

Adaptive Structure-Aware Connectivity-Preserving Loss for Improved Road Segmentation in Remote Sensing Images

Sara Shojaei¹, Trevor Bohl¹, Kannappan Palaniappan¹, Filiz Bunyak¹

¹ Department of Electrical Engineering and Computer Science
 University of Missouri-Columbia, MO, USA

{ssfht, tdbkcx, pal, bunyak}@missouri.edu

Abstract

Accurate road extraction from remote sensing images is a critical step for many applications, including navigation, environmental monitoring, and disaster response. In remote sensing images, roads appear as thin curvilinear structures and are very prone to fragmentation due to challenges such as low-contrast materials and occlusions by surrounding vegetation or buildings. Although deep learning has greatly advanced image segmentation techniques, pixel-level loss functions are insufficient for preserving connectivity and topology of thin curvilinear structures such as roads. Small pixel-level errors can disrupt the extraction of accurate road graphs, complicating subsequent characterizations and analysis tasks. In this paper, we propose a novel, adaptive, structure-aware, connectivity-preserving loss function SAC-Loss. This loss function combines global and local processing to improve structure awareness and uses a proximity-based weighting scheme with asymmetric penalties to fill gaps in the road structures while limiting spurious detections elsewhere. Experimental results demonstrate improved road segmentation in terms of visual quality and quantitative performance.

1. Introduction

Automated road segmentation from remote sensing images is a critical first step for various applications, including improved navigation and mapping, traffic monitoring and management, environmental monitoring, urban planning, and disaster management and response. This task has gained significant importance for two key reasons: first, manual road extraction is time-consuming and labor-intensive, and second, aerial imaging at various scales—from drones to satellites has become increasingly affordable and accessible. Furthermore, advances in deep learning techniques for image analysis now present a wealth of opportunities to tackle this challenge more efficiently,

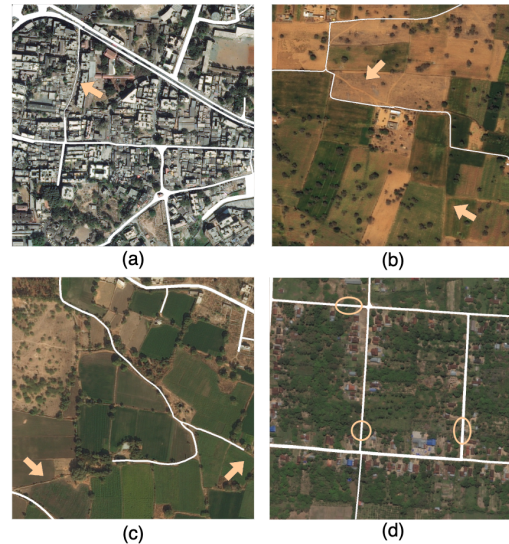


Figure 1. Sample images from the DeepGlobe dataset with road masks (ground truths) highlighting discontinuities and false detections challenges in road extraction caused by: (a) dense urban areas and varying road widths, (b-c) similar textures in agricultural fields, and (d) occlusions from trees or shadows.

more accurately, and with greater automation.

Road segmentation in remote sensing presents significant challenges due to the diverse characteristics of roads, such as variations in color and texture resulting from materials like asphalt, concrete, clay, and gravel; occlusions caused by tall buildings in off-nadir imagery, vegetation, shadows, and cloud cover; and the presence of similar structures like rivers and field boundaries (Figure 1).

Although deep learning has greatly advanced image segmentation techniques [1–5], segmentation of thin, curvilinear structures remains difficult due to their sensitivity to small pixel-level errors. Classical pixel-level segmentation loss functions, such as binary cross-entropy (BCE) or the Dice coefficient, treat all pixels equally, failing to prioritize topological accuracy. Depending on the location, even

small pixel-level errors can cause discontinuities in thin road structures, leading to significant inaccuracies in road graph extraction. These topological errors are particularly problematic for applications that depend on accurate road network data for decision-making, planning, and analysis.

In recent years, deep learning techniques have demonstrated considerable promise in road segmentation and extraction from remote sensing images [6, 7]. Tao et al. [6] proposed a new transformer-based architecture equipped with CNN layers to capture both local and long-range global features. They have also integrated a pixel connectivity structure (PCS) to more effectively preserve the connectivity of the roads. Recent studies [8–10] have demonstrated that modifying the loss functions alone, without altering the network architecture, can significantly enhance both the accuracy and connectivity of narrow and curvilinear structures. Yuan et al. [11] introduced a local loss function called GapLoss, which focuses on addressing the gaps between road segments. Nanni et al. [10] built upon previous work by incorporating both gaps and road direction in their approach. While both GapLoss [11] and its extension [10] demonstrated improvements in connectivity and accuracy, their final loss calculations are based solely on the mean error, which may overlook local details. Additionally, [10] requires four times more computational resources than the original GapLoss [11] to capture road direction, making it computationally expensive. Batra et al. [12] introduced an orientation learning task inspired by how humans trace roads, and developed a stacked multi-branch convolutional module to harness the mutual information between orientation learning and segmentation tasks. Bandara et al. [13] proposed an integrated network that combines multi-scale convolutional features with graph-based long-range spatial reasoning. Li et al. [14] extracted road features at the pixel, edge, and region levels, with each level being supervised by its mask. However, preparing these masks for each category adds significant labor, increases network complexity, and extends training time. Despite these advances, accurate road network extraction remains challenging.

