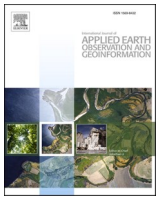
International Journal of Applied Earth Observations and Geoinformation 116 (2023) 103159

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



A comparative study of loss functions for road segmentation in remotely sensed road datasets

Hongzhang Xu [a], Hongjie He [a], Ying Zhang [b], Lingfei Ma [c,*], Jonathan Li [a,d,*]

[a] Geospatial Intelligence and Mapping Laboratory, Department of Geography and Environmental Management, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada
[b] Canada Centre for Mapping and Earth Observation, Natural Resources Canada, 560 Rochester Street, Ottawa, Ontario K1S 5H4, Canada
[c] School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 102206, China
[d] Department of Systems Design Engineering, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada

ARTICLE INFO

Keywords:
Road extraction
Image segmentation
Loss function
Cross-entropy
Dice
D-LinkNet

ABSTRACT

Road extraction from remote sensing imagery is a fundamental task in the field of image semantic segmentation. For this goal, numerous supervised deep learning techniques have been created, along with the employment of various loss functions that play a crucial role in determining the performances of supervised learning models. However, there is a lack of comprehensive analysis of the performance differences between the loss functions for road segmentation in remote sensing imagery. Therefore, this study conducts a comparative study of 12 well-known loss functions used widely in the field of image segmentation by training and evaluating the representative D-LinkNet network for road segmentation tasks with two publicly available remote sensing road datasets consisting of very high-resolution aerial and satellite images. The results show that different loss functions could lead to very different outcomes using the D-LinkNet, with varying focuses such as on overall model performances, *precision*, or *recall*. By dividing the loss functions into the distribution-based, region-based, and compound ones, we found that the region-based loss function type led to generally better model performances than the distribution-based one in terms of F_1, *IoU*, and the road segmentation maps, with the compound loss function type being comparable to the region-based one. This paper eventually tries to offer suggestions for choosing the loss function that best suits the purposes of road segmentation-related studies.

1. Introduction

Automated road extraction from remotely sensed imagery is an integral part of many remote sensing applications, such as intelligent transportation management (Guerrero-Ibañez et al., 2021), image registration processing (Tondewad and Dale, 2020), and topographic database updating (Mena, 2003). Furthermore, the accuracy of road extraction notably affects the detection of other objects such as vehicles (Abraham & Sasikumar, 2013), buildings (Simler, 2011), and oil well pads (He et al., 2022). Therefore, automated road segmentation is of general interest to researchers in the remote sensing community.

Numerous studies have been conducted in recent years to address the challenge of automated road extraction from high to very high spatial resolution remote sensing imagery, with the use of deep learning (DL) techniques becoming the norm (Lian et al., 2020). The DL-based road segmentation algorithms can be understood as to address a binary image classification problem. In other words, the road segmentation is completed by classifying the pixels that represent the road in remote sensing images, in which the road is the foreground and other elements are the background. In more recent years, DL-based road segmentation approaches usually adopt an encoder-decoder structure (Abdollahi et al., 2020). This idea is to first encode and down-sample the input image using convolutional operations, thereby gradually extracting high-level features. The mapping of features is then performed by employing a decoder procedure to recover the classification results of each pixel layer by layer. Therefore, each pixel in the output image will correspond to a target class. The well-known Fully Convolutional Network (FCN) (Long et al., 2015) and U-Net (Ronneberger et al., 2015)

* Corresponding authors at: School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 102206, China (L. Ma) and Geospatial Intelligence and Mapping Laboratory, Department of Geography and Environmental Management, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada (J. Li).
E-mail addresses: l53ma@cufe.edu.cn (L. Ma), junli@uwaterloo.ca (J. Li).

https://doi.org/10.1016/j.jag.2022.103159
Received 21 October 2022; Received in revised form 8 December 2022; Accepted 14 December 2022

are examples of semantic segmentation networks that use this structure. In particular, these networks have been refined and widely employed for the application of automated road extraction tasks. For example, Zhong et al. (2016) used an FCN-4s model for road segmentation in aerial remote sensing imagery, achieving about 66 % *recall*. Zhang et al. (2018) adopts residual blocks in the U-Net, which reduced the number of network parameters and increased computational efficiency. Similarly, Singh & Dash (2019) combines two U-Net networks to perform a two-step road extraction from aerial images.

The training of DL-based road segmentation networks is typically conducted in a supervised fashion (Abdollahi et al., 2020; Lian et al., 2020). This means that the model uses labeled samples (also known as "ground truth") to adjust its parameters so that it can make correct predictions on unknown data. However, the accuracy and efficiency of a supervised learning-based DL model heavily rely on the loss function, which is a measure of the distance between the true values of training samples and those of model predictions. There are many loss functions designed for addressing tasks in the field of image semantic segmentation (Ma et al., 2021). Unfortunately, there is no universal loss function that works perfectly for all types of data. Different loss functions can have different characteristics, such as being more sensitive to certain types of errors or being more efficient to compute. Therefore, it is important to carefully evaluate the performance of different loss functions in the context of a specific model and task.

To the best of our knowledge, there have been few studies on discussing loss functions in the particular field of road segmentation in remote sensing imagery. Considering road extraction plays a crucial role in a wide range of remote sensing applications such as topographic database updating and intelligent transportation management, understanding how different loss functions impact the performance of DL-based road segmentation models can be valuable for improving the accuracy and efficiency of relevant tasks. This study fills this gap in the literature by conducting a thorough evaluation of 12 well-known loss functions on road segmentation in remotely sensed imagery. In particular, the D-LinkNet, a representative DL network developed specifically for road segmentation, with two public remotely sensed road datasets consisting of aerial and satellite images, respectively, are utilized to examine and analyze the differences in the model performance between different loss functions in terms of evaluation metrics such as *accuracy*, *precision*, and *recall*. This research provides valuable insights into the effects of different loss functions on the model performance with regard to road segmentation, contributes to the broader field of image semantic segmentation, and may inform the development of improved road segmentation methods for remote sensing applications. Three detailed contributions of this paper are as follows:

(1) The common loss functions for binary semantic segmentation are thoroughly reviewed and compared.
(2) Comparative analyses of 12 loss functions are presented regarding road extraction from two public remote sensing road datasets using the D-LinkNet model.
(3) Suggestions for the selection of an appropriate loss function in road segmentation tasks are offered based on our findings.

The remainder of this paper is organized as follows. The mathematical expressions and characteristics of the 12 loss functions examined in this paper are summarized in Section 2. The D-LinkNet architecture for road segmentation, two publicly available road datasets, and quantitative evaluation metrics are all covered in Section 3. Section 4 elaborates on the designed experiments and the evaluation results of the models with different loss functions. Section 5 provides a summary of key findings in this study and provides suggestions for the choice of loss functions regarding road segmentation in remote sensing imagery.

2. Loss functions

The loss function is a fundamental concept in fields such as statistics, economics, and machine learning (ML)/DL, which is used to map the values of a random event or its associated random variables to non-negative real numbers, representing the function of "risk" or "loss" of that event (Goodfellow et al., 2016). In supervised learning, the *loss* represents the deviation between the true and model-predicted values, which is a measure of the model performance during training. The more the loss function can expand the inter-class distance and shrink the intra-class distance of samples, the higher the model's predictive performance will be. The use of loss functions is an indispensable component of the training process, and it is usually placed in the output layer of DL networks, which is responsible for feeding back the calculated *loss* to prior layers and updating the network parameters according to the *loss* value in order to adjust the model for a better fit on the training data. Typically, there is a large loss between the predicted and true values at the start of training, but after updating the network parameters through gradient descent methods in the backpropagation process, the predicted values gradually approach the true values until the *loss* is reduced below an acceptable threshold or no longer decreases.

The design and selection of loss functions are crucial to the application of DL networks in road extraction tasks. In this study, we will examine 12 loss functions in semantic segmentation. The reasons why these loss functions are chosen are as follows. First, we concentrate on the loss functions for general use instead of those designed for specific applications with less common usage. Second, the loss functions for multi-class segmentation are excluded given road segmentation is essentially a binary classification problem. In addition, the selected loss functions in this paper can be categorized into three very popular groups: distribution-based, region-based, and compound losses (Ma et al., 2021). The distribution-based loss category is a measure of the distance between the predicted and true values in a pixel-by-pixel fashion, whereas the region-based loss type measures the non-overlap between the road segmentation map and the ground truth map. On the other hand, the compound loss type is a combination of both types, thereby leveraging pixel- and region-level losses (Zhou et al., 2018). The commonly used and representative loss functions in the three categories are accordingly presented in detail in the following sub-sections.

2.1. Distribution-based loss functions

2.1.1. Binary Cross-Entropy (BCE)

The pixel-wise cross entropy loss is the most fundamental one and commonly used for the task of image segmentation in ML/DL. For road extraction, there are only a road class and a non-road class. It is therefore necessary to use a binary cross-entropy (BCE) loss, which represents the deviation of the predicted probability distribution from the true one, as shown below:

$$\mathscr{L}_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log(\widehat{p}_i) + (1-y_i)\log(1-\widehat{p}_i)\right) \tag{2.1}$$

where $y_i \in \{0,1\}$ is a binary class label for pixel i with 1 being positive class and 0 being negative class. While $\widehat{p}_i \in [0,1]$ refers to the predicted likelihood that pixel i will be classified into the positive class, $(1-\widehat{p}_i) \in [0,1]$ is the probability that pixel i will be categorized into the negative one. The input image's total number of pixels is denoted by the letter N.

