

Road Segmentation in High-Resolution Images Using Deep Residual Networks

Dhanashri Patil

Department of Electronics and Telecommunication, D. Y. Patil College of Engineering, Akurdi, Pune, India and
Department of Electronics and Telecommunication, Army Institute of Technology, India
dpdhanashripatil@gmail.com

Sangeeta Jadhav

Department of Information Technology
Army Institute of Technology
Pune, India
djsangeeta@rediffmail.com

Received: 6 August 2022 | Revised: 12 September 2022 | Accepted: 20 September 2022

Abstract-Automatic road detection from remote sensing images is a vital application for traffic management, urban planning, and disaster management. The presence of occlusions like shadows of buildings, trees, and flyovers in high-resolution images and misclassifications in databases create obstacles in the road detection task. Therefore, an automatic road detection system is required to detect roads in the presence of occlusions. This paper presents a deep convolutional neural network to address the problem of road detection, consisting of an encoder-decoder architecture. The architecture contains a U-Net with residual blocks. U-Net allows the transfer of low-level features to the high-level, helping the network to learn low-level details. Residual blocks help maintain the network's training performance, which may deteriorate due to a deep network. The encoder and decoder structures generate a feature map and classify pixels into road and non-road classes, respectively. Experimentation was performed on the Massachusetts road dataset. The results showed that the proposed model gave better accuracy than current state-of-the-art methods.

Keywords-U-network; residual block; encoder; decoder

I. INTRODUCTION

Remote Sensing (RS) is one of the most essential applications. It is related to the analysis and identification of features by measuring the physical parameters of an object at some distance with the help of electromagnetic radiation. Target discrimination is possible using different characteristics of sensors such as temporal, spatial, and spectral [1]. Due to the improvements in RS technology, high-resolution images with adequate spectral and temporal information are available for different applications such as disaster management, water source, land utilization, marine, glacier monitoring, urban city planning, etc. These high-resolution images are considered a primary database for automatic road extraction methods.

Automatic road detection from remotely sensed images is an essential task for many applications such as urban planning, geographic information systems, map updates, and automatic vehicle navigation [2]. Various methods to address the problem of road detection have been presented, but it is still a challenging task due to the presence of occlusions, building shadows, and complex backgrounds in high-resolution images.

These noise elements result in a false and discontinuous segmentation of roads. Manual road detection is time-consuming, complex, and involves human interpretation. Artificial intelligence and deep learning have made automatic road detection more efficient and accurate than manual methods. Road detection methods are of two types: road detection and centerline extraction [3]. The main objective of a road detection system is to detect road pixels from RS images, while centerline extraction locates the skeleton of a road using mathematical morphological operations. The RS characteristics of the roads depend on weather conditions and sensors. Road features are classified into geometric, functional, topological, and contextual [4]. Geometric features are continuous and include the length-to-width ratio of roads, where the width is always narrower than the length. Functional features constitute connectivity in different areas. Characteristics of road intersections and continuity in length without interruption are called topological features. Contextual features include the shadows of trees, buildings, and flyovers. Some road detection methods used multiple features to improve the segmentation accuracy. But it is complex to identify all the features due to occluded high-resolution images.

Figure 1 shows a general block diagram of road extraction from remote sensing images. Several road datasets are available, including high-resolution satellite images with two classes: roads and non-roads [4]. The Massachusetts road dataset [5] has 1124 VHR images with 1500×1500 resolution. As a model's performance depends on the data quality, the preprocessing step ensures that the model receives correct data in a suitable format. The filtering technique helps to remove unwanted data from the image. Thresholding and binarization methods are used for grayscale conversion and normalization. The extraction of road features helps to produce different road aspects (geometric, radiometric, and contextual). In the classification technique, each pixel in an image is categorized in the road or non-road class. Various supervised or unsupervised methods are used for classification. Model accuracy can be evaluated using precision, recall, and IoU score. The segmented map represents the output of the model containing only the road pixels of the image.

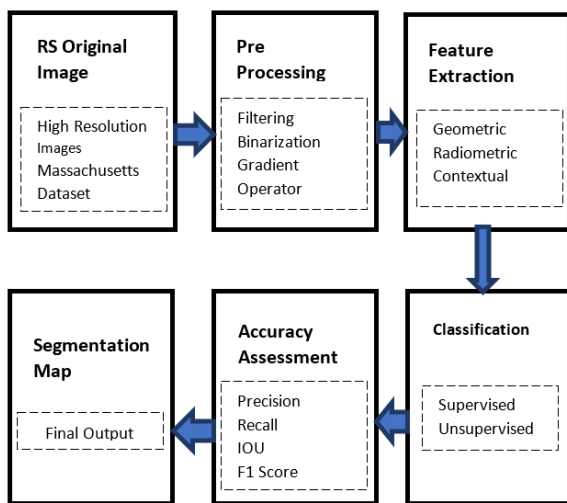


Fig. 1. Block diagram of road extraction from remote sensing images.