In this paper, we propose a novel structure-aware, connectivity-preserving loss function for semantic road segmentation in remote sensing images. The proposed loss function aims to reconnect disconnected road segments by locally penalizing gaps in the predicted road areas. This loss function can be integrated into other object segmentation networks with similar goals, such as the segmentation of rivers in remote sensing images. Unlike the previous works inspired by the GapLoss, our approach takes advantage of the structural characteristics of roads both locally and globally. We achieve this by incorporating a weight map to penalize local gaps, which is then integrated into the Focal Tversky Loss to address global penalties. The main contributions of this paper can be summarized as follows:

1. We introduce a new structure-aware connectivity-preserving loss function, *SAC-Loss*, for segmenting curvilinear structures that are susceptible to connectivity errors. This function combines both global (gap identification) and local (adaptive weighting) processing to improve structure awareness. Distance-based weighting enables shape-aware, precise, and adjustable segmentation, improving accuracy.
2. The proposed loss function is adaptable and can be integrated into other currently available segmentation networks to analyze various curvilinear structures. Unlike computationally expensive local topology losses, such as the Betti loss [15], this function is efficient enough to be trained on less advanced hardware.
3. We evaluate the segmentation results using two base network architectures with several different loss functions across three road datasets: 1) the Massachusetts Road Dataset, 2) the DeepGlobe Road Dataset, and 3) the SpaceNet Challenge Road Dataset. Our results show that the proposed *SAC-Loss* function improves network performance, enabling it to outperform other CNN-based methods in road segmentation.

2. The Proposed Method

In this study, we introduce *SAC-Loss*, a novel adaptive structure-aware connectivity-preserving loss function, designed to improve road segmentation, particularly in terms of connectivity. The following sections describe the formulation of *SAC-Loss*, its integration into model architectures, the training procedures, and the evaluation metrics used in our experiments.

2.1. Conventional Binary-Cross Entropy Loss

Binary Cross Entropy (BCE) is a widely used and highly effective loss function for classification and segmentation tasks. It is defined as a measure of the difference between two probability distributions for a given random variable [16]. BCE loss is formulated as follows:

$$\text{BCE Loss} = -\frac{1}{N} \sum_{i=0}^N y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i) \quad (1)$$

where N represents the number of pixels in the image and y_i and $\hat{y}_i$ denotes the pixel values at position i in the ground truth and predicted masks respectively.

As a result of using the BCE loss, a segmentation network learns to assign probabilities closer to 1 for the foreground and closer to 0 for the background pixels marked in the ground truth. This allows the network to focus on pixel-level differences, leading to more accurate binary masks. While the BCE loss is commonly used in segmentation due

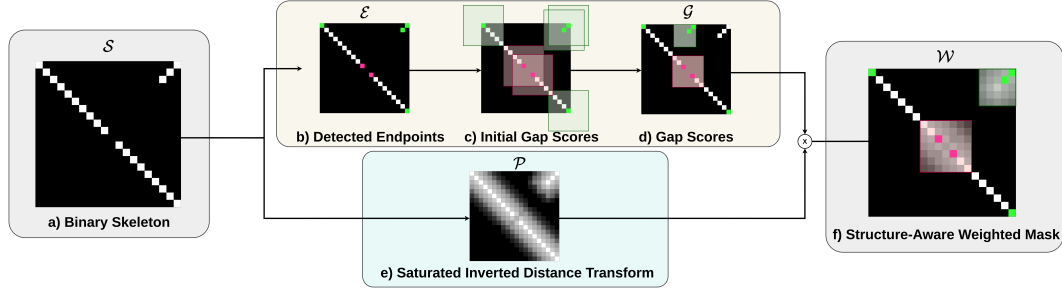


Figure 2. SAC-Loss calculation procedure: a) Binary skeleton extracted from the prediction/ground truth, b) Endpoints detected from skeleton S , c) Adding 9×9 weight window around the endpoints, d) Keeping the overlapped windows to generate gap matrix G , e) Inverted saturated distance transform of skeleton S , f) Distance transform P and the gap Map G multiplication as the final weight map W .

to its effectiveness and simplicity, it has limitations such as poor performance with highly imbalanced data and equal treatment of false positive (FP) and false negative (FN) errors. The BCE loss also suffers from the lack of structural awareness. It focuses on pixel-level classification and does not account for the spatial or structural relationships between pixels. In tasks where the connectivity or spatial continuity of features is important such as road segmentation, BCE may not capture long-range dependencies between pixels, leading to fragmented predictions.