2.1.2. Weighted Cross-Entropy (WCE)

Considering that most of the pixels in a large remote sensing image in relation to road extraction are usually negative examples (i.e., non-road class, labeled as "0″) and only a very few pixels belong to the road, it is difficult to fit the model well on the training data using the BCE loss function. This is because even if in the worst case the classifier predicts all labels as the negative class, it is still acceptable in terms of

classification accuracy due to the dominant number of negative example pixels in the input image, but this result is senseless in terms of road extraction and results in poor predictive performance of the trained model. Thus, the Weighted Cross-Entropy (WCE) loss (Pihur et al., 2007) is proposed to account for such a class imbalance.

The major improvement of the WCE on the BCE loss is adding weights to the positive examples, i.e., assigning a smaller weight to the loss of the dominating background, and a larger weight to that of the foreground. It can be described as:

$$\mathscr{L}_{WCE} = -\frac{1}{N}\sum_{i=1}^{N}(\beta y_i \log(\widehat{p}_i) + (1-y_i)\log(1-\widehat{p}_i)) \tag{2.2}$$

where β is the coefficient that weighs the loss of positive examples. Whereas setting $\beta > 1$ reduces false negatives and consequently increases *recall*, $\beta < 1$ cuts false positives, thus increasing *precision*.

2.1.3. Balanced Cross-Entropy (BalanCE)

The balanced cross-entropy (BalanCE) (Xie and Tu, 2017) is a further improvement on the WCE loss by adding the $(1-\beta)$ coefficient to the negative examples, which can be defined as follows:

$$\mathscr{L}_{BalanCE} = -\frac{1}{N}\sum_{i=1}^{N}(\beta y_i \log(\widehat{p}_i) + (1-\beta)(1-y_i)\log(1-\widehat{p}_i)) \tag{2.3}$$

When $\beta = 0.5$, it is equivalent to the BCE loss.

2.1.4. Focal loss

The Focal loss (Lin et al., 2020) can also be considered as a variation of the BCE, which can be written as:

$$\mathscr{L}_{\text{Focal}} = -\frac{1}{N}\sum_{i=1}^{N}(\alpha(1-\widehat{p}_i)^{\gamma} y_i \log(\widehat{p}_i) + (1-\alpha)\widehat{p}_i^{\gamma}(1-y_i)\log(1-\widehat{p}_i)) \tag{2.4}$$

where $\gamma \geq 0$, and when $\gamma = 0$, the Focal loss function becomes the BalanCE. Similar to the BalanCE loss, α is a parameter that accounts for class imbalance.

With the Focal loss, the examples with smaller errors (called easy examples) are downweighed, which drives the model to concentrate more on learning difficult ones (with larger errors) (Lin et al., 2020). In general, the prediction probability of easy examples is higher than that of hard ones. Here, assume the former be $\widehat{p}_1 > 0.5$ and the latter $\widehat{p}_2 < 0.5$. For positive examples, when $\gamma > 1$, $(1-\widehat{p}_1)^{\gamma}$ does not decrease as fast as $(1-\widehat{p}_2)^{\gamma}$. Therefore, adding the parameter γ can make the loss of easy examples smaller than that of hard ones. In other words, the network reduces the influence of easy examples and pays more attention to hard ones.

2.2. Region-based loss functions

2.2.1. Jaccard loss

The Jaccard index is a measure of the similarity between two finite sets. It is calculated as the ratio between the intersection of the positive instances between two sets to their mutual combined values:

$$J(X,Y) = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|} \tag{2.5}$$

where X and Y stand for the predicted and true values, respectively. $|X \cap Y|$ represents the common elements between set X and set Y, and $|X|$ refers to the number of elements in set X (likewise for $|Y|$).

With a value range of $[0,1]$, if the Jaccard index is closer to 1, then the prediction is closer to the ground truth. It can be modified to act as a loss function as follows:

$$\mathscr{L}_{Jaccard} = 1 - J = 1 - \frac{\sum_{i=1}^{N} y_i \widehat{p}_i + 1}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \widehat{p}_i - \sum_{i=1}^{N} y_i \widehat{p}_i + 1} \tag{2.6}$$

It is worth noting that a constant value, 1, also known as a "smooth" parameter, is added to both the numerator and the denominator to avoid instability of the function when $y = \widehat{p} = 0$.

2.2.2. Dice loss

Like the Jaccard coefficient, the Dice coefficient (Sudre et al., 2017) is a similarity index defined as follows:

$$D(X,Y) = \frac{2|X \cap Y|}{|X| + |Y|} \tag{2.7}$$

Correspondingly, the Dice loss function is formulated as follows:

$$\mathscr{L}_{Dice} = 1 - D = 1 - \frac{2\sum_{i=1}^{N} y_i \widehat{p}_i + 1}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \widehat{p}_i + 1} \tag{2.8}$$

2.2.3. Squared Dice (sDice) loss

The Dice loss has another form of expression, in which a squared sum instead of a simple sum of values in the denominator is used (see Eq. (2–9)). Some researchers, such as Milletari (2018), speculated in favor of this formulation of the Dice loss because its derivative is zero when the prediction equals the ground truth while that of the former Dice loss (i.e., Eq. (2–8)) is not. In this study, both Dice losses will be examined for comparison, and in order to distinguish one another, we term the second one the "Squared Dice (sDice)" loss.

$$\mathscr{L}_{sDice} = 1 - \frac{2\sum_{i=1}^{N} y_i \widehat{p}_i + 1}{\sum_{i=1}^{N} y_i^2 + \sum_{i=1}^{N} \widehat{p}_i^2 + 1} \tag{2.9}$$

2.2.4. Log-Cosh Dice (lcDice) loss

The Log-Cosh Dice (lcDice) loss function (Jadon, 2020) is a refinement of the Dice loss. According to the author, by applying a log of cosh function to the Dice loss, the lcDice loss can become tractable while incorporating the characteristics of the Dice coefficient. Its mathematical expression can be written as:

$$\mathscr{L}_{lcDice}(x) = \log(\cosh x) = \log\left(\frac{e^x + e^{-x}}{2}\right) \tag{2.10}$$

where $x = \mathscr{L}_{Dice}$ is the Dice loss.

2.2.5. Tversky loss

The Tversky coefficient (Salehi et al., 2017) is a generalization of the Jaccard and Dice coefficients, which can be expressed as:

$$T(X,Y) = \frac{|X \cap Y|}{|X \cap Y| + \alpha|X - Y| + \beta|Y - X|} \tag{2.11}$$

where $|X - Y|$ indicates modifying set X by removing its elements that also belongs to set Y, and likewise for $|Y - X|$. α and β are the weight added to the penalties for false negatives and false positives, respectively, and by adjusting the two parameters, the balance between *recall* and *precision* can be controlled. When $\alpha = \beta = 0.5$, the Tversky coefficient becomes the Dice coefficient, and when $\alpha = \beta = 1$, the Tversky coefficient degrades to the Jaccard coefficient.

Setting $\beta = 1 - \alpha$, the Tversky loss is written as:

$$\begin{aligned}\mathscr{L}_{Tversky} &= 1 - T \\ &= 1 - \frac{\sum_{i=1}^{N} y_i \widehat{p}_i + 1}{\sum_{i=1}^{N} y_i \widehat{p}_i + \alpha\sum_{i=1}^{N} y_i(1-\widehat{p}_i) + (1-\alpha)\sum_{i=1}^{N}(1-y_i)\widehat{p}_i + 1}\end{aligned} \tag{2.12}$$

where a larger α would favor *recall* over *precision*.

2.2.6. Focal Tversky(fTversky) loss

Like the Focal loss, the Focal Tversky (fTversky) loss (Abraham & Khan, 2019) aims to leverage hard examples by using a γ modifier, which downweights easy examples in favour of hard ones as shown below:

$$\mathscr{L}_{fTversky} = (1-T)^{\gamma} = \mathscr{L}_{Tversky}^{\gamma} \tag{2.13}$$

2.3. Compound loss functions

2.3.1. BCE-Dice loss

It is also feasible to combine different loss functions to create a single loss function. The BCE-Dice loss is one example, which refers to a weighted sum of the BCE and Dice losses:

$$\mathscr{L}_{BCE-Dice} = (1-\alpha)\mathscr{L}_{BCE} + \alpha\mathscr{L}_{Dice} \tag{2.14}$$

where α is the coefficient that weighs the Dice loss, $\mathscr{L}_{Dice}$, against the BCE loss, $\mathscr{L}_{BCE}$. It is quite a popular loss function in data competitions (Zhou et al., 2018).

2.3.2. Combo loss

A similar compound loss to the BCE-Dice is the Combo loss (Taghanaki et al., 2019), which is calculated as a weighted sum of the BalanCE and Dice losses:

$$\mathscr{L}_{Combo} = (1-\alpha)\mathscr{L}_{BalanCE} + \alpha\mathscr{L}_{Dice} \tag{2.15}$$

where α is a weighting coefficient for the BalanCE loss, $\mathscr{L}_{BalanCE}$, and the Dice loss, $\mathscr{L}_{Dice}$. By replacing the BCE with the BalanCE, the Combo loss tends to account for more class imbalance than the BCE-Dice loss.

