



Road Network Detection from Aerial Imagery of Urban Areas Using Deep ResUNet in Combination with the B-snake Algorithm

Hafiz Suliman Munawar¹ · Ahmed W. A. Hammad¹ · S. Travis Waller² · Danish Shahzad³ · Md. Rafiqul Islam⁴

Received: 2 October 2022 / Accepted: 11 January 2023 / Published online: 25 January 2023
© The Author(s) 2023, corrected publication 2023

Abstract

Road network detection is critical to enhance disaster response and detecting a safe evacuation route. Due to expanding computational capacity, road extraction from aerial imagery has been investigated extensively in the literature, specifically in the last decade. Previous studies have mainly proposed methods based on pixel classification or image segmentation as road/non-road images, such as thresholding, edge-based segmentation, k-means clustering, histogram-based segmentation, etc. However, these methods have limitations of over-segmentation, sensitivity to noise, and distortion in images. This study considers the case study of Hawkesbury Nepean valley, NSW, Australia, which is prone to flood and has been selected for road network extraction. For road area extraction, the application of semantic segmentation along with residual learning and U-Net is suggested. Public road datasets were used for training and testing purposes. The study suggested a framework to train and test datasets with the application of the deep ResUNet architecture. Based on maximal similarity, the regions were merged, and the road network was extracted with the B-snake algorithm application. The proposed framework (baseline + region merging + B-snake) improved performance when evaluated on the synthetically modified dataset. It was evident that in comparison with the baseline, region merging and addition of the B-snake algorithm improved significantly, achieving a value of 0.92 for precision and 0.897 for recall.

Keywords Deep ResUNet · Region merging · Object-based image analysis (OBIA) · B-snake · Road network extraction

1 Introduction

Recent advancements in aerial imagery have allowed the provision of high-resolution images that can distinguish roads. Road network extraction from aerial images has been applied for transportation management, road navigation, updating geographic information and urban planning. Road extraction from aerial imagery has been carried out using

different methods, known as road area extraction/detection [2, 10, 11, 22]. Image segmentation and pixel classification have been the widely used methods for sorting road or non-road images. For instance, a shape index feature, support vector machine (SVM), angular texture feature, and a fuzzy classifier have been proposed for road area extraction [3, 15, 29]. A framework based on SVM facilitated road feature extraction from multi-spectral images [8]. Similarly, Yuan et al. [32] proposed a multi-stage road extraction method involving road grouping, segmentation, and medial axis point selection. Hierarchical graph-based image segmentation has also been proposed for unsupervised extraction [21]. Furthermore, the conditional random field (CRF) model has also been implemented for the said purpose [17]. However, the destruction caused by cars, trees, or surrounding features results in poor accuracy of these methods [21].

1.1 Related work

Modern image segmentation and classification techniques are powered by deep learning technology. Deep learning

✉ Hafiz Suliman Munawar
h.munawar@unsw.edu.au

¹ School of the Built Environment, University of New South Wales, Sydney, NSW 2052, Australia

² Lighthouse Professor and Chair of Transport Modelling and Simulation “Friedrich List” Faculty of Transport and Traffic Sciences, Technische Universität Dresden, Dresden, Germany

³ Department of Visual Computing, University of Saarland, 66123 Saarbrücken, Germany

⁴ Data Science Institute (DSI), University of Technology Sydney (UTS), Sydney, Australia

methods have progressed immensely in recent times. They have been utilised for interpreting remote sensing data, computer vision and solving other complex problems with higher performance and achieving better results [24, 26, 27]. In deep learning, applying multilayered models allows the processing of different levels of visual information on each layer. Local features are processed by lower layers, while higher layers assist in inferring more complex features. Road network extraction has been improved using deep learning methods. The First attempt to detect roads by deep learning methods was proposed by Mnih and Hinton [19] utilising restricted Boltzmann machines (RBMs). Other studies have also suggested better outcomes with the application of deep architecture. However, it is challenging to train due to the vanishing gradient. To facilitate training, a deep residual learning framework was suggested based on identity mapping to overcome training issues [35].

Similarly, to enhance segmentation accuracy, U-Net was proposed to concentrate map features instead of fully convolutional Networks with skip connection [1]. U-Net architecture is for semantic segmentation consisting of the expansive and contracting path. Zhang et al. [34] proposed a deep residual U-Net combining the strengths of deep learning methods and U-Net, built on the residual unit instead of basic neural units and removed the cropping operation from the network.

Image segmentation by analysing images at the object level instead of working at the pixel level is a well-adapted approach for high-resolution images that is robust and less noisy [6]. Therefore, the "Object-based Image Analysis (OBIA) approach improves the quality of segmentation results [13, 16, 18].

Road boundary detection techniques utilise lane patterns (features) and road models. These techniques should be capable of maintaining the quality of road detection without being affected by the shadows, processing painted and unpainted roads, detecting the curved road, and detecting both sides of the lane markings utilising parallel constraints [24, 26, 27]. Wang et al. [25] addressed these constraints by proposing a novel B-snake algorithm. This algorithm can define a wide range of lane structures rather than only straight and parabolic models. It utilises parallel knowledge of roads and is robust against external factors like noise, shadow, and missing and incorrect markings. B-snake

exhibits local control and forms arbitrary shapes which assist in describing a different range of road shapes while retaining compact representation. For instance, by increasing the control points, more complex shapes of roads with corner turns can be explained by the B-snake algorithm [28].