2.2. Conventional Focal Tversky Loss

Tversky loss is a generalization of the Dice coefficient. It introduces asymmetric penalties for false positive (FP) and false negative (FN) errors, which is particularly useful in cases of class imbalance where the foreground class is rare compared to the background (as in road segmentation) or when one type of error may be more critical than the other. Tversky loss is formulated as below:

$$\text{Tversky Score}(GT, Pred) = \frac{TP}{TP + \alpha FP + \beta FN} \quad (2)$$

Where TP , α , and β are True Positives, the parameters controlling the importance of false positives and false negatives, respectively. Focal Tversky Loss combines the Tversky loss with the concept of focal loss to penalize harder-to-classify pixels more than easier ones. It is formulated as:

$$\text{Focal Tversky Loss} = (1 - \text{Tversky Loss})^\gamma \quad (3)$$

Where γ is the parameter controlling the aggressiveness of the loss function on hard-to-classify pixels. Larger γ biases the model to learn from difficult examples, which is helpful when the dataset is highly imbalanced or the boundaries between foreground and background are challenging.

2.3. Adaptive Structure Aware Connectivity Preserving Loss (SAC-Loss)

Road segmentation in remote sensing imagery suffers from class imbalance since roads occupy only a small por-

tion of the images. Road segmentation in remote sensing imagery also suffers from fragmentation since roads are thin curvilinear structures and few false negative pixels can convert a single connected road into multiple disconnected road segments. Factors such as occlusions (i.e. due to vegetation cover), low contrast (i.e. due to road surface material), or shadows (i.e. shadow of tall buildings or other structures surrounding the roads) lead to weak network responses, false negatives, and subsequent gaps in road masks. While the focal Tversky loss enables asymmetric penalties for false positive and false negative pixels and helps with class imbalance, the α and β parameters are set globally for the entire image before training. Increasing penalties for false negatives reduces gaps in road masks but raises false positives, causing spurious detections and inaccurate road networks.

We propose a new adaptive loss function to reduce fragmentation in thin curvilinear structures. Unlike the focal Tversky loss or GapLoss, the proposed loss function alters the penalties (α and β parameters in the focal Tversky loss) locally based on local structure. The main processing steps of the proposed structure-aware connectivity-preserving loss are as follows:

1. Preliminary segmentation and road structure analysis:

Using the likelihood map $\mathcal{L}$ predicted by the segmentation network, we first generate a preliminary binary segmentation mask $\mathcal{M}$ using thresholding:

$$\mathcal{M}(x, y) = \begin{cases} 0 & \text{if } \mathcal{L}(x, y) < 0.5 \\ 1 & \text{otherwise} \end{cases} \quad (4)$$

We then generate the skeleton $\mathcal{S}$ (Figure 2a) of the binary mask $\mathcal{M}$ using mathematical morphology operations. Using saturated distance transform, we compute a proximity map $\mathcal{P}$ (Figure 2e) :

$$D = \text{DistanceTransform}(1 - \mathcal{S}) \quad (5)$$

$$D(x, y) = \{d_{max} | D(x, y) \geq d_{max}\} \quad (6)$$

$$\mathcal{P}(x, y) = 1 - \frac{D(x, y)}{d_{max}} \quad (7)$$

Where d_{max} is the upper bound of the pixel distance in D (set to 10 pixels in our experiments). $\mathcal{P}$ indicates the proximity of the background pixels to the object of interest (roads) and acts like an attention map.

2. Gap detection: To detect the endpoints of road segments (potential gap boundaries), we convolve the skeleton S , with a 3×3 kernel, where all the values are set to one except for the center. We then locate all foreground pixels ($\mathcal{S}(x, y) = 1$) with a single neighbor (pixels where the output of the convolution is one). The process results in a binary matrix $\mathcal{E}$ (Figure 2b), the same size as the original image, where ones indicate the skeleton endpoints.

A gap matrix $\mathcal{G}$, the same size as the original image is initialized with the predicted mask. For each endpoint in $\mathcal{E}$, a $w \times w$ window (w empirically set to 9) in $\mathcal{G}$ centered around the endpoint is increased with a value of K . The process is cumulative, if a pixel (x, y) is covered by n endpoint windows, $\mathcal{G}(x, y)$ is assigned $n \times K$, increasing the probability of a gap in the region.

3. Adaptive structure-aware connectivity-preservation: Gap matrix $\mathcal{G}$ as the likelihood of gap presence, and proximity matrix $\mathcal{P}$ as proximity to a road structure are fused using element-wise multiplication to generate a locally adaptive structure-aware weighted mask $\mathcal{W}$ (Figure 2f):

$$\mathcal{W}(x, y) = \mathcal{G}(x, y) \cdot \mathcal{P}(x, y) \quad (8)$$

Finally, SAC-Loss is computed by feeding the above local structure-aware weighted mask to the Focal Tversky Loss.

$$\text{SAC-Loss} = \text{FocalTversky}(GT, \mathcal{W}) \quad (9)$$

Where GT is the ground truth binary mask. The pseudo-code for the proposed SAC-Loss is shown in Algorithm 1. This method can be easily implemented without adding any significant complexity to the training procedure or increasing the training time. The final weighted mask $\mathcal{W}$ is calculated outside the body of the network and only changes the weight coefficients of each pixel in the gradient loss calculation. Finally, our proposed structure-aware loss function—locally and globally—propagates through the network and optimizes the weights.