3. Data and method

3.1. Public road datasets

This paper makes use of two publicly available road datasets: the Massachusetts roads dataset (Mnih, 2013) and the DeepGlobe road extraction dataset (Demir et al., 2018). Both are popular and extensively used remote sensing road datasets (Buslaev et al., 2018; He et al., 2019; Panboonyuen et al., 2018), with the former containing aerial RGB images and the latter containing satellite RGB images.

3.1.1. Massachusetts roads dataset

The Massachusetts roads dataset contains 1171 aerial RGB images of Massachusetts in the U.S. With a resolution of 1.0 m/pixel and a size of 1500 × 1500 pixels, each image is coupled with a mask image in grayscale, with white pixels standing for roads and black ones representing the background.

A training set of 1108 images, a validation set of 14 images, a test set of 49 images, and associated labelled images make up the dataset. This dataset covers an area of approximately 2600 km^2 with diverse landscapes, containing roads in urban, suburban, and rural regions.

3.1.2. DeepGlobe road extraction dataset

The DeepGlobe road extraction dataset is obtained from "Road Extraction Challenge Track" in "DeepGlobe 2018 Challenge". It is sampled from DigitalGlobe's "+Vivid", a high-quality base map covering Thailand, Indonesia, and India, with satellite images collected from the WorldView-2/3 satellites. The dataset contains 6226, 1243, and 1001 images for training, validation, and test, respectively, covering different types of road data in rural and urban areas. However, only training images (72.7 % of the whole dataset) are provided with corresponding mask images of road labels. The satellite images are in RGB combination, while the mask images are grayscale. Both satellite and mask images are in a size 1024 × 1024 and have 0.5 m/pixel resolution.

In this paper, since the original validation and test sets are not annotated, only 6226 images in the original training set and their paired mask images are utilized as the sample data for our road segmentation experiments. Further, these 6226 images and their corresponding labeled images are randomly split into 4368, 928, and 930 pairs as the training, validation, and test set, respectively.

3.2. Methodology

In this study, the methodology for comparing different loss functions is developed by configuring the D-LinkNet model with each of the 12 loss functions for training on the Massachusetts and DeepGlobe datasets, respectively. Each trained model is then evaluated on the test set from the two road datasets, respectively, in a quantitative manner based on evaluation metrics including *precision*, *recall*, F_1 score, and *IoU*, as well as in a qualitative manner by generating road segmentation maps using the trained models and comparing them to the ground truths.

3.2.1. D-LinkNet

D-LinkNet is a convolutional neural network (CNN)-based network that was developed specifically for the task of road segmentation in remote sensing imagery (Zhou et al., 2018). It is derived from the popular U-Net architecture (Ronneberger et al., 2015), which has been widely used for tasks such as medical image segmentation, satellite image analysis, and object detection. D-LinkNet incorporates several improvements over U-Net, including the use of dilated convolutions, which allows it to maximize the extent of the perceptual field and facilitate multi-scale feature fusion.

The reasons for the adoption of D-LinkNet in this study are as follows. First, in addition to dilated convolution layers, D-LinkNet also features an encoder-decoder structure, residual blocks, and skip connections, which enables efficient and accurate road segmentation in high-resolution aerial and satellite remote sensing imagery. Second, D-LinkNet has been shown to perform well in road segmentation tasks compared to its counterparts. Apart from its championship at the DeepGlobe 2018 Road Extraction Challenge (Demir et al., 2018), D-LinkNet has also been proven by articles in more recent years for its advantages over other DL methods for road segmentation in remote sensing images. For example, Chen et al. (2022) showed that D-LinkNet outperformed other well-known DL-based road segmentation networks, including SegNet (Badrinarayanan et al., 2017), DCS-TransUperNet (Zhang et al., 2022), DeepLabV3+ (Chen et al., 2018), U-Net, CRAE-Net (Li et al., 2022), Res-UNet (Zhang et al., 2018), and DiResNet (Ding and Bruzzone, 2021) on both the Massachusetts and DeepGlobe road datasets in terms of F_1 score. Similar results were reported by Jie et al. (2022), in which D-LinkNet achieved better overall performance than RoadNet (Liu et al., 2019), SegNet, NL-LinkNet (Wang et al., 2021), U-Net, DeepLabV3+, PP-LinkNet (Tran et al., 2020), and LinkNet (Chaurasia and Culurciello, 2017) also on these two datasets. Finally, D-LinkNet has been widely used and validated in various other fields. Some notable examples include its use in segmenting brain tumors in medical images (Bi et al., 2022), detecting driver behavior in video surveillance footage (Zhang et al., 2019), and extracting buildings (Zhu et al., 2020) and helping detect oil well pads (He et al., 2022) from satellite imagery. These applications highlight the robustness, generalizability, and effectiveness of D-LinkNet in solving complex segmentation tasks in various domains. Overall, the adoption of D-LinkNet in this study is justified by its proven performance and versatility in related tasks.

Fig. 1 depicts the architecture of D-LinkNet, which is composed of three parts: the encoding part, the central part, and the decoding part. The encoding part transforms the information in input images into features by employing several combinations of a maxpooling layer, a convolutional layer, and 3 to 6 residual blocks after an initial convolutional filter of size 7 × 7 with a stride of 2. The residual block features skip connections, which could help improve the model's generalization ability, and this encoder structure is from a pre-trained ResNet34 (He et al., 2016). The central part consists of dilated convolutional layers in both cascaded and parallel mode, which could enlarge the receptive field while retaining spatial information. The decoding part uses a "bottleneck" building block structure (He et al., 2016) for residual modules, which utilizes a 1 × 1 convolutional kernel to enhance the computational efficiency of the network. Finally, using transposed

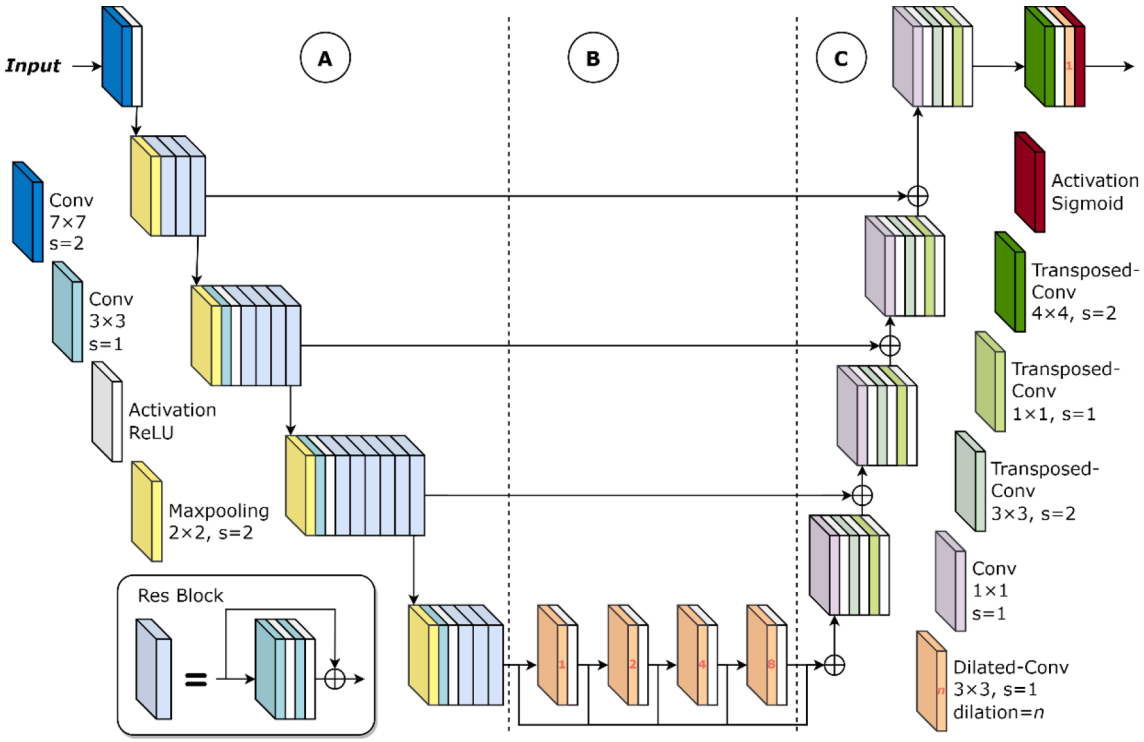


Fig. 1. Architecture of D-LinkNet (adapted from Zhou et al. (2018)). The circled A, B, and C denote the encoding, central, and decoding part, respectively.

convolution-based up-sampling strategies, the encoded feature maps are restored to the original image size (Chen et al., 2017). It is important to note that the input would be down-sampled 32 times after the encoding part because of the initial 7×7 convolutional filter with stride 2 and three 2×2 maxpooling filters with stride 2; therefore, in order to restore the encoded features to its original size, the size of the input image is required to be a multiple of 32.

3.2.2. Evaluation metrics

The commonly used evaluation metrics for road segmentation are *precision*, *recall*, F_1 score, (pixel) *accuracy*, and *IoU*:

$$Precision = \frac{TP}{TP + FP} \tag{3.1}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.2}$$

$$F_1 = \frac{2 \times (Precision \times Recall)}{Precision + Recall} = \frac{2TP}{2TP + FP + FN} \tag{3.3}$$

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{3.4}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{3.5}$$

where *TP* and *FP* denote true and false positives, indicating outcomes in which the model correctly and incorrectly predicts the positive class, respectively, while *FP* and *FN* stand for true and false negatives that are correctly and incorrectly predicted outcomes regarding the negative class, respectively.