II. RELATED WORKS

Road detection from remote sensing images can be addressed using manual and automatic techniques. Currently, it is based on a manual process that is tedious and requires human intervention, while its accuracy depends upon the interpreter's skills. Automatic methods are more structured and economic. A neural network classifier approach was proposed in [6] to discriminate road and non-road pixels from IKONOS images. The accuracy of segmentation was improved using texture and spectral information. This method used a backpropagation neural network and a Gray Level Co-occurrence Matrix (GLCM) for the detection of road pixels. A backpropagation neural network with GLCM features for semantic segmentation was proposed in [7], where its functionality was optimized by incorporating homogeneity, contrast, and entropy.

Many computer vision applications use deep neural networks and mark state-of-the-art performance. Several studies applied deep neural networks to the semantic segmentation of roads, using different algorithms like U-Net, fully convolutional networks, CasNet, and deep Convolutional Neural Networks (CNN) for the semantic segmentation of remotely sensed images. A CNN approach to detect roads from noisy labeled samples was proposed in [8]. This algorithm generated the training labels and used texture descriptors to obtain road areas, using a combination of the supervised and unsupervised methods for post- and pre-training respectively. A novel cascade end-to-end CNN was presented in [3] to detect roads and their centerlines from high-resolution images. This study used two networks for the semantic segmentation of road pixels. The first network detected road pixels and generated a feature map. At the same time, the second network was cascaded with the first one to generate the centerline of the road area. This method marked the fastest computation time compared to existing networks. Deep residual U-nets were proposed in [9-10], using the U-net architecture [11] with skip connections and residual blocks. The skip connection helps to propagate the information from encoder to decoder and improves the segmentation result with fewer network parameters.

A linear integral convolutional algorithm was used in [12] to label road pixels using probability maps generated by CNN. Post-processing was performed using the linear convolutional algorithm to join the road labels, overcoming the limitation of spectral intensity variation in high-resolution images. CNN shows improved performance in semantic segmentation. An early fusion model using a VGG-16-based CNN was proposed in [13] for the classification and detection of 16 classes, showing better training and test validity. A fully CNN for automatic road extraction was presented in [14], using ResNet-34 on the encoder and basic U-Net on the decoder. The encoder was pre-trained on the ImageNet dataset and used a data augmentation technique to increase the dataset size. In [15], semantic segmentation was proposed using an eleven-layered CNN and images of the San Francisco bay area for road detection. The result was compared using average and max pooling and showed that CNN is a suitable method for remote sensing image analysis. A pre-trained CNN model was used in [16] for segmentation, proposing transfer learning using CNN. In [17], the conditional random field and landscape metrics were utilized with a deep CNN, where over-segmented roads were removed using landscape metric thresholding.

Unsupervised techniques such as the knowledge-based method, mean shift, graph theory, and mathematical morphology have been used to address the problem of semantic segmentation from remotely sensed images. In [18], semi-automatic road detection was presented using a mean shift. The seed points were prescribed in the images in advance by the user. Then, road and non-road segments were classified using the thresholding technique and the seed points were connected to generate a road network. Some studies used knowledge-based methods for road detection. The structural element can be evaluated using the energy function, which results in object detection, and the road tracker algorithm can be applied to gray-level images to find long ribbons, but this method is not suitable for occluded remote sensed images [19]. The toe-finding algorithm has been used to detect road footprints, eliminating non-road segments, but had over-segmentation limitations and was also sensitive to shadows and occlusions in an image [20].

In [21], morphological operations were used for the segmentation of the image and restored the road area by filtering out noise elements. This method resulted in discontinuous segmentation, and some road gaps existed in output. In [22], the multiscale retinex algorithm was used to enhance a high-resolution image, a canny operator was used for edge detection, and the skeleton of the road was detected using morphological operations. Mathematical morphology gives results for road segmentation, but accuracy depends on the size and shape of the road. There are no fixed road construction standards, as it varies from region to region. Therefore, integrating mathematical morphology with other techniques can help improve the overall result.