2.4. Network Architecture

We integrated the proposed SAC-Loss into two deep learning networks, the classical U-Net [17], and the more recent SemSeg [18] segmentation networks. U-Net employs a symmetric encoder-decoder structure with skip connections, enabling precise localization and context preservation. Its ability to perform well even with limited training

Algorithm 1 SAC-Loss Computation

Require: Binary segmentation mask $\mathcal{M}$, d_{max} , K , w

Ensure: SAC-Loss values

```

1: // Step-1: Road Structure Analysis
2:  $\mathcal{S} \leftarrow \text{GenerateSkeleton}(\mathcal{M})$ 
3:  $D = \text{DistanceTransform}(1 - \mathcal{S})$  // distance to skeleton
4: for each pixel  $(x, y)$  do
5:    $D(x, y) = \{d_{max} | D(x, y) \geq d_{max}\}$  //saturate dist
6:    $\mathcal{P}(x, y) = 1 - \frac{D(x, y)}{d_{max}}$  // proximity to roads
7: end for
8: // Step-2: Gap Detection
9:  $\mathcal{G} \leftarrow 0$  // Gap score
10:  $\mathcal{E} \leftarrow \text{DetectEndpoints}(\mathcal{S})$ 
11: for each pixel  $(x, y)$  where  $\mathcal{E}(x, y) == 1$  do
12:   // Increase  $\mathcal{G}$  around the skeleton endpoints
13:   //Window size around endpoints  $w$  and  $K$  as hyper-
     parameters
14:    $j = \text{Floor}[\frac{w}{2}]$ 
15:    $\mathcal{G}(x - j : x + j, y - j : y + j) += K$ 
16: end for
17: // Step-3: SAC-Loss Computation
18: // Proximity to roads and road endpoints increases the
     structure-aware weight
19:  $\mathcal{W}(x, y) = \mathcal{G}(x, y) \cdot \mathcal{P}(x, y)$ 
20: // Feeding the weighted mask  $\mathcal{W}$  to Focal Tversky Loss
21:  $\mathcal{Z} \leftarrow \text{FocalTversky}(GT, \mathcal{W})$ 
22: return  $\mathcal{Z}$ 

```

data makes it particularly suitable for our application. The SemSeg segmentation network [18] also consists of an encoder and a decoder. The encoder structure features two spatial and two receptive blocks that utilize dilated convolution layers which significantly expand the receptive field without increasing the number of parameters. To further enlarge the receptive field, max-pooling layers are incorporated in receptive blocks, as they increase the actual receptive field compared to dilated convolutions alone more effectively and help mitigate latency issues associated with high spatial dimensions. Transposed convolution layers restore the spatial dimensions after the receptive blocks. Attention blocks generate a 2D attention map to prioritize influential pixels in the image, which is crucial for accurate road segmentation in high-resolution satellite images. Both networks were trained using a weighted sum of the conventional binary-cross entropy (BCE) [16] and the proposed structure-aware connectivity preserving (SAC-Loss) losses:

$$\mathcal{L}_{\text{Total}} = k_1 \mathcal{L}_{\text{BCE}} + k_2 \times \mathcal{L}_{\text{SAC-Loss}} \quad (10)$$

where $k_1 = 0.8$ and $k_2 = 0.2$ control the weights of BCE and SAC-Loss losses, respectively.

3. Experimental Results

3.1. Datasets

We evaluated the performance of our loss function using three well-established remote sensing datasets. The experiments conducted on these datasets demonstrate the effectiveness of our approach, regardless of the unique characteristics of each dataset. The details about the datasets used in this study are as follows:

- **The Massachusetts Road Dataset [7]** collected from Massachusetts, USA, contains 1171 RGB images (1108 for training, 14 for validation, and 49 for testing) with binary ground truth masks. Each image has a size of 1500×1500 pixels with a resolution of 1 meter per pixel. The road labels are consistently 7 pixels thick with no smoothing, regardless of road type. The dataset covers 2600 square kilometers, encompassing rural, urban, and suburban areas.
- **The DeepGlobe Road Dataset [19]** has been introduced in the DeepGlobe Road Extraction Challenge in 2018. The images were captured from various locations such as Thailand and India. These images have a resolution of 50 cm per pixel and the total area covers 1716 square kilometers. The dataset consists of 6226 images (80% for training, 10% for validation, and 10% for testing). All the images are in RGB format with an image size of 1024×1024 pixels.
- **The Spacenet Road Dataset [20]** as part of the SpaceNet challenges, consists of high-resolution RGB satellite images from various global locations. The images range from 30 to 70 cm per pixel, with a resolution of 400×400 pixels. The dataset is divided into training, validation, and test sets, each with corresponding ground truth masks.

3.2. Evaluation Metrics

3.2.1 Topology-Aware Metrics

Conventional pixel-level segmentation metrics like recall, precision, and F1 score, which rely on the exact pixel-by-pixel overlap, face limitations when applied to thin curvilinear structures such as roads. These metrics are overly sensitive to minor pixel misalignments (e.g., small position shifts) and shape variations (e.g., small width changes); while being insufficiently sensitive to structural integrity-related changes (e.g., fragmentation).