Whereas *precision* denotes the ratio of correctly predicted road pixels to all the pixels predicted as road, *recall* indicates the proportion of correctly predicted road pixels to all road pixels in the ground truth image. F_1 is the harmonic mean of *precision* and *recall* (Taha and Hanbury, 2015). While *accuracy* is the percentage of correct predictions to total predictions, *IoU* refers to the proportion of the overlap between predicted road pixels and ground truth pixels to their union.

4. Experiments and results

4.1. Experiment settings

Table 1 lists the 12 loss functions used for road segmentation and their parameters for subsequent experiments in this study. The values of parameters in each loss function are determined based on the recommendation in previous source papers, which are also noted in the table. One reason we used suggested values rather than tuning all the hyperparameters to improve performance for some loss functions is that the hyperparameters obtained through searching may not be generally applicable. This is because the best hyperparameters for one task may not be the best for others. Therefore, using suggested values can provide

Table 1
Summary of loss functions used in this study and their parameter settings for experiments.

Type	Loss function	Expression	Parameter
Distribution-based Loss	BCE	Eq. (2–1)	–
	WCE	Eq. (2–2)	$\beta = \frac{1}{N}\sum_{i=1}^{N}(1-y_i)$(Xie and Tu, 2017)
	BalanCE	Eq. (2–3)	$\beta = \frac{1}{N}\sum_{i=1}^{N}(1-y_i)$(Xie and Tu, 2017)
	Focal	Eq. (2–4)	$\alpha = 0.25, \gamma = 2$(Lin et al., 2020)
Region-based Loss	Jaccard	Eq. (2–6)	–
	Dice	Eq. (2–8)	–
	sDice	Eq. (2–9)	–
	lcDice	Eq. (2–10)	–
	Tversky	Eq. (2–12)	$\alpha = 0.7$ (Salehi et al., 2017)
	fTversky	Eq. (2–13)	$\alpha = 0.7, \gamma = 0.75$(Abraham and Sasikumar, 2013)
Compound Loss	BCE-Dice	Eq. (2–14)	$\alpha = 0.5$(Zhou et al., 2018)
	Combo	Eq. (2–15)	$\alpha = 0.5$(Taghanaki et al., 2019)

a good starting point for choosing hyper-parameters, even though they may not be the optimal choice for all tasks. It is worthwhile to state that according to Xie and Tu (2017), the β coefficient in the BalanCE loss can be defined as $\frac{1}{N}\sum_{i=1}^{N}(1-y_i)$, in which y_i is the true label of pixel i, and N refers to the image's total number of pixels. This setting is believed to alleviate label-imbalanced problems, especially when the non-road pixels greatly outnumber the road ones, because in that case a larger β would be assigned to add more importance to positive samples. For consistency, in this study we set the weighting parameter β in the WCE loss to be the same as that in the BalanCE loss function.

In terms of the settings of experiments, for the Massachusetts dataset, the image size input to the network is 1500 × 1500 × 3, and then a resizing process is performed to resize the image to 1504 × 1504 × 3 using the bilinear interpolation method from the Python Pillow package, in order to be compatible with the input size (i.e., a multiple of 32) of the D-LinkNet (c.f., Section 3.2.1) while retaining the resampling to a minimum degree. The reason why a different input image size from the original D-LinkNet paper (i.e., 1024 × 1024 × 3) is used is to leverage the context of the entire input image while cutting computational costs

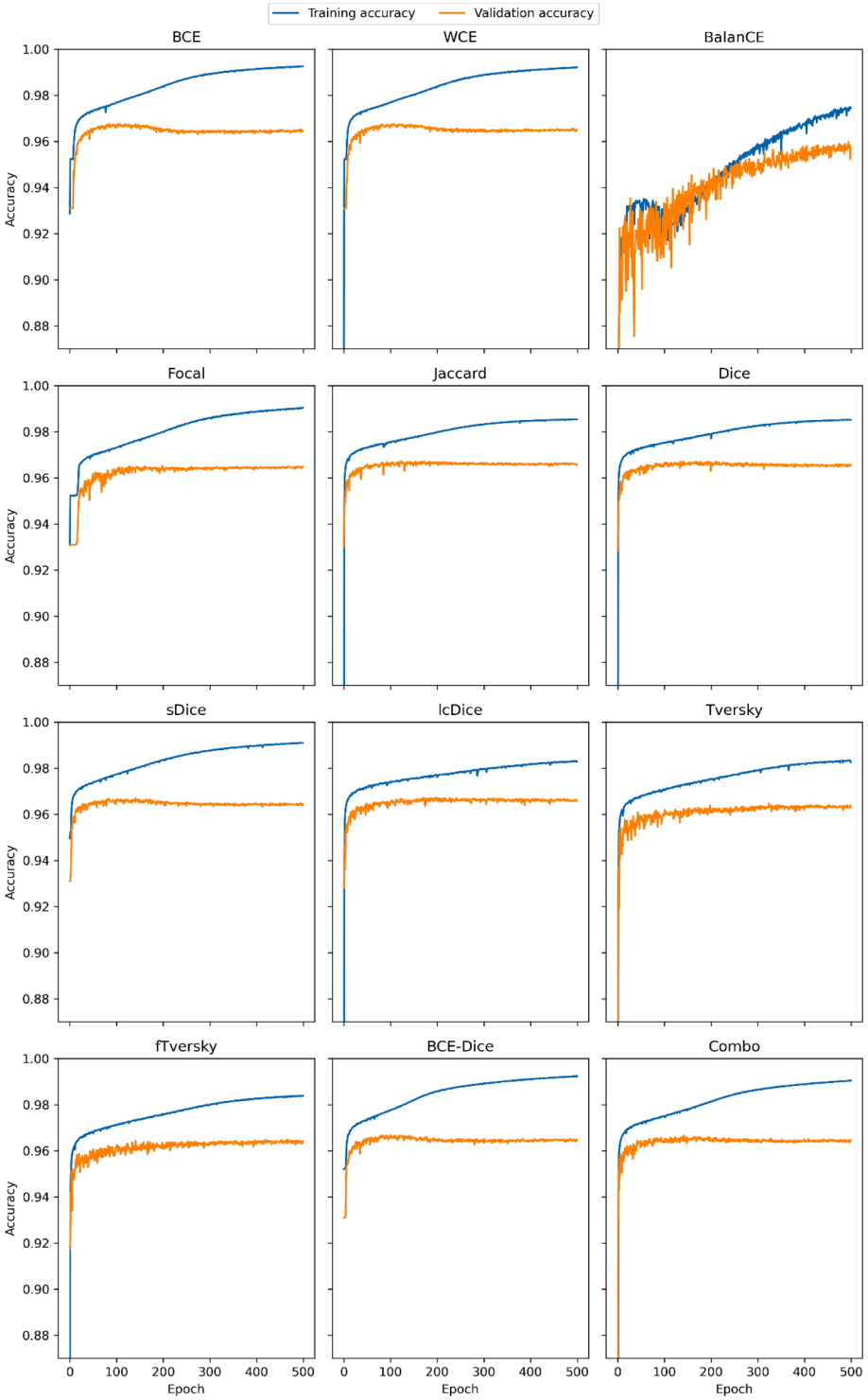


Fig. 2. Training *vs.* validation *accuracy* over epoch for D-LinkNet trained with different loss functions on the Massachusetts roads dataset.

compared to the random cropping strategy that would generate multiple 1024 × 1024 × 3 crops from a single original image. In addition, the fully convolutional structure of D-LinkNet permits differently sized input. In terms of the DeepGlobe dataset, the image size input to the network is 1024 × 1024 × 3. The training and validation batch sizes are set to 4 and 14 for the Massachusetts roads and 8 and 32 for the DeepGlobe dataset, respectively. The batch sizes are chosen based on the maximum available GPU memory size. In addition to the above, other settings for D-LinkNet models trained on the two datasets with different loss functions are the same: the Adam optimizer (Kingma and Ba, 2017) is selected for model optimization with an initial learning rate of 0.0001, the input image data during training are augmented with horizontal and vertical flips, and the training and validation epochs are all set to 500. All the models are trained and evaluated using the Ubuntu 20.04 operating system with 32 GB of RAM, along with the Tensorflow DL framework and dual Nvidia GeForce GTX 1080 Ti cards for accelerated computing.