III. METHODOLOGY

This paper proposes a road segmentation architecture to extract roads from remotely sensed images. The main challenge was to identify design components that regulate the performance of the road segmentation architecture. This model

was established using U-Net and residual blocks [23]. The U-Net architecture consists of a contraction and an expansion path, which help to propagate the feature map in all possible ways from the encoder to the decoder. Therefore, the network accuracy can be improved using low-level details and high-level semantic information. The residual block introduces identity mapping in the network, facilitates smooth training, and helps to prevent the problem of vanishing gradients. Another task was to preprocess the high-resolution satellite images to produce the dataset to train and test the network.

A. Network Architecture

Road detection from remote sensing images is a semantic segmentation task, and U-Net is the best-suited architecture. The proposed model was inspired by the U-Net architecture. Figure 2 shows the detailed flow chart of the Deep Res-UNet. The architecture consists of three parts: the encoder, the decoder, and the bridge.

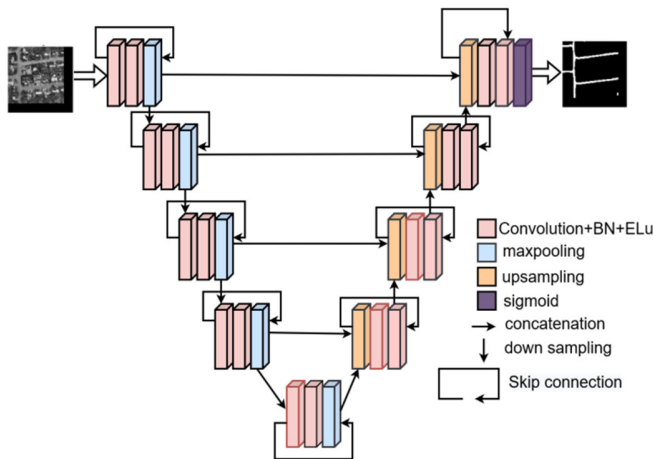


Fig. 2. Network architecture of the proposed model.

An input image sized 256×256 pixels is considered at the encoder's input. Four residual blocks with a 3×3 filter are implemented in the encoder part. The filter sizes of the residual blocks are respectively 16, 32, 64, and 128. An additional residual block with a filter size of 256 is connected between the encoder and the decoder, forming the bridge. The encoder helps to extract road features and generates feature maps, while the decoder classifies each pixel into a specific class. The encoder path uses the typical architecture of a convolutional network. It consists of repeated 3×3 convolutional layers followed by an Exponential Linear Unit (ELU) [24]. It also carries out a 2×2 max pooling operation with a stride of 2. The max pooling operation is required for the downsampling of the feature map and size reduction. Each residual block in the encoding part is connected to the decoding part using concatenation. The residual block consists of identity mapping, which represents connections between the input and output of the residual module. The proposed architecture replaces the Rectified Linear Unit (ReLU) with an Exponential Linear Unit (ELU) and adds the BN layer [25] after each convolutional layer. The ELU activation function is defined as:

$$f(x) = \begin{cases} x, & x > 0 \\ \alpha(e^x - 1), & x \leq 0 \end{cases} \quad (1)$$

The ELU and BN layers have become widely used components in CNN due to their high performance during network training. The number of kernels per output layer was set to 1. The sigmoid function, shown in (2), was used for classification instead of softmax:

$$\text{sigmoid} = \frac{1}{1+e^{-x}} \quad (2)$$

The classification cross-entropy was minimized using the Adam optimizer. Cross entropy is used to represent the difference between the distribution of input data and the distribution obtained from model training. Convolutional and upsampling operations were carried out in the decoder path. Each decoder module first performs the up convolutional operation with a feature map provided by the concatenation path, doubling the size of the feature map and improving the resolution. The same size of the input and output images can be ensured by upsampling. Then, 3×3 convolutional operation was carried out with the ELU activation function. The filter size of a convolutional layer in the decoder is 128, 64, 32, and 16 respectively. The last layer is 1×1 convolutional and is used to map 128 component feature vectors to their desired classes. This layer uses a sigmoid as an activation function. Table I shows the details of the network layers.