This study proposes a deep residual dense U-Net method along with (1) region merging (merging the regions formed by segmentation) and (2) a B-snake algorithm for road detection. The regions which were road-like were assembled for the study. The merging criterion in the region merging algorithm defines the cost of merging two regions which should be considered. The proposed framework for road network extraction is shown in Fig. 1. Moreover, the study utilises a boundary loss function in combination with BCE-dice loss [binary cross entropy criteria (BCE) and dice loss] for segmentation to merge pixels along the road network and cancel pixels that were across the road network.

The study considers the case study of the Hawkesbury Nepean Valley, located northwest of Sydney, New South Wales (NSW), Australia for road detection during disaster scenarios. The organisation of the paper is as follows. Section 2 defines the deep residual dense U-Net architecture and the lane boundaries modelling by the B-snake model. Section 3 describes the results of the pre-processing, training of the data sets, and road extraction. Section 4 summarises the outcome of the proposed framework, which depicted better results on synthetically modified data sets.

2 Methodology

This study proposes a Deep ResUnet architecture for training and testing on the datasets. The regions were merged based on maximal similarity, followed by road network extraction through the B-snake model. The system used during this project had the following specifications: 12th Gen Intel® Core™ i7-12700H (24 MB cache, 14 cores, 20 threads, up to 4.70 GHz Turbo).

2.1 U-Net

U-net is a convolution neural network consisting of max pooling, ReLU activation, concatenation, and convolution operations [33]. Collecting finer details while obtaining

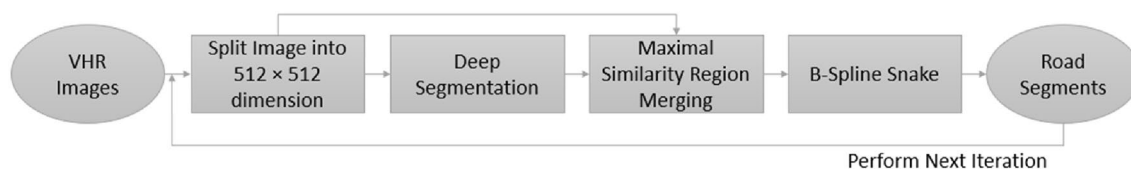


Fig. 1 Proposed framework for road network extraction

high-precision results in semantic segmentation while keeping semantic knowledge is vital. It is difficult to train a deep neural network with limited training datasets. It can be overcome by applying a pre-trained network to the desired datasets. The extensive data augmentation in U-Net is another way to overcome the training issues. Its key contribution is the creation of shortcut connections and is found to be useful for tasks where the output and input are of similar size, and the output requires spatial resolution. The U-net efficiently creates segmentation masks. Replacing the basic unit with the residual unit significantly enhances the performance of U-Net.

2.1.1 Residual Unit

The residual neural network is composed of units stacked in a sequence that assists in the training of the U-Net model and overcomes the degradation issues [9]. There is a stack of residual units in between. The residual unit is composed of ReLU activation, convolutional layers, batch normalisation, no pooling layer and using 3 × 3 kernels and preserving spatial dimensions; these combinations impact the processing of the data. The residual unit is given below:

$$y_m = h(x_m) + F(x_m, W_m)$$

$$x_{m+1} = f(y_m)$$

where x_m , and x_{m+1} are the input and output of the m th residual unit, the residual function $F(\cdot)$, activation and identity mapping function $f(y_m)$ and $h(x_m)$, respectively, for a characteristic one residual unit is given as $h(x_m) = x_m$.

2.1.2 Deep ResUnet

The combination of U-net and residual neural networks has many benefits. It provides ease of training of the network; the skip connection enhances the information propagation and minimises degradation [19] (Fig. 2). It enables designing a neural network with fewer parameters and enhanced performance. For road area extraction, a Deep ResUnet 7-level architecture has been proposed [31]. The Deep ResUnet network comprises encoding, decoding, and bridge. The input image is encoded into a compact representation converted into a pixel-wise image. The bridge connects encoding and decoding. The three components are residual units consisting of identity mapping (input and output units) and convolution blocks (consisting of ReLU activation, convolutional, and a BN layer).

Encoding and decoding path consists of 3 residual units (Fig. 2). For encoding path, instead of using the pooling