Topology-aware metrics Completeness, Correctness, and Quality [21] have been introduced to address these issues. These metrics evaluate the global structure of the segmented objects, tolerating small pixel-level errors, which makes them more suitable for thin curvilinear structures. The computation of topology-aware metrics generally involves

three steps: (1) skeletonization, (2) buffer setting for misalignment, and (3) distance-based matching. These metrics compare the skeletons extracted from the predicted and the ground truth masks using a tolerance buffer set around the skeleton pixels. Matches are established using a distance-based criterion instead of the exact overlap of points. Figure 3 illustrates the matching process and computation of completeness and correctness metrics. In this work, we used a 5-pixel buffer around the skeletons. The evaluation metrics used in this paper are described below:

- **Completeness:** This metric calculates the fraction of the reference (ground truth) foreground that is within the buffer area of the skeletonized predicted foreground, as shown in Figure 3a. Completeness is a more flexible version of the conventional Recall measure.

$$Completeness = \frac{Length(Matched\ Reference)}{Length(Reference)} \quad (11)$$

- **Correctness:** This metric computes the fraction of predicted foreground that is within the buffer area of the skeletonized reference (ground truth) foreground, as shown in Figure 3b. Correctness is a more flexible, relaxed version of the conventional Precision measure.

$$Correctness = \frac{Length(Matched\ Extracted)}{Length(Extracted)} \quad (12)$$

- **Quality:** The Quality metric corresponds to the relaxed Intersection of Union (IoU), which is obtained from the Correctness and the Completeness metrics.

$$Quality = \frac{Compl. \times Correct.}{Compl. + Correct. - Compl. \times Correct.} \quad (13)$$

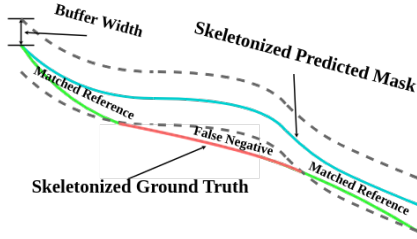
- **Relaxed F1-score:** The relaxed F1-score is also calculated based on the Correctness and the Completeness.

$$RelaxedF_1 = \frac{2 \times Compl. \times Correct.}{Compl. + Correct.} \quad (14)$$

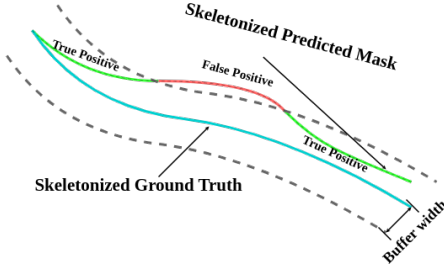
3.2.2 Graph-level Metrics

To assess the connectivity and the accuracy of the resulting road graphs, we also employed the Average Path Length Similarity (APLS) metric [20]. This graph-level metric is based on the dissimilarity of the shortest paths between all pairs of nodes within the road graph. For this evaluation, the ground truth and the prediction graphs are extracted and matched for comparison. The $S_{P \rightarrow T}$ calculates the average shortest path between each pair of nodes in the ground truth graph $G = (V, E)$ and the prediction graph $\hat{G} = (\hat{V}, \hat{E})$.

$$S_{P \rightarrow T} = 1 - \frac{1}{|V|} \sum \min(1, \frac{|L(a, b) - L(\hat{a}, \hat{b})|}{L(a, b)}) \quad (15)$$



(a) Matched Reference: GT to Pred. Mask-with-Buffer



(b) Matched Extraction: Pred. Mask to GT-with-Buffer

Figure 3. Relaxed matching protocol of the ground-truth skeleton (reference) and prediction skeleton (extracted) to compute completeness (a) and correctness (b) metrics [21].

where $a, b \in V$ and $\hat{a}, \hat{b} \in \hat{V}$, denote pairs of nodes in ground truth and prediction graphs; and $|V|$ is the number of nodes in the ground truth graph. The operators $L(a, b)$ and $L(\hat{a}, \hat{b})$ calculate the path lengths from $a \rightarrow b$ and $\hat{a} \rightarrow \hat{b}$, respectively. To account for penalties from false positives, the APLS is symmetrically calculated by adding the $S_{T \rightarrow P}$, where the ground truth and prediction graphs are swapped.

$$APLS = \frac{1}{N} \sum_{y, \hat{y}} \left(\frac{1}{\frac{1}{S_{P \rightarrow T(G, \hat{G})}} + \frac{1}{S_{T \rightarrow P(\hat{G}, G)}}} \right) \quad (16)$$

Where y and $\hat{y}$ are ground truth and prediction masks, and N is the number of images in the test set.