TABLE I. SIZE OF NETWORK LAYERS

Input	Unit Level	Conv. Layer	Filter	Stride	Output Size
Encoding	Level 1	Conv 1	$3 \times 3 / 16$	2	$256 \times 256 \times 3$
		Conv 2	$3 \times 3 / 16$	2	$256 \times 256 \times 16$
	Level 2	Conv 3	$3 \times 3 / 32$	2	$128 \times 128 \times 32$
		Conv 4	$3 \times 3 / 32$	2	$128 \times 128 \times 32$
	Level 3	Conv 5	$3 \times 3 / 64$	2	$64 \times 64 \times 64$
		Conv 6	$3 \times 3 / 64$	2	$64 \times 64 \times 64$
	Level 4	Conv 7	$3 \times 3 / 128$	2	$32 \times 32 \times 128$
		Conv 8	$3 \times 3 / 128$	2	$32 \times 32 \times 128$
Bridge	Level 5	Conv 9	$3 \times 3 / 256$	1	$16 \times 16 \times 256$
		Conv 10	$3 \times 3 / 256$	1	$16 \times 16 \times 256$
Decoder	Level 6	Conv 11	$3 \times 3 / 128$	2	$32 \times 32 \times 128$
		Conv 12	$3 \times 3 / 128$	2	$32 \times 32 \times 128$
	Level 7	Conv 13	$3 \times 3 / 64$	2	$64 \times 64 \times 64$
		Conv 14	$3 \times 3 / 64$	2	$64 \times 64 \times 64$
	Level 8	Conv 15	$3 \times 3 / 32$	2	$128 \times 128 \times 32$
		Conv 16	$3 \times 3 / 32$	2	$128 \times 128 \times 32$
	Level 9	Conv 17	$3 \times 3 / 16$	2	$256 \times 256 \times 16$
		Conv 18	$3 \times 3 / 16$	2	$256 \times 256 \times 16$
Output		Conv 19	1×1	1	$256 \times 256 \times 16$

IV. EXPERIMENT

A. Dataset and Preprocessing

The Massachusetts dataset was used, where each image contains $3 \times 1500 \times 1500$ pixels. The dataset consists of 1711 aerial images, and covers around 2600 km^2 . The dataset was divided into training, validation, and testing sets. The experimental procedure used 1191 images for training, 109 images for validation, and 273 images for testing. Figure 3 shows some samples of cropped satellite images with ground-truth labels. As the size of the original image is too large, it

would be resource intensive to feed it in the first layer, and the input image was cropped into 256×256 to avoid it. All the cropped images were then processed with the data augmentation technique to increase the number of images. This was carried out using random rotation and vertical and horizontal flips. The data augmentation technique helps to prevent the model from overfitting.

B. Implementation Details

The implementation was carried out on the TensorFlow platform. A deep learning framework requires lots of memory and high-performance machines. The proposed method was

implemented on an NVIDIA GPU with 24GB of memory. The Adam optimizer was used during training and the learning rate was set to 0.0001 to reduce the errors during training and improve performance. Semantic segmentation is a binary classification problem. The binary cross entropy was used as a loss function:

$$L(y, p(y)) = -\frac{1}{N} \sum_{i=0}^N (y * \log p(y)) + (1 - y) \times \log(1 - p(y)) \quad (3)$$

where $p(y)$ is the predicted value, y is the true label, and N is the number of samples.