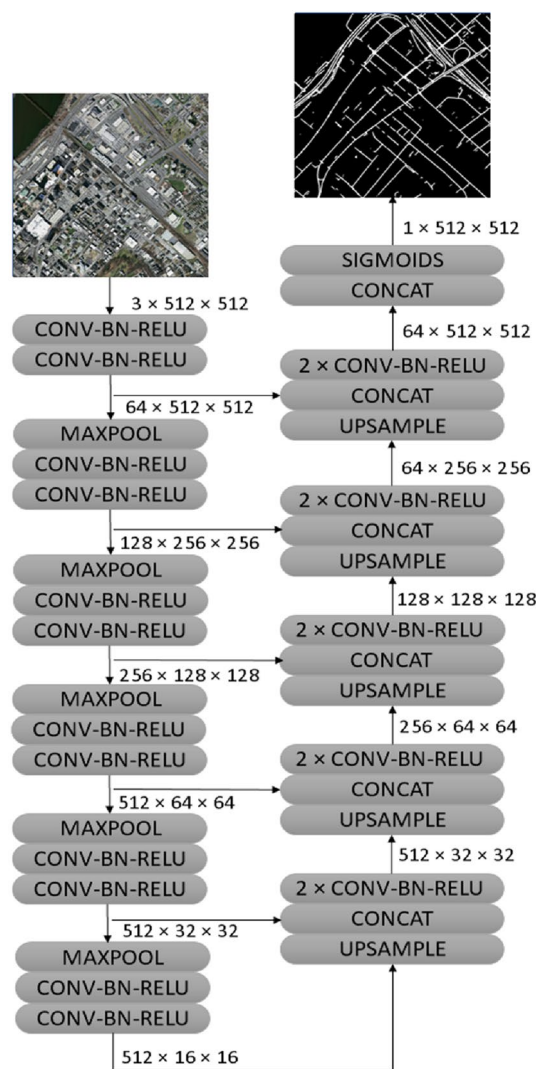


Fig. 2 The proposed deep Residual UNET architecture

operation, each unit is applied with a stride of two to the first convolution block. This reduces the feature map to its half size for multiscale learning. Stride alters the volume of movement over the image and compresses it. The encoded output volume is affected by the size of the filter. Before each unit chain of features, maps are up sampled from the corresponding encoding path for decoding. The multi-channel feature maps are converted into desired segmentation through a 1 × 1 convolution and a sigmoid activation layer after the last level of the decoding path (Fig. 2) [30]. The deconvolutional layers are utilised by the decoder to increase feature map size to the dimensions of the input image [12].

2.1.3 Loss Function

Boundary loss for road boundaries (highly unbalanced segmentation) is being used. The loss function aims to get smoother outputs at the boundaries and enhance model output for two close parallel roads. To resolve the issue of highly unbalanced segmentation, a distance metric on the space of contours is formed. The boundary loss function was combined with BCE-Dice Loss [7].

2.2 Region Merging

Region merging can be defined as the assembly of the raw regions produced by segmentation [14]. The grouping of similar regions is given as follows:

$$G = \{G_i, i = 1, 2, \dots, P\}, \quad P < O,$$

where the region after grouping relates to, P and before grouping the number of all segments is given as O .

The region merging algorithms are classified into

- Non-purposive grouping (NPG).
- Purposive grouping (PG).

Non-purposive grouping involves merging small regions into larger regions based on efficient segmentation. It merges with regions based on related characteristics such as pixel segmentation and marker refinement. It also merges regions relating to similar objects based on expected connections of joints between parts of the same object. On the other hand, PG is based on the distinct properties of the objects. Maximal similarity based region merging (MSRM) was introduced as a region-merging approach by Ning et al. [20]. When the similarity rate is ascertained, an approach for locating image objects for merging is necessitated. Various heuristics can be applied to merging arbitrary object A with adjacent object B. Four strategies were proposed by Baatz [4]. These are (1) fitting, (2) best fitting, (3) local mutual best fitting and (4) global mutual best fitting. The roads appear as connected road segments in remote sensing images. The application of MSRM will assemble road segments and distinguish them from the rest [20]. The similarity between the arbitrary objects C and D is given as

$$Sim(C, D) = \sum_1^P \sqrt{NH_c^i \cdot HD},$$

where NH_c and NHD give the normalised histograms of C and D, the quantity of bins for each colour channel is given by b , $P=b^3$ and the element of histogram is given by i superscript. The similarity measure is given as:

$$Sim(C, D) = \cos^{-1}(\overline{NH_C} \cdot \overline{NH_D}).$$

MSRM belongs to the second category i.e., best fitting and the merging strategy implies that two arbitrary regions C and D can only merge when the following condition is applied:

$$Sim()^{on, D=I_{max}}(Sim(C, N_c^i)),$$

where N_c gives C's adjacent regions.

2.3 B-spline Snake

The B-spline snake algorithm is efficient for rapid and spontaneous contour outlining. The application of the snake algorithm is varied and has been used for segmentation, edge detection, shape modelling, and tracking motion. The active contours or snakes move under the impact of forces (both internal and external) from the curve and image data, respectively [23]. Cubic B-spline with fewer state variables provides more economical recognition of snake and are piecewise polynomial functions. They give local proximation to contours with limited control points or parameters. Four or more control points can represent the curves. With the addition of more control points, the flexibility of the curves enhances, which either permits variation in the curve or reduces continuity at certain points when multiple knots are utilised [5].

The segmented image calculates B-Spline by defining the control points after every connected ($n=64$) pixel.