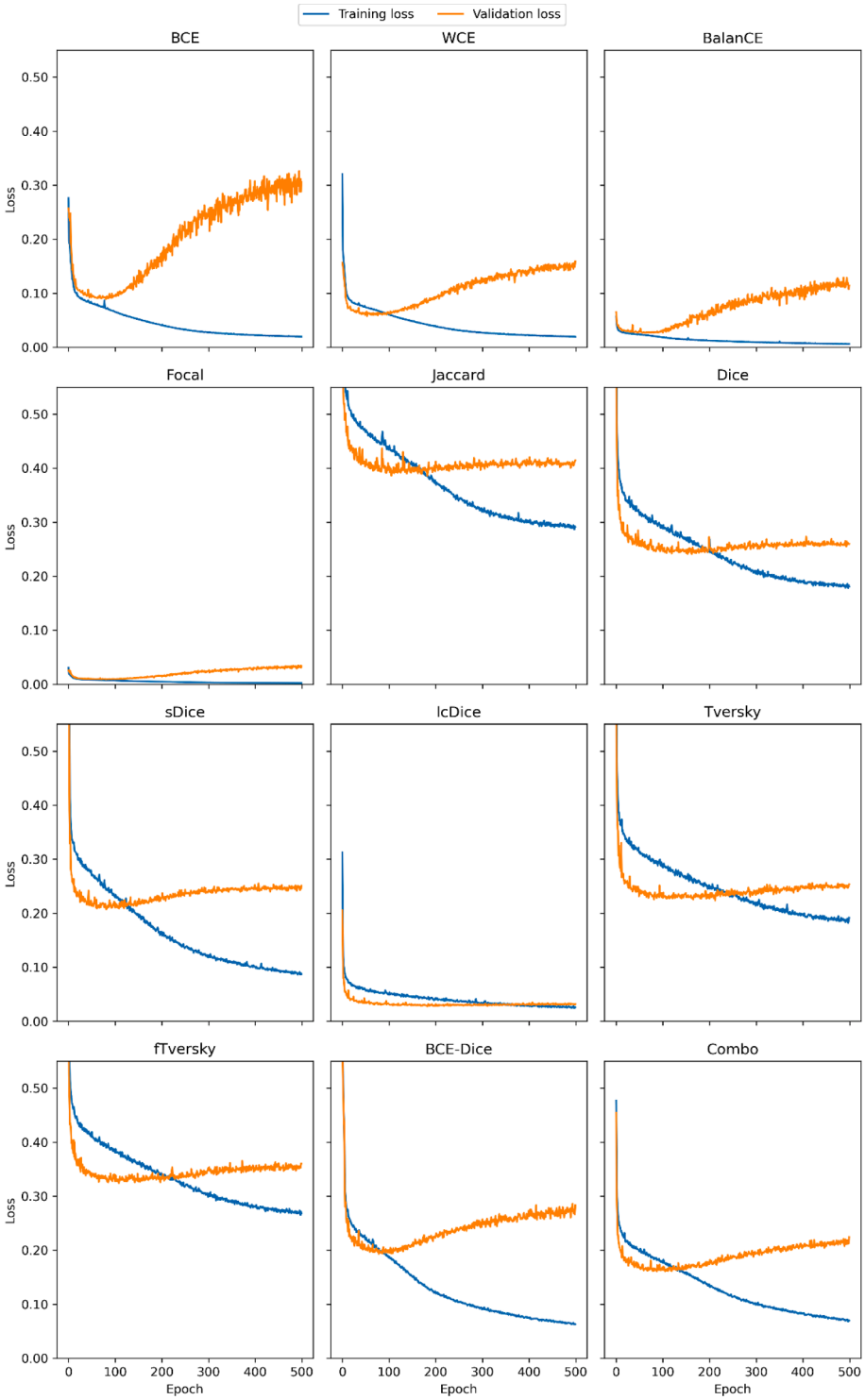


Fig. 3. Training *vs.* validation *loss* over epoch for D-LinkNet trained with different loss functions on the Massachusetts roads dataset.

4.2. Comparison of loss functions in road segmentation

4.2.1. Evaluation of model performance

Figs. 2-5 illustrate the training and validation progresses of D-LinkNet models with the 12 loss functions for the Massachusetts and DeepGlobe datasets, respectively. In general, the *accuracy* in the training phase keeps increasing over time, whereas in the validation it first rises and then remains constant or slightly decreases before stabilizing. The *loss* metric exhibits the opposite trend: the training *loss* steadily declines, while the validation *loss* reaches its lowest point at an early epoch before rising. The *accuracy* and *loss* values in the validation phase are subject to wide oscillations compared to in the training.

For the Massachusetts dataset (Figs. 2 & 3), all models achieve high *accuracy*, with the best validation *accuracy* over epoch exceeding 0.96 except for the BalanCE loss function, in which the maximum *accuracy* is only slightly lower than 0.96. In addition, it can be observed that large oscillations of validation *accuracy* occur for the model with the BalanCE loss. However, such oscillations slightly diminish with regard to its validation *loss* values. On the other hand, the performances in terms of the *loss* metric vary greatly from model to model. The D-LinkNet trained

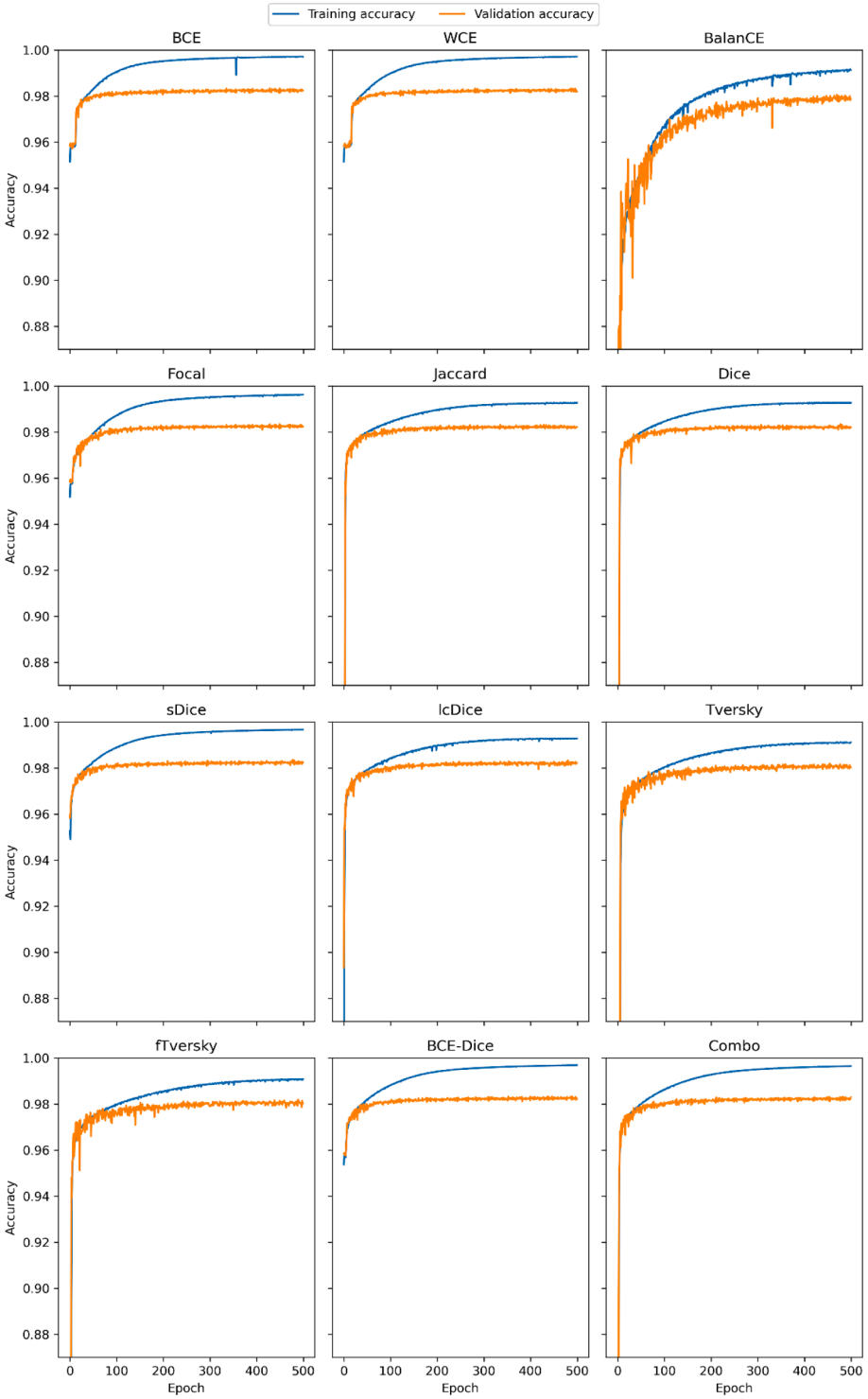


Fig. 4. Training *vs.* validation *accuracy* over epoch for D-LinkNet trained with different loss functions on the DeepGlobe road extraction dataset.

with the Focal loss obtains the lowest validation *loss* value, which is smaller than 0.01. In contrast, the lowest points on the validation *loss* curves of the other loss functions are much higher than that. Moreover, the region-based loss functions, except for the lcDice loss, tend to result in higher *loss* values than the other two types, with the Jaccard loss being the worst, whose *loss* values level off around 0.40. The models trained on the DeepGlobe dataset yield similar results (Figs. 4 & 5), with the *accuracy* for all models surpassing 0.98 and the Focal loss function obtaining the lowest *loss* value. It is worth noting that there are fewer oscillations in validation *loss* for the model trained with the BalanCE loss on the DeepGlobe dataset than on the Massachusetts dataset.

After the training, for each loss function, only the model with trained weights from the epoch when the lowest *loss* is achieved is saved for use in the test phase. In consequence, a total of 24 trained D-LinkNet models are obtained, i.e., one for each loss function for each road dataset.