3.3. Implementation Details

For both network architectures, UNet [17] and SemSeg [18], we started the training process in each experiment with only the BCE loss, then added the various state-of-the-art losses. In all experiments, the UNet and SemSeg networks were trained for 20 and 200 epochs, respectively. The weighted sum of the individual loss functions (Eq. 10), has been used for training. In the experiments with two losses, the weights of the BCE loss (k_1) and SAC-Loss (k_2) were set to 0.8 and 0.2 respectively. For the Focal Tversky loss (Eq. 3), we have set the parameters α to 1.0 and β to

0.4. This configuration penalizes the false positives more than the false negatives. To place greater emphasis on more complex samples, we have set the parameter γ to $\frac{4}{3}$. During training and testing, the input images were resized to 512×512 before being fed to the networks. We used Adam optimizer [22] with a learning rate of 10^{-4} and weight decay of 10^{-3} . In our evaluations, we used a buffer of 5 pixels to calculate topology-level metrics.

3.4. Quantitative Results

3.4.1 Comparison of Loss Functions

We compared the performance of the proposed SAC-Loss to the other loss functions using the topology-aware and graph-level metrics described above. The results are summarized in Table 1. Out of six experiments (involving two network architectures and three datasets) and three evaluation metrics (relaxed F1, Quality (Relaxed IoU), and APLS), the proposed SAC-Loss was ranked highest 14 out of 18 cases and second highest 3 out of 18 cases. In terms of the graph-based APLS metric that measures the connectivity and accuracy of the predicted graph, the proposed SAC-Loss was ranked highest 5 out of 6 cases corresponding to two networks and three datasets with considerable increases in score ranging from 0.02 to 0.17 points.

For the Massachusetts dataset, adding the proposed SAC-Loss to the UNet architecture resulted in the highest scores for relaxed F1, Quality (Relaxed IoU), and graph-based APLS, with values of 0.9004, 0.8193, and 0.7289, respectively (over 0.02, 0.03, and 0.09 points increase over second-highest scores). When integrated into the SemSeg architecture, the proposed SAC-Loss resulted in the highest scores for relaxed F1 and Quality with values of 0.8899 and 0.8026 respectively, but fell slightly short in terms of APLS score with a value of 0.7036 (0.04 points lower compared to the best score of 0.7454 by BCE only).

For the DeepGlobe dataset, adding the proposed SAC-Loss to the UNet architecture resulted in the second-highest scores for relaxed F1 and Quality, with values of 0.8161 and 0.6835 respectively. While these results were slightly lower in terms of relaxed F1 and Quality scores (0.01 and 0.02 points lower compared to the highest scores respectively), they considerably improved the connectivity and accuracy of the predicted graph indicated by a 0.06 points increase in the APLS score. When integrated into SemSeg architecture, the proposed SAC-Loss resulted in the highest scores for all three metrics, relaxed F1, Quality, and APLS, with values of 0.8058, 0.6729, and 0.5796, respectively (over 0.02 points increase per metric compared to the second highest scores by BCE + FocalTversky).

For the SpaceNet dataset, adding the proposed SAC-Loss to the UNet network resulted in the highest scores for the relaxed F1 with a value of 0.7674 and the second highest score for Quality with a value of 0.6284 (only 0.0122 points lower

Table 1. Quantitative comparison of the various loss functions integrated to the UNet and the SemSeg networks tested on three remote sensing road datasets. Bold and underlined numbers represent the highest and the second-highest performance per category, respectively.

Dataset	Selected Loss Function	Compl. (Rlx. Rec.)	Corr. (Rlx. Pre.)	Rlx. F1-Score	Qual. (Rlx. IoU)	APLS
UNet Network Results						
Massachusetts	BCE	0.9635	0.4808	0.6415	0.4719	0.5601
	BCE + FocalTversky	0.9590	0.7195	0.8221	0.6983	0.6282
	BCE + GapLoss	0.8047	0.9697	<u>0.8795</u>	<u>0.7857</u>	<u>0.6353</u>
	BCE + SAC-Loss (proposed)	0.9318	0.8710	0.9004	0.8193	0.7289
DeepGlobe	BCE	0.9413	0.3648	0.5258	0.3560	<u>0.4573</u>
	BCE + FocalTversky	0.7334	0.8551	0.7896	0.6433	0.2754
	BCE + GapLoss	0.7983	0.8600	0.8280	0.7035	0.4456
	BCE + SAC-Loss (proposed)	0.8283	0.8042	<u>0.8161</u>	<u>0.6835</u>	0.5169
SpaceNet	BCE	0.6399	0.8853	0.7429	0.6017	0.2315
	BCE + FocalTversky	0.8868	0.6253	0.7334	0.5859	0.2637
	BCE + GapLoss	0.6958	0.8532	<u>0.7665</u>	0.6406	<u>0.3186</u>
	BCE + SAC-Loss (proposed)	0.8336	0.7109	0.7674	<u>0.6284</u>	0.4904
SemSeg Network Results						
Massachusetts	BCE	0.9282	0.8466	<u>0.8849</u>	<u>0.7959</u>	0.7454
	BCE + FocalTversky	0.9374	0.8247	0.8781	0.7822	<u>0.7392</u>
	BCE + GapLoss	0.8316	0.9144	0.8710	0.7726	0.7136
	BCE + SAC-Loss (proposed)	0.9189	0.8628	0.8899	0.8026	0.7036
DeepGlobe	BCE	0.9005	0.6437	0.7507	0.6001	0.5138
	BCE + FocalTversky	0.8999	0.6957	<u>0.7847</u>	<u>0.6441</u>	<u>0.5593</u>
	BCE + GapLoss	0.6763	0.8281	0.7445	0.5904	0.3105
	BCE + SAC-Loss (proposed)	0.8766	0.7455	0.8058	0.6729	0.5796
SpaceNet	BCE	0.9066	0.6086	0.7283	0.5738	<u>0.5677</u>
	BCE + FocalTversky	0.9129	0.5813	0.7103	0.5539	0.5359
	BCE + GapLoss	0.7865	0.6845	<u>0.7363</u>	0.5832	0.2177
	BCE + SAC-Loss (proposed)	0.8737	0.7175	0.7879	0.6552	0.5910