A cubic B-spline can be specified by $m+1$ control point Q_0, Q_1, \dots, Q_m and comprises m^2 cubic polynomial curve segments, where each segment of the B-Spline is derived from its four neighbouring control points. The knots in B-spline curve are the joints between the two segments of a curve. The equation for each curve segment is:

$$Ui(s) = \left(\frac{1}{6}\right) [s^3 s^2 s^1] \begin{vmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 3 \\ 1 & 4 & 1 & 0 \end{vmatrix} \begin{vmatrix} Q[i-3] \\ Q[i-2] \\ Q[i-1] \\ Q[i] \end{vmatrix},$$

where “ s ” is the curve segment with a value 0–1 and “ i ” corresponds to curve segments. Applying B-splines as active contours is effective as they are continuous at each point and knot and smooth out the extracted features in the images. The number of control points controls the splines' flexibility or curvature. They also exhibit local control since changing a single control point will only change a small contour section. The pseudo-code for B-snake is given as below:

1. Get output/segmentation mask using the proposed deep residual UNET architecture.
2. Apply region merging (Sect. 2.2), whereas maximal similarity-based region merging (MSRM) involves merging small regions into larger regions based on efficient segmentation.
3. The segmented image calculates B-spline by defining the control points after every connected ($n = 64$) pixel.
4. Perform minimisation using the non-maximal suppression on control points to calculate optimised B-spline segments.
 - a. For the 4 splines the control point was recalculated to check if the control point needs to be moved.
 - b. The distances (in the expected direction to the spline) of the sample points along the 4 splines to the closest edge were evaluated
 - c. The control point was moved to the neighbourhood point which had the smallest sum of distances.

2.4 Minimisation Algorithm

The minimisation algorithm detects the minimum of the objective function in the n -dimensional parameter space. Using the non-maximal suppression, control points were generated to calculate the B-spline segments. Control points were generated on the segmented image using non-maximal suppression as follows:

- a. The maximum distance between the peaks ($n = 64$) was defined.
- b. For every row in the image, perform a sliding window operation and, on each step, all non-maximum values were inverted to a fixed negative number.
- c. The same operation (b) was used to handle the non-maximum values per column.
- d. The pixels with a negative value to were set to zero.

From the control points, the initial B-spline segments were calculated. Sample $k = 20$ points were taken along each spline segment. The sample points' distances (in the expected direction to the spline) were calculated along the 4 splines to the closest edge. The above steps were repeated till less than $k\%$ of control points were moved ($k = 60$). Cycle through each control point to find the contribution to 4 spline segments. For each pixel in a neighbourhood surrounding the current control point following steps were followed:

3 Experiments and Results

The Hawkesbury Nepean Valley region, NSW, Australia was selected for road extraction. For this, the Massachusetts roads dataset's online data source was used (Table 1). The training datasets (Fig. 3) contained 1105 images with a corresponding labelled mask. While in the test dataset, there were 13 images with 13 corresponding labelled masks.

4 Data Collection and Pre-processing

This study selected Hawkesbury Nepean valley, NSW, Australia to detect road networks because this region is prone to floods each year. With road network detection, disaster response can be enhanced and a safe route for evacuation could be selected. Additionally, Massachusetts road datasets were used.

During pre-processing, the training dataset contained 1105 images of size (1500*1500), but we had the corresponding labelled mask for only 804 (73%) images. So only images having corresponding masks were utilised for training (Fig. 4). The Table 1. Describes the statistical overview of Massachusetts roads dataset.

Out of 804 images with masks, there are images with white patches in them but had labelled data for those white patch regions. Such images diminish the model performance and therefore were not used during training. Each of the remaining images and masks was then resized to (1536*1536) and then broken into nine images of size (512*512). The benefits of splitting images were a more extensive training dataset and more options for

Table 1 Describe the statistical overview of the Massachusetts Roads Dataset

	Height	Width	Area	road_pixels	bg_pixels	road_pixels_per	bg_pixels_per
Count	804.0	804.0	804.0	804.000000	8.040000e+02	804.000000	804.000000
Mean	1500.0	1500.0	2,250,000.0	117,340.655473	2.132659e+06	5.215140	94.784860
Std	0.0	0.0	0.0	75,938.547720	7.593855e+04	3.375047	3.375047
Min	1500.0	1500.0	2,250,000.0	2913.000000	1.807388e+06	0.129467	80.328356
25%	1500.0	1500.0	2,250,000.0	55,667.750000	2.082287e+06	2.474122	92.546100
50%	1500.0	1500.0	2,250,000.0	101,626.500000	2.148374e+06	4.516733	95.483267
75%	1500.0	1500.0	2,250,000.0	167,712.750000	2.194332e+06	7.453900	97.525878
Max	1500.0	1500.0	2,250,000.0	442,612.000000	2.247087e+06	19.671644	99.870533

