcomes to all its correctly predicted positive outcomes. In other words, precision quantifies the fraction of positive forecasts that are accurate. A model with high precision is adept at avoiding false positives. Nevertheless, this does not necessarily imply that the model has a high recall (i.e., that it is good at recognising all genuine

positive instances). The recall is the proportion of correctly predicted positive cases to all the actual positive instances in the data set. In other words, recall measures how many genuine positive cases the model was able to recognise. A high recall suggests that all genuine positive events are detected by the model.

$$\text{mIoU} = \frac{\text{Class 1+Class 2+}\cdots\text{+Class n}}{\text{n}} \text{ ----------(1)}$$

$$\text{F1} - \text{score} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \text{ ----------(2)}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \text{ ----------(3)}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \text{ ----------(4)}$$

where TP is the number of rightly predicted road pixels, FP represents the number of incorrectly predicted road pixels, and FN represents the number of incorrectly predicted non-road pixels for the remote sensing image.

3.5. RESULTS AND ANALYSIS

On SpaceNet datasets, we compared our proposed model to three existing cutting-edge semantic segmentation techniques. The three cutting-edge approaches for semantic segmentation are U-Net [28], DeepLabV3+ [29], and the original LinkNet [30]. During model training, the training settings were identical to our recommended model settings.

Table 1: Comparison results with various segmentation methods experimented on SpaceNet

Proposed Model	mIoU	Precision	Recall	F1score
U-Net	0.71	0.72	0.65	0.68
DeepLabV3+	0.70	0.71	0.62	0.66
LinkNet	0.78	0.77	0.76	0.77
Modified LinkNet	0.81	0.79	0.80	0.79

Table 1 compares the mIoU, Precision, Recall, and F1score between our Modified LinkNet and three other cutting-edge semantic segmentation techniques. Our proposed model gets the highest mIoU score (0.81) among the other three semantic segmentation approaches. The mIoU scores for the remaining procedures are approximately 0.71, 0.70, and 0.78, respectively. In the experiment, the proposed technique achieved the highest mIoU score, which is approximately 0.1, 0.11, and 0.03 points higher than the other three state-of-the-art methods, respectively.

The visual findings indicate the potential of the proposed deep learning algorithm for retrieving high-resolution satellite images. To avoid misclassification of edges, segmentation is performed using a kernel size of 512 x 512 with a 23 per cent overlap. We adjusted the settings of a convolutional neural network model to conduct binary segmentation and distinguish road features.

Figure 3 depicts a visual comparison of the expected road maps based on the comparison of architectures. The majority of road areas can be accurately anticipated by all four methods. Based on exact observation, we can distinguish the forecasts of Unet, DeepLabV3+, and LinkNet, which tend to disrupt road connectivity. Existing models failed to maintain conformance with the road map amid the urban environment's complexity. Using the different layer configurations of the enhanced Modified LinkNet might preserve the road's topology and significantly enhance the road's connectivity and smoothness. The observed performance metrics conclusively illustrate that our suggested Modified LinkNet offers increased performance. The efficacy of the model was demonstrated by the fact that its accuracy was better than 97%. However, the other three models achieved greater than 70 per cent precision but were unable to preserve the road's smoothness.