Table 2 and Table 3 summarize the values of the five evaluation metrics obtained by applying the best trained D-LinkNet models with different loss functions to the test set of the two road datasets, respectively. Overall, the distribution-based loss functions achieve better *precision* than the region-based ones, but an opposite trend can be seen in

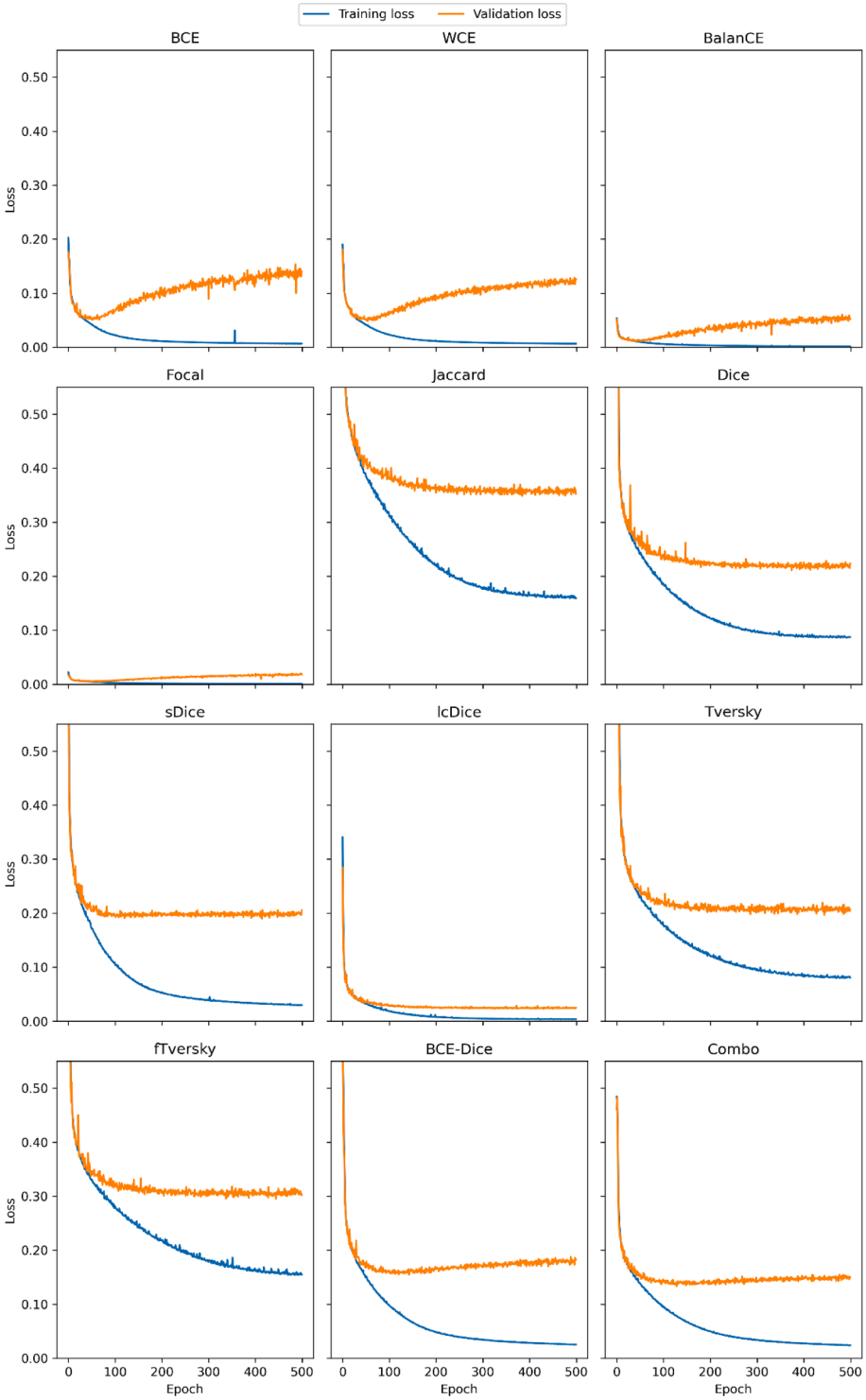


Fig. 5. Training *vs.* validation *loss* over epoch for D-LinkNet trained with different loss functions on the DeepGlobe road extraction dataset.

Table 2
Comparison of different loss functions for the evaluation of D-LinkNet trained on the Massachusetts roads dataset.

Dataset	Loss function	*Precision* (%)	*Recall* (%)	F_1(%)	*Accuracy* (%)	*IoU* (%)
Massachusetts roads dataset	BCE	81.70	71.07	76.01	**97.89**	61.31
	WCE	85.08	64.36	73.28	97.80	57.83
	BalanCE	45.94	**93.41**	61.59	94.53	44.50
	Focal	**91.35**	51.46	65.83	97.49	49.07
	Jaccard	77.36	77.34	77.35	97.87	63.06
	Dice	76.69	78.60	77.63	97.87	63.44
	sDice	78.43	75.49	76.93	97.87	62.51
	lcDice	76.35	79.52	**77.90**	97.88	**63.80**
	Tversky	69.10	84.42	76.00	97.49	61.29
	fTversky	69.87	84.32	76.42	97.55	61.83
	BCE-Dice	74.14	79.51	76.73	97.73	62.25
	Combo	75.21	79.88	77.48	97.82	62.51

Table 3
Comparison of different loss functions for the evaluation of D-LinkNet trained on the DeepGlobe road extraction dataset.

Dataset	Loss function	*Precision* (%)	*Recall* (%)	F_1(%)	*Accuracy* (%)	*IoU* (%)
DeepGlobe road extraction dataset	BCE	78.42	71.38	74.73	97.94	59.66
	WCE	78.55	71.11	74.64	97.94	59.54
	BalanCE	40.73	**95.27**	57.07	93.89	39.93
	Focal	**87.05**	60.28	71.23	97.92	55.32
	Jaccard	79.81	76.55	78.15	98.18	64.13
	Dice	78.70	77.82	78.26	98.16	64.28
	sDice	79.25	77.80	**78.52**	**98.19**	**64.64**
	lcDice	80.80	75.44	78.03	**98.19**	63.98
	Tversky	74.61	81.40	77.86	98.03	63.74
	fTversky	72.78	83.08	77.59	97.95	63.38
	BCE-Dice	77.97	77.78	77.87	98.12	63.76
	Combo	75.37	79.69	77.47	98.02	63.22

terms of *recall*, with the exception of the BalanCE loss, which results in much better *recall* and worse *precision* than its counterparts. As for the region-based loss functions, the Jaccard, the Dice, the sDice, and the lcDice losses, which resemble each other in mathematical expressions, bring about comparable results, with lcDice being slightly better. The two compound loss functions (i.e., the BCE-Dice and the Combo) accomplish equal or slightly worse model performances to the region-based ones in terms of *precision*, *recall*, and F_1. All the loss functions realize a satisfying and comparable pixel *accuracy* of around 97 % to 98 %, except for the BalanCE loss, whose *precision* is about 3 % to 4 % less. This result of high *accuracy* is reasonable given that the background (i.e., non-road pixels) in the image predominantly outnumber the foreground (i.e., road pixels), especially in rural areas with a scarcity of roads. The order of *IoU* is identical to that of F_1, with its value being smaller than the latter.

For the Massachusetts roads dataset, the highest *precision* and *recall* are achieved by the D-LinkNet models trained with the Focal and the BalanCE loss functions, respectively. The lcDice loss results in the best model performance on the test set in respect to F_1 and *IoU*. As for the DeepGlobe dataset, the Focal and the BalanCE losses still top the *precision* and the *recall*, respectively. The sDice loss realizes the highest score of F_1, *accuracy*, and *IoU*, while the lcDice loss function ties for the first place in terms of *accuracy*. It is important to note that the models with the BalanCE and the Focal losses do not perform well on the test set in terms of F_1 and *IoU*, though they ranked in the top two in terms of validation *loss* during training, indicating the models overfit the training data. However, it is worthwhile to note that the overfitting problem in the BalanCE loss is alleviated in the Combo loss, which is a combination of the BalanCE and the Dice losses. We also find that the two forms of Dice losses (i.e., the Dice and the sDice) lead to similar model performances, with each win over the other in terms of F_1 and *IoU* across the two road datasets. In addition, the improvement of the fTversky loss over the Tversky is very limited.

Another notable result is that different loss functions differ in their focus on either *precision* or *recall*. For example, the WCE and the Focal loss functions weigh in favor of *precision*, while the BalanCE and the Tversky losses pay more attention to *recall*. This phenomenon is closely related to the values of their functions' parameters used in this study, as noted in Table 1. To start with, the value of the β coefficient in the WCE (c.f., Eq. (2–2)) is defined as $\frac{1}{N}\sum_{i=1}^{N}(1-y_i)\langle 1$, which downweighs *FP*s, thus increasing the *precision*. On the other hand, the BalanCE loss, which is an improved WCE loss function by adding the $(1-\beta)$ coefficient to the negative samples (c.f., Eq. (2–3)), suppresses the obtained *precision* while encouraging the *recall*. This is because given the non-road pixels greatly outnumber the road ones in the datasets used in this study, $\beta > 0.5 > 1-\beta$ would mean more emphasis on cutting *FN*s and consequently increase *recall*. Similarly, setting $\alpha = 0.25 < 1-\alpha$ in the Focal loss would promote *precision*. Moreover, as the Tversky coefficient is a generalization of the Jaccard and Dice coefficients (c.f., Section 2.8), setting a larger α (i.e., 0.7 in this study) in the Tversky loss reduces *FP*s, thus increasing the *precision* values.