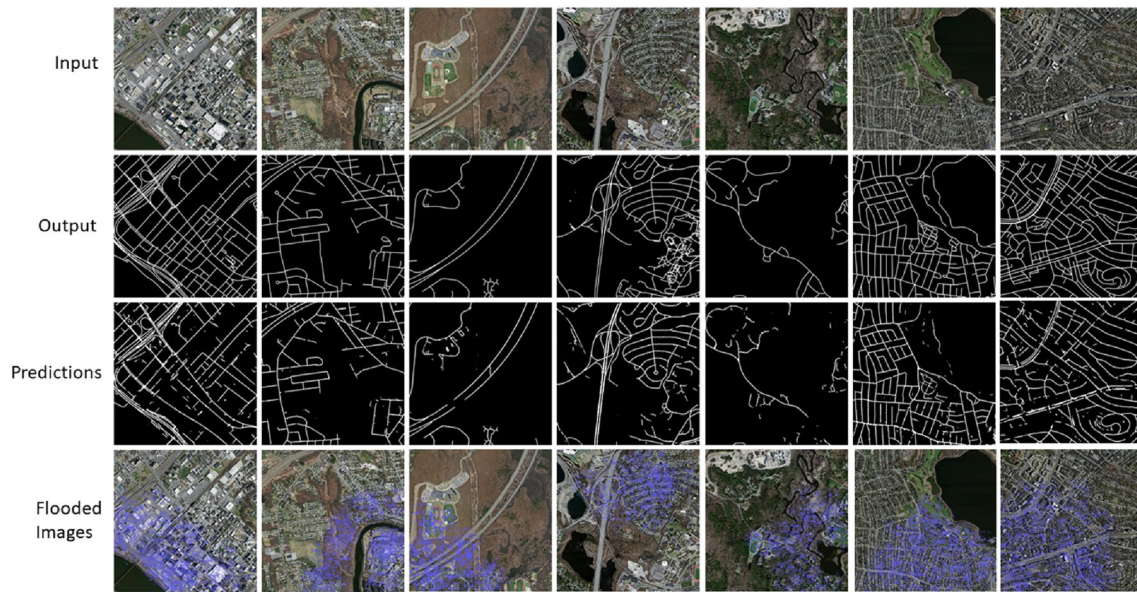


Fig. 3 Synthetically added flood region using Massachusetts Roads Dataset

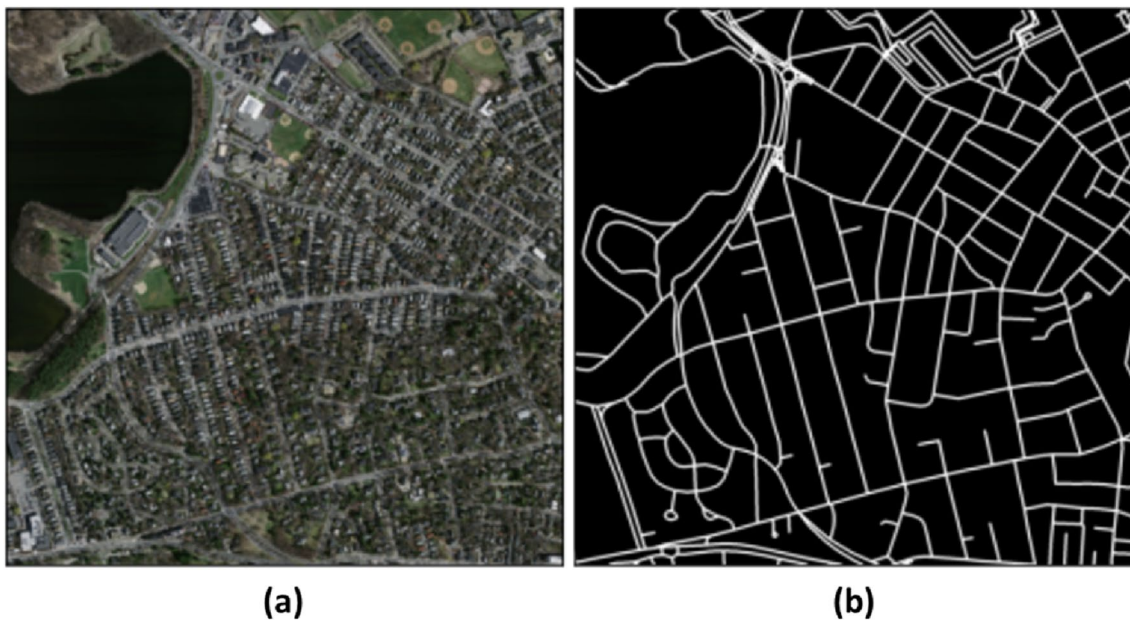


Fig. 4 Massachusetts roads dataset, image and its segmented mask used for training

augmentation. Each of the nine images can have different augmentation at run time, reducing the chance of overfitting. Also, it resulted in a bigger batch size as more images of smaller size can be loaded into limited GPU memory compared to larger images. A few more random crops of size (512*512) from size (1536*1536) images were also taken to increase the dataset. To avoid data duplication, the random crops do not overlap with nine cropped images. These images were randomly rotated by either 90 degrees or 270

degrees. After pre-processing, a total of 7240 images were obtained for training. All 13 images from the testing set were correct and used directly during model performance evaluation.

4.1 Training

The training set was divided into an 85:15 ratio to obtain 6150 training images and 1090 validation images.

Tensorflow v2 and TensorFlow Keras were used to build the UNET model [33]. Around 15–20% image synthesis was achieved.

4.2 Augmentation

A Runtime augmentation was performed on the training dataset to increase dataset variety with a combination of horizontal and vertical flips having a probability of 0.5. Brightness augmentation was done to improve the model deal in low-light situations. Tensorflow dataset API is used to pre-process data before feeding it into the model.

4.3 Model architecture

The model uses U-net architecture to segment small objects from large images. This capability makes U-net an excellent candidate for satellite imagery segmentation problems [36]. The benefit of using this model for road extraction is that the residual units ease the training of deep networks. The connections within the network ease the propagation

of information without degradation, thus allowing the designing of a network with few parameters with better performance.

4.4 Training schedule

At the outset, the model was trained for the first ten epochs with a combination of boundary loss and BCE-dice loss, as shown in Fig. 5. Later the model was only optimised using BCE-dice loss for image mask prediction, as shown in Fig. 6.