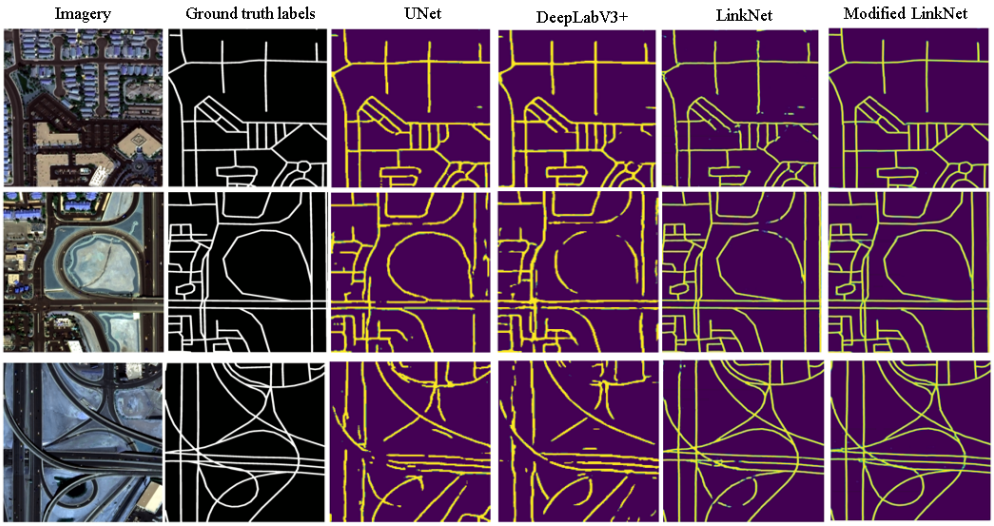


Figure 3: Visual comparison results on the SpaceNet dataset

As can be seen, the first image depicts a business area with tall buildings, but the generated model accurately anticipated all the roadways. All of the streets are present in the results of the prediction, with continuous and end-to-end linkages. The model accurately predicted all road continuity elements in the second image, which depicted a region containing freeways and intersections. From a thorough inspection of the

result, we can conclude that the model can preserve the connectedness between road pixels, which is one of our primary objectives.

According to the available literature, connection performance is crucial when addressing the topology of a road network. Accordingly, the experimental analysis shows that the proposed model is sufficiently fundamental and robust for use in real-time applications. The model also can do topological and global knowledge-based semantic segmentation of remote-sensing images of any dimension. The proposed model could preserve the structural integrity of the road maps which in turn reduces the distortion enormously. The quantitative outcomes demonstrate the model's effectiveness in extracting roads from high-resolution remote-sensing images.

4. CONCLUSIONS AND FUTURE WORK

This study developed a convolutional neural network model based on deep learning to extract road properties from high-resolution remote sensing images. This paper proposes a modified LinkNet architecture to extract road features from high-resolution remote sensing images to improve the efficiency of the existing LinkNet model, which was influenced by Unet. The objective is to maximise the effectiveness of the network design layers while maximising the usage of limited resources.

We have established a bridge connection and used concatenation and addition to efficiently recover the lost features. Unlike full-fledged deep learning architectures, this model is designed to make effective use of network parameters, allowing it to execute flawlessly with fewer computational resources. Because our architecture has more encoder and decoder blocks than the baseline architecture, it has more trainable parameters, allowing the model to acquire knowledge more precisely.

The improved LinkNet model can be used in real-time applications and performs large-scale computations more quickly. The model’s performance was evaluated using the SpaceNet benchmark dataset and compared to other conventional deep learning architectures for picture segmentation. The quantitative and visual examination demonstrates the model's capability for semantic segmentation. The proposed architecture was capable of capturing context information and maintaining the road network's connectivity. The redesigned network's mIoU score is 0.03 percentage points higher than the baseline architecture. Consequently, the study makes a successful effort to illustrate the efficient application of deep learning network parameters.

Future road features will allow researchers to examine the texture and structure of road networks. Given the availability of road topology, additional research on deep learning-based models can be conducted to broaden their usefulness in the

transportation sector for testing the extracted road features for optimal planning and maintenance.

ACKNOWLEDGMENT

We thank the Indian Institute of Technology Kharagpur and IHub mobility of IIIT Hyderabad for infrastructural and financial support.

REFERENCES

[1] B. H. Aithal and R. T. V, "*Urban growth patterns in India: Spatial analysis for sustainable development*," S.l.: CRC PRESS, 2023, doi:10.1.1.368.2094.