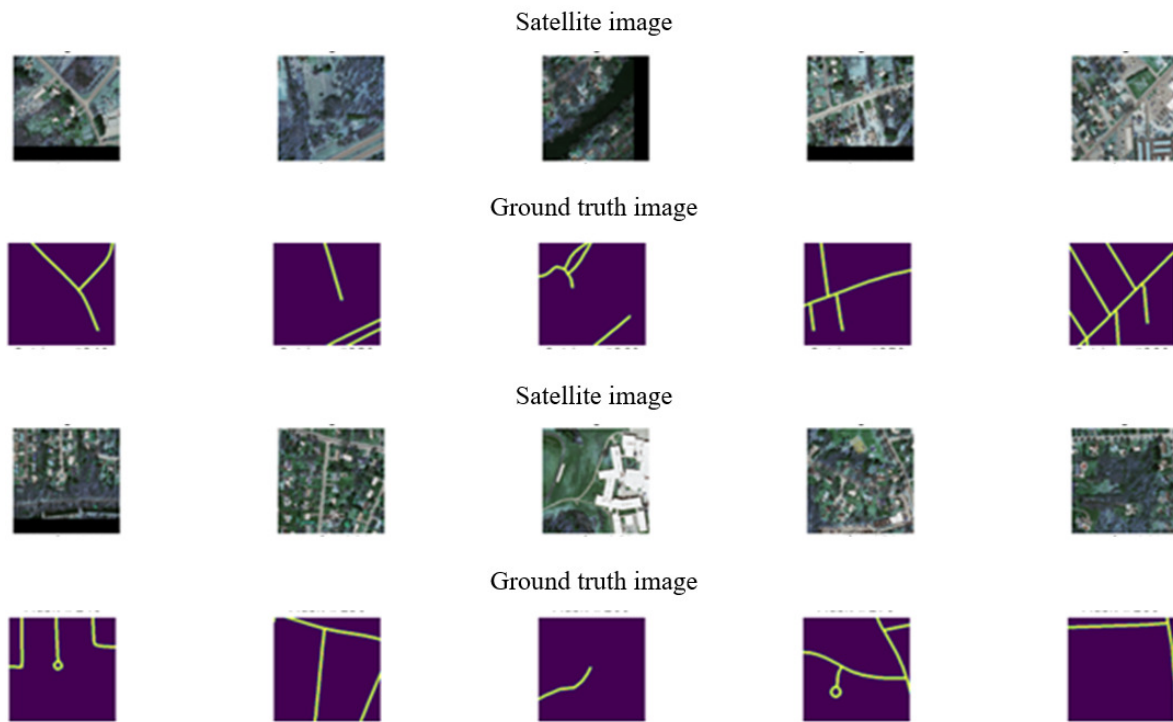


Fig. 3. Cropped satellite images with ground truth labels.

C. Evaluation Metrics

The qualitative performance of the model was assessed using the Intersection over Union (IoU), Completeness (COM), Correctness (COR), and F1 score. The IoU expresses the correctness of generated maps and is the ratio of intersection over the predicted map and ground-truth labels to a total area of the ground-truth labels and predicted map. The IoU can be calculated using the union operation. COM measures the percentage of areas matched on the reference map. COR measures the percentage of the matched road area in the segmentation map, and F1 defines the overall metric by combining COM and COR.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of union}} \quad (4)$$

$$\text{COM} = \frac{TP}{TP+FN} \quad (5)$$

$$\text{COR} = \frac{TP}{TP+FP} \quad (6)$$

The proposed model was trained from scratch without using any pre-trained modules. Table II shows the results of the proposed architecture and mentions the external accuracy results from [9]. This network uses skip connections and dense blocks. The results showed that the IoU score on the Massachusetts road dataset was 74.47%. Figure 4 shows the IoU and loss graphs of the proposed model. The COR metric was 96.42% and IoU was 81.57%. There were two problems associated with the ground-truth labels: sometimes, an area in an image was marked as a road but it was not. Some side or local roads were incorrectly labeled in the ground-truth image. The first problem increases the difficulty of extracting precise edges for the road networks. Figure 5 represents the flow chart of the proposed model.

TABLE II. RESULTS OF THE PROPOSED ARCHITECTURE

	Accuracy Metric			
	COR	COM	IoU	F1 score
Proposed model	96.42%	81.03%	81.57%	88.02%
External accuracy [9]	78.25%	70.41%	74.47%	74.07%

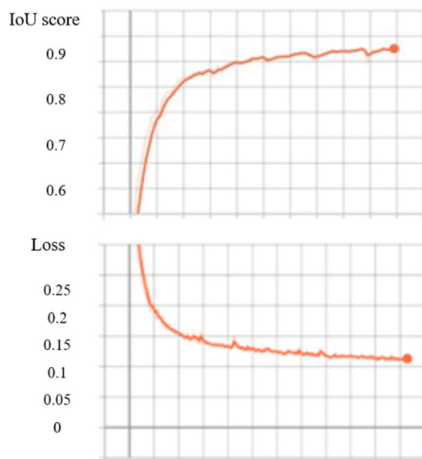


Fig. 4. IoU and loss metrics of the proposed model.

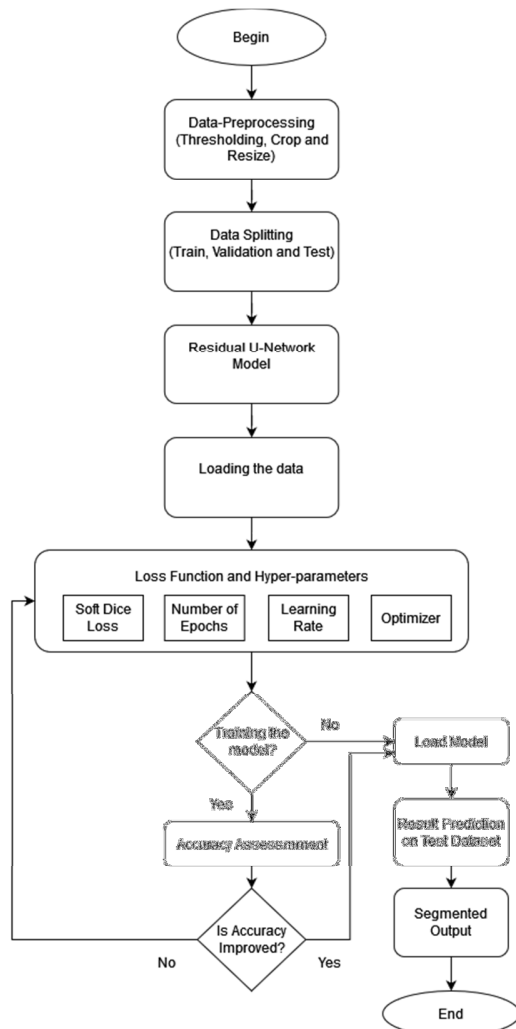


Fig. 5. Flow chart of the proposed model.