Learning Rate Decay is an advanced technique to optimise and generalise Deep Neural Networks (DNN), and its methods are widely applied. In our approach, we observed a decay of 20% in the learning rate after a cycle size of 5 epochs, as shown in Fig. 5. Whilst training, after every batch update, the cyclical learning rate decay slowly increases the learning rate.

Following graphs in Fig. 5 show training progress:

- Redline implies validation data.
- Orange line implies training data.

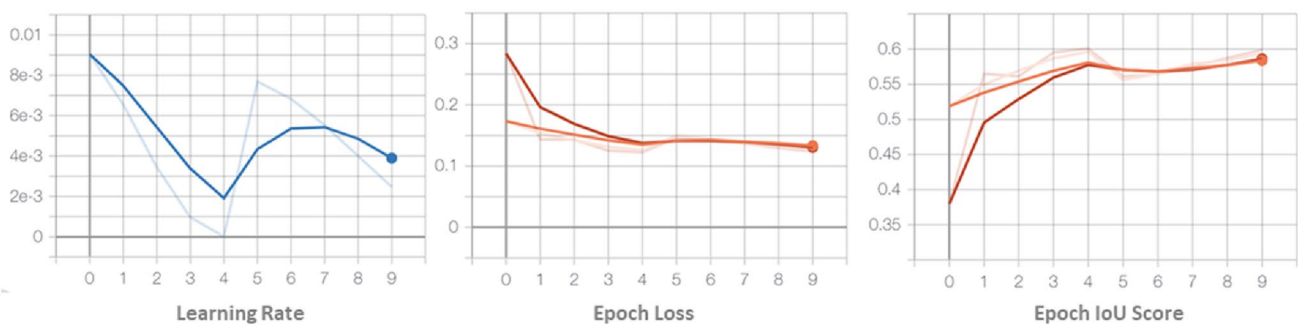


Fig. 5 a–c Initial training for ten epochs with both boundary loss and BCE-dice loss



Fig. 6 Image mask prediction

Table 2 (a–c): Proposed methods evaluated on (a) without synthetically modified dataset (b) with synthetically modified datasets, (c) Performance of proposed methods

(a)									
	Height	Width	Area	Road_pixels	Water_pixels	Bg_pixels	Road_pixels_per	Water_pixels_per	Bg_pixels_per
Count	793.0	793.0	793.0	793.0	793.0	793.0	793.0	793.0	793.0
Mean	1500.0	1500.0	2,250,000.0	117,524.1	0.0	2,132,475.9	5.2	0.0	94.8
Std	0.0	0.0	0.0	76,277.6	0.0	76,277.6	3.4	0.0	3.4
Min	1500.0	1500.0	2,250,000.0	2913.0	0.0	1,807,388.0	0.1	0.0	80.3
25%	1500.0	1500.0	2,250,000.0	55,487.0	0.0	2,080,765.0	2.5	0.0	92.5
50%	1500.0	1500.0	2,250,000.0	101,582.0	0.0	2,148,418.0	4.5	0.0	95.5
75%	1500.0	1500.0	2,250,000.0	169,235.0	0.0	2,194,513.0	7.5	0.0	97.5
max	1500.0	1500.0	2,250,000.0	442,612.0	0.0	2,247,087.0	19.7	0.0	99.9
(b)									
	Height	Width	Area	Road_pixels	Water_pixels	Bg_pixels	Road_pixels_per	Water_pixels_per	Bg_pixels_per
Count	817.0	817.0	817.0	817.0	817.0	817.0	817.0	817.0	817.0
Mean	1500.0	1500.0	2,250,000.0	115,577.2	46,763.1	2,087,659.7	5.1	2.1	92.8
Std	0.0	0.0	0.0	74,616.5	99,861.0	75,293.7	3.3	0.4	3.3
Min	1500.0	1500.0	2,250,000.0	2844.8	3,156.2	1,779,748.6	0.1	1.3	79.1
25%	1500.0	1500.0	2,250,000.0	54,566.7	38,302.9	2,039,278.2	2.4	1.7	90.6
50%	1500.0	1500.0	2,250,000.0	100,285.0	46,863.0	2,103,427.7	4.5	2.1	93.5
75%	1500.0	1500.0	2,250,000.0	165,262.6	55,033.2	2,148,424.4	7.3	2.4	95.5
max	1500.0	1500.0	2,250,000.0	432,499.9	64,554.8	2,208,715.8	19.2	2.9	98.2
(c)									
Method	Mean_iou	Mean_precision	Mean_recall	Mean_F1					
Epoch—10	0.6041	0.7204	0.7907	0.7521					
Epoch—30 (last)	0.6067	0.7275	0.7862	0.7534					
Epoch—24 (baseline)	0.6171	0.7315	0.7975	0.7612					
Region merging	0.8334	0.863	0.875	0.867					
Region merging + B-snake	0.894	0.92	0.897	0.908					

As seen in Table 2a–c, performance was significantly lost when the proposed methods were evaluated on an unseen dataset. It is due to different abilities to generalise knowledge between seemingly identical tasks, as the area on the image was synthetically modified for a flood. However, the proposed framework (baseline + region merging + B-snake) achieved better performance when evaluated on a synthetically modified dataset. It is evident that in comparison with baseline region merging and the addition of B-snake, significant improvement was achieved through the proposed framework with a value of 0.92 for precision and 0.897 for recall. A Tensor board visualisation example for validation samples is shown in Fig. 7.