4.2.2. Road segmentation results

The best trained D-LinkNet models with the 12 loss functions are further evaluated by generating road segmentation maps with white pixels (value of 255) representing the foreground (i.e., roads) and black ones (value of 0) representing the background. Fig. 6 and Fig. 7 display several segmentation results from the Massachusetts and DeepGlobe road datasets, respectively, along with the original input images and the ground truths. These selected examples cover urban, suburban, and rural areas.

In general, the distribution-based loss function type appears to generate inferior road segmentation maps to the other two types. As can be seen, the region-based and the compound loss functions seem to reconstruct more complete road network structures than the distribution-based ones, with the exception of the BalanCE loss. The D-LinkNet model trained with the BalanCE loss, however, appears to overestimate the road pixels from the input test images, showing that there are many misidentifications (i.e., *FP*s) in the maps. On the contrary, the model with the Focal loss tends to make underestimates, producing discontinuous road paths compared to the other models.

Another notable result is that the densely interconnected road structure in urban regions is exceedingly difficult to restore from the images for all the three loss function types, as can be seen from the red rectangles drawn in the first column of the segmentation maps in Fig. 6. On the other hand, better results are from images with fewer roads, such as in sub-urban and rural areas. This can be well demonstrated by the example from the fourth column in Fig. 7, in which there is only one line of road in the input image, leading to similar road segmentation maps by

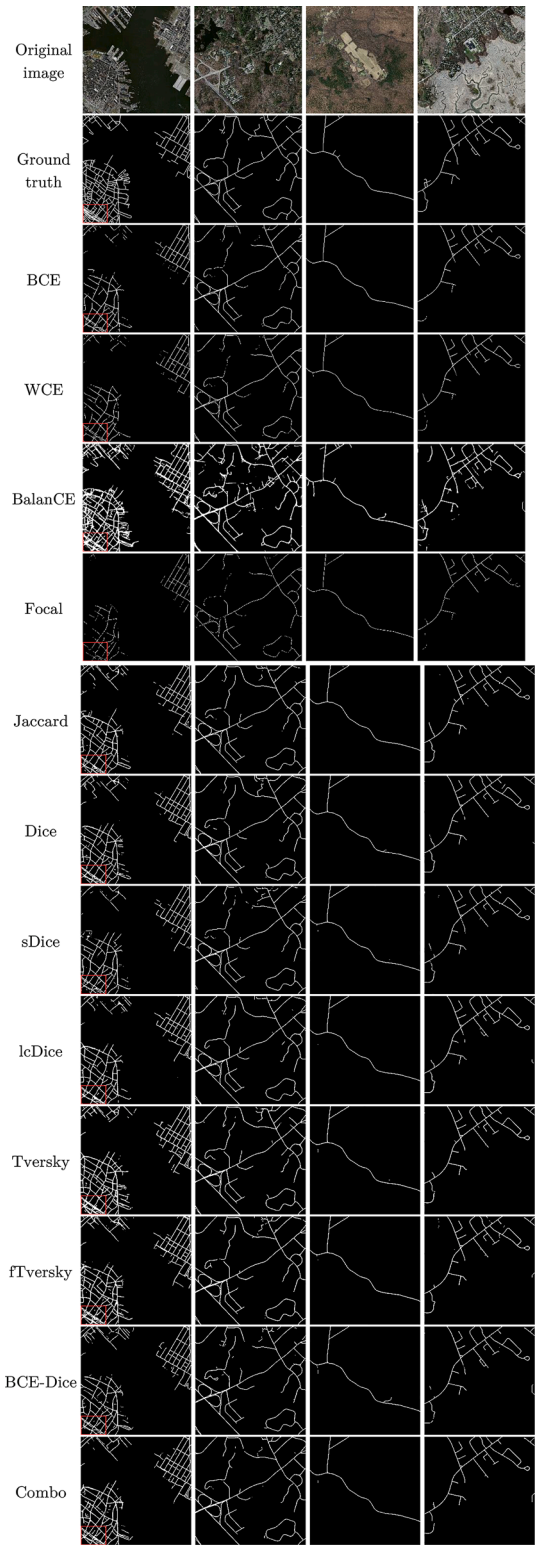


Fig. 6. Four examples of road segmentation results for the Massachusetts roads dataset using D-LinkNet trained with different loss functions.

nearly all the models to the ground truth. It is also worthwhile to note that the models with different loss functions can correctly distinguish roads from streams (see the bottom right of maps in the last column in Fig. 6). This finding could reveal that road networks in aerial and satellite RGB images share common characteristics regardless of the image background (vegetated or non-vegetated areas, rural or urban regions).

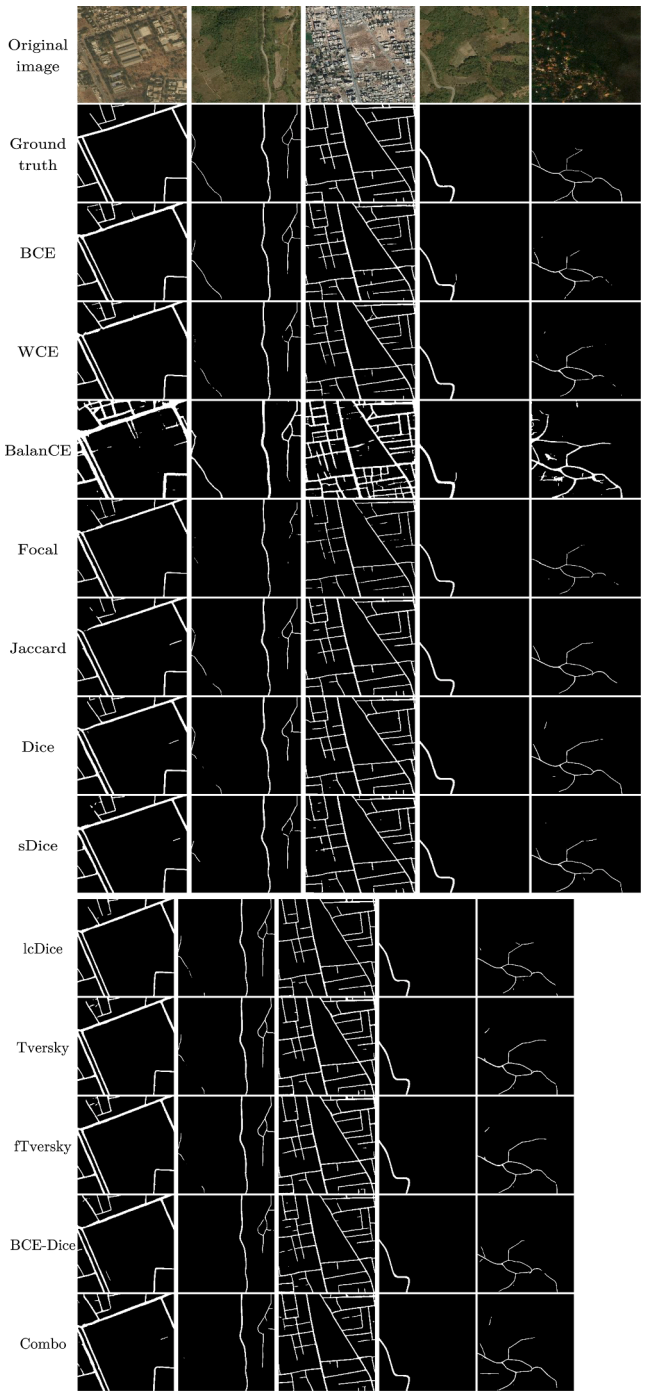


Fig. 7. Five examples of road segmentation results for the DeepGlobe road extraction dataset using D-LinkNet trained with different loss functions.

5. Conclusion

The paper conducted a comparative study on the effects of 12 general loss functions on road segmentation in high-resolution aerial and satellite remote sensing imagery. The D-LinkNet network was chosen to extract roads from two public remotely sensed road datasets, the Massachusetts roads dataset and the DeepGlobe road extraction dataset. For each dataset, D-LinkNet models with different loss functions were trained for 500 epochs using the same experiment settings, and then the best trained models for each loss function were evaluated by comparing their predictive performances between different models using common quantitative evaluation metrics and assessing their generated road

segmentation maps on the test set.

The results showed that different loss functions could result in different model performances with regard to road segmentation in remote sensing imagery. The region-based loss functions (i.e., the Jaccard, the Dice, the sDice, the lcDice, the Tversky, and the fTversky losses) led to generally better model performances than the distribution-based ones (i.e., the BCE, the WCE, the BalanCE, and the Focal losses) in terms of F_1, *IoU*, and the generated road segmentation maps, with the compound loss functions (i.e., the BCE-Dice and the Combo losses) being comparable to the region-based ones. The best F_1 scores of 77.90 % and 78.52 % are realized by the D-LinkNet model trained with the lcDice loss for the Massachusetts dataset and the one trained with the sDice loss for the DeepGlobe dataset, respectively. As for *precision* and *recall*, however, the distribution-based losses took the first place, with the Focal loss and the Balance loss topping *precision* and *recall*, respectively, for both datasets.