D. Comparison with Other Algorithms

The proposed architecture was compared with state-of-the-art models to test its validity. U-Net, FCN, and SegNet models

were considered and trained on the Massachusetts road dataset by keeping the same hyperparameters, as they are used extensively in semantic segmentation. The U-Net architecture was proposed for medical image segmentation, implementing a four-layer U-Net architecture using skipped connections without identity mapping [11]. A basic 7-layer FCN algorithm was implemented to assess accuracy, which was proposed for semantic segmentation in [26]. The SegNet model was introduced for multiscale segmentation [27]. The proposed model was used as is, following the encoder and decoder architecture with pixel-wise labeling and using pooling indices for the maximum pooling feature map at the decoder. Table III shows the qualitative performance of the models for the road extraction task using COR, COM, IoU, and F1 scores. The bold and underlined values indicate the best and second-best scores per metric respectively. The proposed model had 96.42% COR and 81.57% IoU values. The proposed model's IoU was about 5% higher than the U-Net's, showing that the more supervised information in the proposed model helps extracting roads more precisely. The COR metric of the proposed model is nearly 8% higher than that of SegNet.

TABLE III. COMPARISON OF RESULTS

Model	COR	COM	IoU	F1 Score	Detection time (s)
Proposed model	96.42%	81.03%	81.57%	88.02%	11.331s
Segnet [27]	88.25%	76.92%	71.03%	82.13%	14.528s
Unet [11]	93.31%	79.03%	77.49%	85.54%	11.432s
FCN [26]	83.13%	75.16%	68.66%	78.80%	9.144s

V. RESULTS AND DISCUSSION

Figures 6-8 show the IoU and loss graphs of the SegNet, U-Net, and FCN models, and Figure 9 shows a chart of their performance.

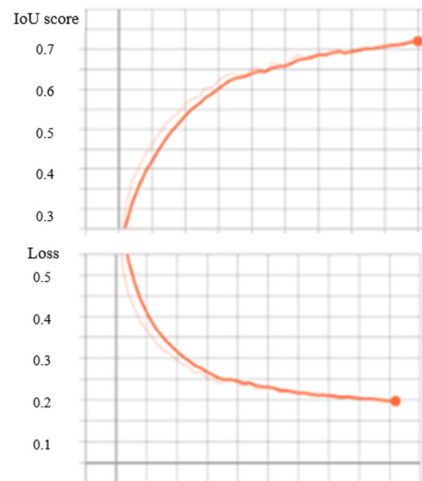


Fig. 6. IoU and loss of SegNet.

The proposed model used more skip connections in the encoder and decoder to improve the performance of U-Net and propagate the feature map in all possible ways. Identity mapping between the input and output of each residual block improves the network's generalization ability and enables the

identification of deep features. The SegNet model showed 71.03% IoU and 82.02% F1 scores. The U-Net achieved the second highest performance in IoU and COR and had an 85.54% F1 score. FCN had the lowest IoU and F1 scores but had the least detection time compared to the other three models. Figure 10 shows sample results for road detection with original, ground-truth, predicted, and threshold images.

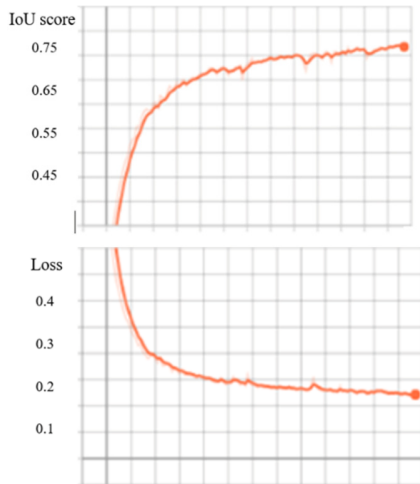


Fig. 7. IoU and loss of U-Net.