4.5 Inference

For inference on test images, each image was divided into (512*512), like the training pre-processing and the model prediction is then stitched together to produce a predicted

mask of size (1500*1500). To get a binary image from the prediction output, a thresholding of 0.5 was applied on each mask. Any pixel with a value above 0.5 was a positive road pixel. Small blobs (white patches) of false positives were removed.

5 Conclusion

Thus, a framework was suggested to enhance road network extraction. The framework was based on deep residual dense U-Net, region merging based on similarity and a B-snake algorithm. The study utilised a boundary loss function in combination with BCE-Dice loss for segmentation to merge the pixels along the road and cancel the pixels across the road network. A case study of Hawkesbury Nepean valley was considered for road network extraction. The Massachusetts roads dataset was used for training and testing the data. In the training datasets, there were 1105



Fig. 7 Comparison of B-splines (B-snake) with a simple snake

images and 804 with a corresponding labelled mask, while in the test dataset, there were 13 images with 13 corresponding labelled masks. Only images having corresponding masks were utilised for training. Tensorflow v2 and TensorFlow Keras were used to build the UNET model. Around 15–20% of image synthesis was achieved for the study. It was observed that network evaluation on unseen datasets experienced a loss in performance. The reason was due to varying abilities to gather information from similar tasks slightly modified for floods. However, the proposed framework depicted better results on synthetically modified data sets. A precision of 0.92 and recall of 0.897 was achieved, respectively. Implementation of the boundary loss function in combination with BCE-Dice loss for segmentation was selected as a learning strategy for the study; however, if higher weightage is applied to the proposed method, the non-road regions also start to merge the pixel resulting in poor segmentation.

Acknowledgements The authors would like to thank CDRI and Natural Hazards Research Australia for their support in conducting this research.

Author contributions Methodology, HSM and AWAH; investigation, AWAH and STW; writing—original draft preparation, HSM, DS, RI and AWAH; writing—review and editing, AWAH, MRI, and STW; supervision AWAH and STW. All authors have read and agreed to the published version of the manuscript.

Funding This research received no external funding.

Data availability statement Codes are available and will be provided upon reasonable request to the corresponding author.

Declarations

Conflict of interest The authors declare no conflict of interest.

Institutional review board statement Not applicable.

Informed consent Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Abdollahi A, Bakhtiari HRR, Nejad MP. Investigation of SVM and level set interactive methods for road extraction from google earth images. *J Indian Soc Remote Sens.* 2018;46(3):423–30. <https://doi.org/10.1007/s12524-017-0702-x>.
2. Abdollahi A, Pradhan B. Integrated technique of segmentation and classification methods with connected components analysis for road extraction from orthophoto images. *Expert Syst Appl.* 2021;176:114908. <https://doi.org/10.1016/j.eswa.2021.114908>.
3. Abdollahi A, Pradhan B, Shukla N, Chakraborty S, Alamri A. Deep learning approaches applied to remote sensing datasets for road extraction: a state-of-the-art review. *Remote Sens.* 2020;12(9):1444. <https://doi.org/10.3390/rs12091444>.
4. Baatz M. Multi resolution segmentation: an optimum approach for high quality multi scale image segmentation. Paper presented at the Beutrage zum AGIT-symposium. Salzburg, Heidelberg, 2000. 2000.
5. Bi D. A motion image pose contour extraction method based on B-spline wavelet. *Int J Antennas Propag.* 2021;2021.
6. Calderero F, Marques F. Region merging techniques using information theory statistical measures. *IEEE Trans Image Process.* 2010;19(6):1567–86. <https://doi.org/10.1109/TIP.2010.2043008>.
7. Cheng T, Wang X, Huang L, Liu W. Boundary-preserving mask r-CNN. Paper presented at the European conference on computer vision. 2020.
8. Das S, Mirnalinee TT, Varghese K. Use of salient features for the design of a multi-stage framework to extract roads from high-resolution multi-spectral satellite images. *IEEE Trans*