Based on the above, some suggestions could be made as to the choice of loss functions with respect to road extraction from remotely sensed imagery. First, the region-based loss functions, more specifically the Dice, the sDice, and the lcDice, are the best option for obtaining overall satisfactory model performances in reference to F_1 and *IoU*, thereby generating road segmentation maps with well re-constructed details. Second, the focus on either *precision* or *recall* could be shifted by adjusting the parameter values in the loss functions such as the BalanCE, the Focal, and the Tversky losses. Third, the compound loss functions offer the flexibility of altering parameter values while producing sub-optimal results. Such results could also achieve stability; for example, the overfitting problem in the BalanCE loss was alleviated in the Combo loss, which is a weighted sum of the BalanCE loss and the Dice loss. In addition, because the differences in model performances across the two road datasets are nearly consistent, our suggestions could hold true regardless of image source types (aerial or satellite RGB imagery), image size, and image resolutions.

To conclude, loss functions are crucial in determining the performance of supervised DL models in respect to road segmentation in remotely sensed road datasets. No one single loss function works perfectly in terms of all evaluation criteria across every dataset. As a result, it is critical to select the loss function that best suits the study's objectives, such as obtaining optimal results or focusing on either *precision* or *recall*. Our future research will concentrate on the use of ensemble learning strategies to combine different road segmentation results based on different loss functions. Other powerful DL networks such as the Transformer, could also be taken into account to improve the aggregated results.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This study was partially funded by the Emerging Interdisciplinary Project of Central University of Finance and Economics from China, as well as the Earth Observation for Cumulative Effects program of Canada Centre for Remote Sensing, National Resources Canada. The first author also thanks the China Scholarship Council for their assistance with a doctoral scholarship through the State Scholarship Fund (No. 202006270042).

References

Abdollahi, A., Pradhan, B., Shukla, N., Chakraborty, S., Alamri, A., 2020. Deep learning approaches applied to remote sensing datasets for road extraction: a state-of-the-art review. Remote Sens. 12 (9), 1444–1465.

Abraham, L., Sasikumar, M., 2013. A fuzzy based road network extraction from degraded satellite images. Proc. ICACCI 2032–2036.

Abraham, N., Khan, N.M., 2019. A novel focal tversky loss function with improved attention U-Net for lesion segmentation. Proc. ISBI 683–687.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39, 2481–2495.

Bi, R., Ji, C., Yang, Z., Qiao, M., Lv, P., Wang, H., Bi, R., Ji, C., Yang, Z., Qiao, M., Lv, P., Wang, H., 2022. Residual based attention-Unet combing DAC and RMP modules for automatic liver tumor segmentation in CT. Math. Biosci. Eng. 19, 4703–4718.

Buslaev, A., Seferbekov, S., Iglovikov, V., Shvets, A., 2018. Fully convolutional network for automatic road extraction from satellite imagery. Proc. IEEE CVPRW 197–1973.

Chaurasia, A., Culurciello, E., 2017. LinkNet: Exploiting encoder representations for efficient semantic segmentation. Proc. VCIP 1–4.

Chen, C., Tian, X., Wu, F., Xiong, Z., 2017. UDNet: Up-down network for compact and efficient feature representation in image super-resolution. Proc. ICCVW 1069–1076.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), in: Proc. ECCV, pp. 833–851.

Chen, X., Sun, Q., Guo, W., Qiu, C., Yu, A., 2022. GA-Net: A geometry prior assisted neural network for road extraction. Int. J. Appl. Earth Obs. Geoinf. 114, 103004.

Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., Raskar, R., 2018. DeepGlobe 2018: A challenge to parse the earth through satellite images. Proc. IEEE CVPRW 172–17209.

Ding, L., Bruzzone, L., 2021. DiResNet: Direction-aware residual network for road extraction in VHR remote sensing images. IEEE Trans. Geosci. Remote Sens. 59, 10243–10254.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning. MIT Press, Cambridge, MA, USA, Adaptive Computation and Machine Learning series.

Guerrero-Ibañez, J., Contreras-Castillo, J., Zeadally, S., 2021. Deep learning support for intelligent transportation systems. Trans. Emerg. Telecommun. Technol. 32 (3), e4169–e4190.

He, H., Xu, H., Zhang, Y., Gao, K., Li, H., Ma, L., Li, J., 2022. Mask R-CNN based automated identification and extraction of oil well sites. Int. J. Appl. Earth Obs. Geoinform. 112, 102875.

He, H., Yang, D., Wang, S., Zheng, Y., Wang, S., 2019. Light encoder–decoder network for road extraction of remote sensing images. J. Appl. Remote Sens. 13 (3), 034510.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proc. CVPR 770–778.

Jadon, S., 2020. A survey of loss functions for semantic segmentation. Proc. CIBCB 1–7.

Jie, Y., He, H., Xing, K., Yue, A., Tan, W., Yue, C., Jiang, C., Chen, X., 2022. MECA-Net: A multiscale feature encoding and long-range context-aware network for road extraction from remote sensing images. Remote Sens. 14, 5342.

Kingma, D.P., Ba, J., 2017. Adam: A method for stochastic optimization. arXiv: 1412.6980. [Online]. https://arxiv.org/abs/1412.6980.

Li, S., Liao, C., Ding, Y., Hu, H., Jia, Y., Chen, M., Xu, B., Ge, X., Liu, T., Wu, D., 2022. Cascaded residual attention enhanced road extraction from remote sensing images. ISPRS Int. J. Geo-Inf. 11, 9.

Lian, R., Wang, W., Mustafa, N., Huang, L., 2020. Road extraction methods in high-resolution remote sensing images: a comprehensive review. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 5489–5507.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2020. Focal loss for dense object detection. IEEE Trans. Pattern Anal. Mach. Intell. 42, 318–327.

Liu, Y., Yao, J., Lu, X., Xia, M., Wang, X., Liu, Y., 2019. RoadNet: Learning to comprehensively analyze road networks in complex urban scenes from high-resolution remotely sensed images. IEEE Trans. Geosci. Remote Sens. 57, 2043–2056.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. Proc. CVPR 3431–3440.

Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., Yang, X., Martel, A.L., 2021. Loss odyssey in medical image segmentation. Med. Image Anal. 71, 102035.

Mena, J.B., 2003. State of the art on automatic road extraction for GIS update: a novel classification. Pattern Recognit. Lett. 24, 3037–3058.

Milletari, F., 2018. Hough voting strategies for segmentation, detection and tracking. Technische Universität München. PhD dissertation.

Mnih, V., 2013. Machine learning for aerial image labeling. University of Toronto. PhD dissertation.

Panboonyuen, T., Vateekul, P., Jitkajornwanich, K., Lawawirojwong, S., 2018. An enhanced deep convolutional encoder-decoder network for road segmentation on aerial imagery. Int. Conf. Comput. Inf. Tech. 191–201.

Pihur, V., Datta, S., Datta, S., 2007. Weighted rank aggregation of cluster validation measures: a Monte Carlo cross-entropy approach. Bioinformatics 23, 1607–1615.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation. Proc. MICCAI 234–241.

Salehi, S.S.M., Erdogmus, D., Gholipour, A., 2017. Tversky loss function for image segmentation using 3D fully convolutional deep networks. Proc. MLMI 379–387.

Simler, C., 2011. An improved road and building detector on VHR images. Proc. IGARSS 507–510.

Singh, P., Dash, R., 2019. A two-step deep convolution neural network for road extraction from aerial images. Proc. SPIN 660–664.

Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M., 2017. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. Proc. DLMIA 240–248.
Taghanaki, S.A., Zheng, Y., Kevin Zhou, S., Georgescu, B., Sharma, P., Xu, D., Comaniciu, D., Hamarneh, G., 2019. Combo loss: Handling input and output imbalance in multi-organ segmentation. Comput. Med. Imaging Graph. 75, 24–33.
Taha, A.A., Hanbury, A., 2015. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med. Imaging 15, 1–28.
Tondewad, Ms.P.S., Dale, Ms.M.P., 2020. Remote sensing image registration methodology: review and discussion. Procedia Comput. Sci. 171, 2390–2399.
Tran, A., Zonoozi, A., Varadarajan, J., Kruppa, H., 2020. PP-LinkNet: Improving semantic segmentation of high resolution satellite imagery with multi-stage training. Proc. SUMAC 57–64.
Wang, G., Wang, S., Li, G., Zhao, G., Niu, Y., 2021. Elastic reflection waveform inversion with a nonlinear born scattering operator for multi-parameter reconstruction. IEEE Geosci. Remote Sens. Lett. 1.
Xie, S., Tu, Z., 2017. Holistically-nested edge detection. Int. J. Comput. Vis. 125, 3–18.
Zhang, C., Lu, Y., Feng, M., Wu, M., 2019. Trucker behavior security surveillance based on human parsing. IEEE Access 7, 97526–97535.
Zhang, Z., Liu, Q., Wang, Y., 2018. Road extraction by deep residual U-Net. IEEE Geosci. Remote Sens. Lett. 15, 749–753.
Zhang, Z., Miao, C., Liu, C., Tian, Q., 2022. DCS-TransUperNet: Road segmentation network based on CSwin Transformer with dual resolution. Appl. Sci. 12, 3511.
Zhong, Z., Li, J., Cui, W., Jiang, H., 2016. Fully convolutional networks for building and road extraction: Preliminary results. Proc. IGARSS 1591–1594.
Zhou, L., Zhang, C., Wu, M., 2018. D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. Proc. CVPRW 192–1924.
Zhu, Q., Li, Z., Zhang, Y., Guan, Q., 2020. Building extraction from high spatial resolution remote sensing images via multiscale-aware and segmentation-prior conditional random fields. Remote Sens. 12, 3983.