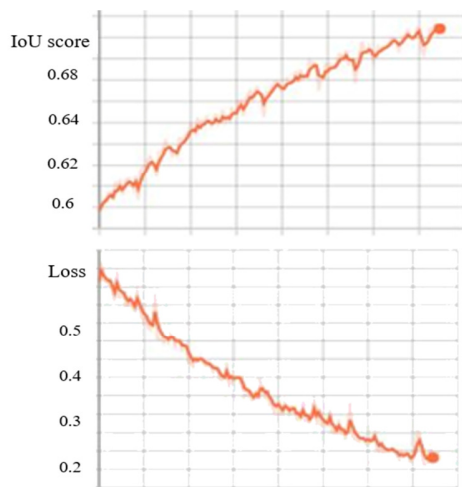


Fig. 8. IoU and loss of FCN.



Fig. 9. Chart of models performance.

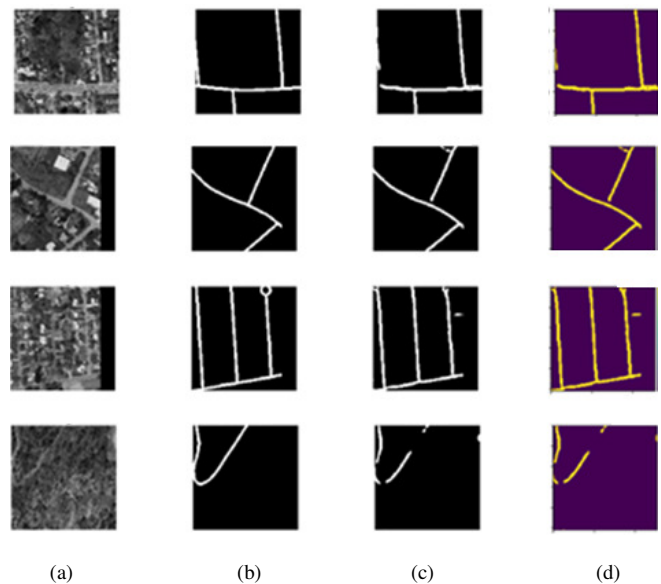


Fig. 10. Road detection results: (a) original, (b) ground-truth, (c) predicted, and (d) threshold images.

VI. CONCLUSION

This paper proposes a deep convolutional ResUnet architecture, using CNN to extract road areas from remotely sensed images. The deep convolutional encoder-decoder architecture serves as the foundation of the network. Data augmentation techniques were used to improve the performance of the network. The encoder helps to generate a feature map, and the decoder classifies each pixel into a particular class. A simulation was carried out on the Massachusetts road dataset. The proposed architecture was compared with other state-of-the-art models. The results showed that the proposed network outperforms SegNet, U-Net, and FCN. The proposed algorithm is robust against occlusions of shadows in the image and can be used to detect roads in forests and farms. It could also be helpful in land analysis and urban planning. The proposed model can be used for the automatic generation of road maps from different datasets. At first, some pre-processing techniques, like cropping and thresholding, need to be carried out on the input data set. These images can be applied as input to the proposed model to generate a road map. In the future, different optimization techniques and activation functions along with different datasets can be used to improve segmentation results.

REFERENCES

- [1] R. R. Navalgund, V. Jayaraman, and P. S. Roy, "Remote sensing applications: An overview," *Current Science*, vol. 93, no. 12, pp. 1747–1766, 2007.
- [2] A. A. Daptardar, "Introduction to Remote Sensors and Image Processing and its Applications," *International Journal of Latest Trends in Engineering and Technology*, Special Issue-IDEAS-2013, pp. 107–114, 2013.
- [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 Convolutional Neural Network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 6, pp. 3322–3337, Jun. 2017, <https://doi.org/10.1109/TGRS.2017.2669341>.