[2] Tran, A. Zonoozi, J. Varadarajan, and H. Kruppa, "PP-LinkNet: Improving semantic segmentation of high-resolution satellite imagery with multi-stage training," *Proceedings of the 2nd Workshop on Structuring and Understanding of Multimedia heritAge Contents*, 2020, doi: 10.1145/3423323.3423407.

[3] A. Abdollahi, B. Pradhan and A. Alamri, "VNet: An End-to-End Fully Convolutional Neural Network for Road Extraction From High-Resolution Remote Sensing Data," in IEEE Access, vol. 8, pp. 179424-179436, 2020, doi: 10.1109/ACCESS.2020.3026658.

[4] B, H.A., Vinay, S., and R, T.V., "Characterization and visualisation of Spatial Patterns of Urbanisation and Sprawl through metrics and modelling, Cities and the Environment (CATE)," vol.10, no. 1, pp. 1–10.

[5] P. P S, J. Soni and B. H A, "Building Extraction from Remote Sensing Images Using Deep Learning and Transfer Learning," IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, 2022, pp. 3079-3082, doi: 10.1109/IGARSS46834.2022.9883898.

[6] P. P S, K. D. Soumya and B. H A, "Urban building extraction using satellite imagery through machine learning," in *IEEE symposium series on computational intelligence 11 (SSCI), Bangalore, India, November* 2018, doi: 10.1109/SSCI.2018.8628919

[7] Madhumita, D. and B. H.A., "Semantic Segmentation of High-Resolution Satellite images: a Deep Learning Approach", *J. of Geomatics*, vol.16, no. 1, pp. 1–10, 2022, doi: 10.6084/m9.figshare.21675008.

[8] A. Abdollahi, B. Pradhan, and A. Alamri, “RoadVecNet: A new approach for simultaneous road network segmentation and vectorisation from aerial and Google Earth imagery in a complex urban set-up,” *GIScience & Remote Sensing*, vol. 58, no. 7, pp. 1151–1174, 2021, doi: 10.1080/15481603.2021.1972713.

[9] A. Abdollahi, B. Pradhan, and A. Alamri, “SC-roaddeepnet: A new shape and connectivity-preserving road extraction deep learning-based network from Remote Sensing Data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022, doi: 10.1109/TGRS.2022.3143855.

[10] P. Ps and B. H. Aithal, “Building footprint extraction from very high-resolution satellite images using Deep Learning,” *Journal of Spatial Science*, pp. 1–17, 2022, doi:10.1080/14498596.2022.2037473.

[11] V. Pereira, S. Tamura, S. Hayamizu and H. Fukai, "A Deep Learning-Based Approach for Road Pothole Detection in Timor Leste," 2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), 2018, pp. 279-284, doi: 10.1109/SOLI.2018.8476795.

[12] N. B. Devi, A. C. Kavida, and R. Murugan, “Feature extraction and object detection using fast-convolutional neural network for Remote Sensing Satellite Image,” *Journal of the Indian Society of Remote Sensing*, vol. 50, no. 6, pp. 961–973, 2022, doi: 10.1007/s12524-022-01506-x.

[13] P. S. Prakash, G. Nimish, M. C. Chandan, and H. A. Bharath, “Urbanization: Pattern, effects and modelling,” *Machine Learning Approaches for Urban Computing*, pp. 1–21, 2021, doi: 10.1007/978-981-16-0935-0_1.

[14] M. Yang, Y. Yuan, and G. Liu, “SDUNet: Road extraction via spatial enhanced and densely connected unet,” *Pattern Recognition*, vol. 126, p. 108549, 2022, doi: 10.1016/j.patcog.2022.108549.