- Geosci Remote Sens. 2011;49(10):3906–31. <https://doi.org/10.1109/TGRS.2011.2136381>.
9. Gao L, Song W, Dai J, Chen Y. Road extraction from high-resolution remote sensing imagery using refined deep residual convolutional neural network. *Remote Sens.* 2019;11(5):552.
 10. Kahraman I, Karas IR, Akay AE. Road extraction techniques from remote sensing images: a review. *ISPRS international archives of the photogrammetry, remote sensing and spatial information sciences*, vol. XLII-4/W9. 2018. p. 339–42. <https://doi.org/10.5194/isprs-archives-XLII-4-W9-339-2018>.
 11. Lian R, Wang W, Mustafa N, Huang L. Road extraction methods in high-resolution remote sensing images: a comprehensive review. *IEEE J Sel Top Appl Earth Observ Remote Sens.* 2020;13:5489–507. <https://doi.org/10.1109/JSTARS.2020.3023549>.
 12. Liu B, Yu X, Zhang P, Tan X, Yu A, Xue Z. A semi-supervised convolutional neural network for hyperspectral image classification. *Remote Sens Lett.* 2017;8(9):839–48. <https://doi.org/10.1080/2150704X.2017.1331053>.
 13. Luo J, Guo C-E. Perceptual grouping of segmented regions in color images. *Pattern Recogn.* 2003;36(12):2781–92. [https://doi.org/10.1016/S0031-3203\(03\)00170-5](https://doi.org/10.1016/S0031-3203(03)00170-5).
 14. Maboudi M, Amini J, Hahn M. Objects grouping for segmentation of roads network in high resolution images of urban areas. *Int Arch Photogramm Remote Sens Spat Inf Sci.* 2016;41:897.
 15. Maboudi M, Amini J, Malihi S, Hahn M. Integrating fuzzy object-based image analysis and ant colony optimisation for road extraction from remotely sensed images. *ISPRS J Photogramm Remote Sens.* 2018;138:151–63. <https://doi.org/10.1016/J.ISPRSJPRS.2017.11.014>.
 16. Maboudi M, Amini J (2015) Object based segmentation effect on road network extraction from satellite images. In: *Proceedings of the 36th Asian conference on remote sensing*, Manila, Philippines, October 2015. pp. 19–23.
 17. Mahdi G. Hierarchical Bayesian regression with application in spatial modeling and outlier detection. University of Arkansas; 2018.
 18. Mayer H, Hinz S, Bacher U, Baltsavias E. A test of automatic road extraction approaches. *Int Arch Photogramm Remote Sens Spat Inf Sci.* 2006;36(3):209–14.
 19. Mnih V, Hinton GE. Learning to detect roads in high-resolution aerial images, Berlin, Heidelberg. 2010.
 20. Ning J, Zhang L, Zhang D, Wu C. Interactive image segmentation by maximal similarity-based region merging. *Pattern Recogn.* 2010;43(2):445–56.
 21. Shuai H, Xu X, Liu Q. Backward attentive fusing network with local aggregation classifier for 3D point cloud semantic segmentation. *IEEE Trans Image Process.* 2021;30:4973–84. <https://doi.org/10.1109/TIP.2021.3073660>.
 22. Steger C, Glock C, Eckstein W, Mayer H, Radig B. Model-based road extraction from images. In: *Automatic extraction of man-made objects from aerial and space images*. Springer; 1995. pp. 275–84.
 23. Wang F, Li Y. Mapping road based on multiple features and B-GVF snake. *Int J Pattern Recognit Artif Intell.* 2020;34(14):2050035.
 24. Wang S, Mu X, Yang D, He H, Zhao P. Road extraction from remote sensing images using the inner convolution integrated encoder-decoder network and directional conditional random fields. *Remote Sens.* 2021;13(3):465. <https://doi.org/10.3390/rs13030465>.
 25. Wang Y, Shen D, Teoh EK. Lane detection using spline model. *Pattern Recogn Lett.* 2000;21(8):677–89.
 26. Wang S, Yang H, Wu Q, Zheng Z, Wu Y, Li J. An improved method for road extraction from high-resolution remote-sensing images that enhances boundary information. *Sensors.* 2020;20(7):2064. <https://doi.org/10.3390/s20072064>.
 27. Wang W, Yang N, Zhang Y, Wang F, Cao T, Eklund P. A review of road extraction from remote sensing images. *J Traff Transp Eng (Engl Ed).* 2016;3(3):271–82. <https://doi.org/10.1016/j.jtte.2016.05.005>.
 28. Wang Y, Teoh EK, Shen D. Structure-adaptive B-snake for segmenting complex objects. Paper presented at the Proceedings 2001 international conference on image processing (Cat. No. 01CH37205). 2001.
 29. Xin J, Zhang X, Zhang Z, Fang W. Road extraction of high-resolution remote sensing images derived from DenseUNet. *Remote Sens.* 2019;11(21):2499. <https://doi.org/10.3390/rs11212499>.
 30. Xu Y, Xie Z, Feng Y, Chen Z. Road extraction from high-resolution remote sensing imagery using deep learning. *Remote Sens.* 2018;10(9):1461.
 31. Yang X, Li X, Ye Y, Zhang X, Zhang H, Huang X, Zhang B. Road detection via deep residual dense u-net. Paper presented at the 2019 international joint conference on neural networks (IJCNN). 2019.
 32. Yuan Y, Xun G, Jia K, Zhang A. A multi-view deep learning framework for EEG seizure detection. *IEEE J Biomed Health Inform.* 2018;23(1):83–94. <https://doi.org/10.1109/JBHI.2018.2871678>.
 33. Zhang Z, Liu Q, Wang Y. Road extraction by deep residual u-net. *IEEE Geosci Remote Sens Lett.* 2018;15(5):749–53.
 34. Zhang Z, Wang Y, Liu Q, Li L, Wang P. A CNN based functional zone classification method for aerial images. In: *2016 IEEE international geoscience and remote sensing symposium (IGARSS)*. pp. 5449–52. 2016. <https://doi.org/10.1109/IGARSS.2016.7730419>.
 35. Zhao J, Fang Y, Li G. Recurrence along depth: deep convolutional neural networks with recurrent layer aggregation. *Adv Neural Inf Process Syst.* 2021;34:10627–40.
 36. Zhuang L, Zhang Z, Wang L. The automatic segmentation of residential solar panels based on satellite images: a cross learning driven U-Net method. *Appl Soft Comput.* 2020;92: 106283.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.