- [4] W. Wang, N. Yang, Y. Zhang, F. Wang, T. Cao, and P. Eklund, "A review of road extraction from remote sensing images," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 3, no. 3, pp. 271–282, Jun. 2016, <https://doi.org/10.1016/j.jtte.2016.05.005>.
- [5] V. Mnih, "Machine Learning for Aerial Image Labeling," Ph.D. dissertation, University of Toronto, Toronto, Canada, 2013.
- [6] A. Kirthika and A. Mookambiga, "Automated road network extraction using artificial neural network," in *2011 International Conference on Recent Trends in Information Technology (ICRTIT)*, Chennai, India, Jun. 2011, pp. 1061–1065, <https://doi.org/10.1109/ICRTIT.2011.5972323>.
- [7] A. N. Saeed, "A Machine Learning based Approach for Segmenting Retinal Nerve Images using Artificial Neural Networks," *Engineering, Technology & Applied Science Research*, vol. 10, no. 4, pp. 5986–5991, Aug. 2020, <https://doi.org/10.48084/etasr.3666>.
- [8] V. Mnih and G. E. Hinton, "Learning to Detect Roads in High-Resolution Aerial Images," in *Computer Vision – ECCV 2010*, Heraklion, Greece, 2010, pp. 210–223, https://doi.org/10.1007/978-3-642-15567-3_16.
- [9] J. Xin, X. Zhang, Z. Zhang, and W. Fang, "Road Extraction of High-Resolution Remote Sensing Images Derived from DenseUNet," *Remote Sensing*, vol. 11, no. 21, Jan. 2019, Art. no. 2499, <https://doi.org/10.3390/rs11212499>.
- [10] T. Alshaikhli, W. Liu, and Y. Maruyama, "Automated Method of Road Extraction from Aerial Images Using a Deep Convolutional Neural Network," *Applied Sciences*, vol. 9, no. 22, Jan. 2019, Art. no. 4825, <https://doi.org/10.3390/app9224825>.
- [11] 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*, Munich, Germany, 2015, pp. 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.
- [12] P. Li *et al.*, "Road network extraction via deep learning and line integral convolution," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Beijing, China, Jul. 2016, pp. 1599–1602, <https://doi.org/10.1109/IGARSS.2016.7729408>.
- [13] S. Nuanmeesri, "A Hybrid Deep Learning and Optimized Machine Learning Approach for Rose Leaf Disease Classification," *Engineering, Technology & Applied Science Research*, vol. 11, no. 5, pp. 7678–7683, Oct. 2021, <https://doi.org/10.48084/etasr.4455>.
- [14] A. V. Buslaev, S. S. Seferbekov, V. I. Iglovikov, and A. A. Shvets, "Fully Convolutional Network for Automatic Road Extraction from Satellite Imagery," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, UT, USA, 2018, pp. 4321–4324.
- [15] S.-H. Wang, J. Sun, P. Phillips, G. Zhao, and Y.-D. Zhang, "Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units," *Journal of Real-Time Image Processing*, vol. 15, no. 3, pp. 631–642, Jul. 2018, <https://doi.org/10.1007/s11554-017-0717-0>.
- [16] L. Loyani and D. Machuve, "A Deep Learning-based Mobile Application for Segmenting Tuta Absoluta's Damage on Tomato Plants," *Engineering, Technology & Applied Science Research*, vol. 11, no. 5, pp. 7730–7737, Oct. 2021, <https://doi.org/10.48084/etasr.4355>.
- [17] T. Panboonyuen, K. Jitkajornwanich, S. Lawawirojwong, P. Srestasathien, and P. Vateekul, "Road Segmentation of Remotely-Sensed Images Using Deep Convolutional Neural Networks with Landscape Metrics and Conditional Random Fields," *Remote Sensing*, vol. 9, no. 7, Jul. 2017, Art.no 680, <https://doi.org/10.3390/rs9070680>.
- [18] Z. Miao, B. Wang, W. Shi, and H. Zhang, "A Semi-Automatic Method for Road Centerline Extraction From VHR Images," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 11, pp. 1856–1860, Aug. 2014, <https://doi.org/10.1109/LGRS.2014.2312000>.
- [19] J. Shen, X. Lin, Y. Shi, and C. Wong, "Knowledge-Based Road Extraction from High Resolution Remotely Sensed Imagery," in *2008 Congress on Image and Signal Processing*, Sanya, China, Feb. 2008, vol. 4, pp. 608–612, <https://doi.org/10.1109/CISP.2008.519>.
- [20] J. Hu, A. Razdan, J. C. Femiani, M. Cui, and P. Wonka, "Road Network Extraction and Intersection Detection From Aerial Images by Tracking Road Footprints," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 4144–4157, Sep. 2007, <https://doi.org/10.1109/TGRS.2007.906107>.
- [21] M. M. Awad, "A Morphological Model for Extracting Road Networks from High-Resolution Satellite Images," *Journal of Engineering*, vol. 2013, Feb. 2013, Art. no e243021, <https://doi.org/10.1155/2013/243021>.
- [22] Z. Ma, J. T. Wu, Z. H. Luo, Z.H., "Road extraction from RS image based on level set method" presented at the 13th National Academic Conference on Image and Graphics, Nanjing, China, 2006.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, <https://doi.org/10.1109/CVPR.2016.90>.
- [24] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)." arXiv, Feb. 22, 2016, <https://doi.org/10.48550/arXiv.1511.07289>.
- [25] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *Proceedings of the 32nd International Conference on Machine Learning*, Lille, France, Jun. 2015, pp. 448–456.
- [26] 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, Apr. 2017, <https://doi.org/10.1109/TPAMI.2016.2572683>.
- [27] 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, Sep. 2017, <https://doi.org/10.1109/TPAMI.2016.2644615>.