[15] X. Lu, Y. Zhong, D. Chen, Y. Su, A. Ma and L. Zhang, “Cascaded Multi-Task Road Extraction Network for Road Surface, Centerline, and Edge Extraction,” in *IEEE Transactions on Geoscience and Remote Sensing,* vol.60, pp. 1-14, doi: 10.1109/TGRS.2022.3165817

[16] Z. Hong, D. Ming, K. Zhou, Y. Guo and T. Lu, "Road Extraction From a High Spatial Resolution Remote Sensing Image Based on Richer Convolutional Features," in IEEE Access, vol. 6, pp. 46988-47000, 2018, doi: 10.1109/ACCESS.2018.2867210.

[17] X. Li, Y. Wang, L. Zhang, S. Liu, J. Mei and Y. Li, "Topology-Enhanced Urban Road Extraction via a Geographic Feature-Enhanced Network," in IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 12, pp. 8819-8830, Dec. 2020, doi: 10.1109/TGRS.2020.2991006.

[18] K. Zhou, Y. Xie, Z. Gao, F. Miao, and L. Zhang, “FuNet: A novel road extraction network with fusion of location data and remote sensing imagery,” *ISPRS International Journal of Geo-Information*, vol. 10, no. 1, p. 39, 2021, 39, doi: 10.3390/ ijgi10010039.

[19] R. Kestur, S. Farooq, R. Abdal, E. Mehraj, O. Narasipura, and M. Mudigere, “UFCN: A fully convolutional neural network for road extraction in RGB imagery acquired by remote sensing from an unmanned aerial vehicle,” *Journal of Applied Remote Sensing*, vol. 12, no. 01, p. 1, 2018, doi: 10.1117/1.JRS.12.016020.

[20] X. Lu, Y. Zhong, Z. Zheng, and J. Wang, "Cross-domain road detection based on global-local adversarial learning framework from very high-resolution satellite imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 180, pp. 296–312, 2021, doi: 10.1016/j.isprsjprs.2021.08.018.

[21] Y. Li, S. Ouyang, and Y. Zhang, “Combining deep learning and ontology reasoning for Remote Sensing Image Semantic segmentation,” *Knowledge-Based Systems*, vol. 243, p. 108469, 2022, doi: 10.1016/j.knosys.2022.108469.

[22] J. Song, S. Gao, Y. Zhu, and C. Ma, “A survey of remote sensing image classification based on cnns,” *Big Earth Data*, vol. 3, no. 3, pp. 232–254, 2019, DOI: 10.1080/20964471.2019.1657720.

[23] L. Xie, X. Feng, C. Zhang, Y. Dong, J. Huang, and K. Liu, “Identification of urban functional areas based on the multimodal deep learning fusion of high-resolution remote sensing images and Social Perception Data,” *Buildings*, vol. 12, no. 5, p. 556, 2022, doi:10.3390/ buildings12050556.

[24] L. Wang, R. Li, C. Zhang, S. Fang, C. Duan, X. Meng, and P. M. Atkinson, "UNETFORMER: A unet-like Transformer for efficient semantic segmentation of remote sensing urban scene imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 190, pp. 196–214, 2022, doi: 10.1016/j.isprsjprs.2022.06.008.

[25] Y. Liu, Q. Ren, J. Geng, M. Ding, and J. Li, "Efficient patch-wise semantic segmentation for large-scale remote sensing images," *Sensors*, vol. 18, no. 10, p. 3232, 2018, doi: 10.3390/s18103232.

[26] A. Van Etten, D. Lindenbaum, and T.M. Bacastow, "SpaceNet: A remote sensing dataset and challenge series," 2018, doi: arXiv:1807.01232v3.

[27] B. Yang and W. Lee, "Human Body Odor Based Authentication Using Machine Learning," 2018 IEEE Symposium Series on Computational Intelligence (SSCI), 2018, pp. 1707-1714, doi: 10.1109/SSCI.2018.8628697.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Lecture Notes in Computer Science*, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4_28.

[29] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for Semantic Image segmentation," *Computer Vision – ECCV 2018*, pp. 833–851, 2018

[30] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," *2017 IEEE Visual Communications and Image Processing (VCIP)*, 2017, doi: arXiv:1707.03718v1.