compared to the best score of BCE + GapLoss). This integration resulted in considerable improvement in connectivity and accuracy of the predicted graph indicated by a 0.17 points increase in the APLS. When integrated into SemSeg, the proposed SAC-Loss resulted in the highest scores for all three metrics, relaxed F1, Quality, and APLS, with values of 0.7879, 0.6552, and 0.5910 respectively (increase of 0.0516, 0.0720, 0.0233 points compared to the second best).

3.4.2 Comparison to Road Segmentation Networks

To further evaluate the performance of the proposed SAC-Loss function, we have compared its performance to recent state-of-the-art road detection approaches from [6] in terms of standard IoU metric. The proposed SAC-Loss function integrated into the SemSeg convolutional neural network (CNN) architecture outperforms other CNN-based

networks. It has a comparable IoU to Seg-Road-s as a more complex transformer-based architecture and with more similar model size (memory size required for transformer-based networks and SemSeg in Mb: SemSeg: 5.75, Seg-Road-s: 4.18, Seg-Road-m: 14.12, Seg-Road-l: 28.67) [6].

3.5. Qualitative Results

Figure 4 illustrates the impact of each loss function on the predicted road masks from the Massachusetts and DeepGlobe Road Datasets. The figure shows two sample representative patches from the test set of these datasets along with segmentation results using BCE, BCE + GapLoss, and the proposed BCE + SAC-Loss approaches.

In the first row of Figure 4, we can observe that the BCE loss resulted in over-detection, high numbers of false positives, and lower precision; addition of the GapLoss to the BCE loss decreased the number of false positives but re-

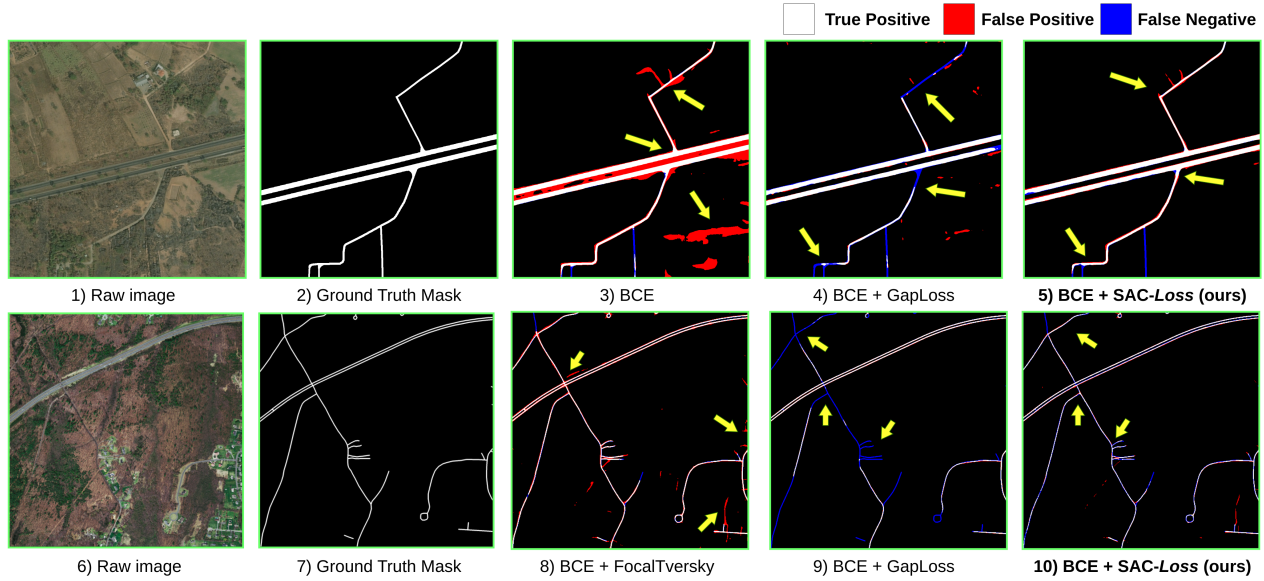


Figure 4. Sample predictions: 1-5) From Deepglobe dataset and SemSeg Network, 6-10) From Massachusetts dataset and UNet Network.

Table 2. State-of-the-art segmentation methods results on Massachusetts dataset [6].

Network type	Approach	IoU
CNN-based	SemSeg: BCE + SAC (Ours)	64.10
	D-LinkNet [23]	61.45
	DeepRoadMapper [2]	59.66
	RoadCNN [3]	62.54
	PSPNet [24]	58.91
	LinkNet34 [25]	61.35
	CoANet [4]	61.67
Transformer-based	Seg-Road-s [6]	64.78
	Seg-Road-m [6]	66.29
	Seg-Road-l [6]	68.38

sulted in under-detection, high number of false negatives, lower recall, and higher levels of fragmentation; whereas addition of the proposed SAC-Loss to the BCE loss resulted in accurate prediction the roads with a low number of false positives while still preserving connectivity.

In the second of Figure 4, the results from BCE + Focal Tversky loss have resulted in over-prediction and having more false positives. In contrast, the BCE + GapLoss has underpredictions with various fragmentation within the road network. However, the most significant improvement is observed when the BCE + SAC-Loss is used with having the most connected road network while minimizing the false positives, as expected from the quantitative results with the highest F1-score, Quality, and APLS metric.

4. Conclusion

In this paper, we proposed a novel adaptive, structure-aware, connectivity-preserving loss function designed to improve road segmentation in remote sensing images. This loss function combined a locally adaptive, structure-aware, weighting scheme based on gap evidence and proximity to road structures with Focal Tversky loss formulation that allows asymmetric penalties for false positives and false negatives to fill the gaps in the road structures while limiting spurious detections. When integrated into two segmentation networks, the proposed loss function improved the segmentation of thin curvilinear road structures that are prone to fragmentation. The resulting network reduced false positives while maintaining the topology and connectivity of the road networks, leading to more accurate road graphs that are critical for follow-up road network analysis tasks.

Acknowledgement

This material is based upon work supported in part by the U.S. Army Corps of Engineers, Engineering Research and Development Center—Information Technology Laboratory (ERDC-ITL) under Contract W912HZ23C0041. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U. S. Government or agency thereof. Computational resources for this research have been supported by the NSF National Research Platform, as part of GP-ENGINE (award OAC #2322218).

References

- [1] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 7, pp. 3523–3542, 2021. 1
- [2] G. Mátyus, W. Luo, and R. Urtasun, "Deeproadmapper: Extracting road topology from aerial images," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 3438–3446. 1, 8
- [3] F. Bastani, S. He, S. Abbar, M. Alizadeh, H. Balakrishnan, S. Chawla, S. Madden, and D. DeWitt, "Roadtracer: Automatic extraction of road networks from aerial images," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4720–4728. 1, 8
- [4] J. Mei, R.-J. Li, W. Gao, and M.-M. Cheng, "Coanet: Connectivity attention network for road extraction from satellite imagery," *IEEE Transactions on Image Processing*, vol. 30, pp. 8540–8552, 2021. 1, 8
- [5] J. Li, Y. Cai, Q. Li, M. Kou, and T. Zhang, "A review of remote sensing image segmentation by deep learning methods," *International Journal of Digital Earth*, vol. 17, no. 1, p. 2328827, 2024. 1
- [6] J. Tao, Z. Chen, Z. Sun, H. Guo, B. Leng, Z. Yu, Y. Wang, Z. He, X. Lei, and J. Yang, "Seg-road: a segmentation network for road extraction based on transformer and cnn with connectivity structures," *Remote Sensing*, vol. 15, no. 6, p. 1602, 2023. 2, 7, 8
- [7] V. Mnih, "Machine learning for aerial image labeling," Ph.D. dissertation, University of Toronto, 2013. 2, 5
- [8] H. Xu, H. He, Y. Zhang, L. Ma, and J. Li, "A comparative study of loss functions for road segmentation in remotely sensed road datasets," *International Journal of Applied Earth Observation and Geoinformation*, vol. 116, p. 103159, 2023. 2
- [9] M. Attari, N. P. Nguyen, K. Palaniappan, and F. Bunyak, "Multi-loss topology-aware deep learning network for segmentation of vessels in microscopy images," in *2023 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*. IEEE, 2023, pp. 1–7. 2
- [10] L. Nanni, S. Brahmam, and A. Loreggia, "An enhanced loss function for semantic road segmentation in remote sensing images," *IEEE Access*, 2024. 2
- [11] W. Yuan and W. Xu, "Gaploss: A loss function for semantic segmentation of roads in remote sensing images," *Remote Sensing*, vol. 14, no. 10, p. 2422, 2022. 2
- [12] A. Batra, S. Singh, G. Pang, S. Basu, C. Jawahar, and M. Paluri, "Improved road connectivity by joint learning of orientation and segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10385–10393. 2
- [13] W. G. C. Bandara, J. M. J. Valanarasu, and V. M. Patel, "Spin road mapper: Extracting roads from aerial images via spatial and interaction space graph reasoning for autonomous driving," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 343–350. 2
- [14] X. Li, Y. Wang, L. Zhang, S. Liu, J. Mei, and Y. Li, "Topology-enhanced urban road extraction via a geographic feature-enhanced network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 12, pp. 8819–8830, 2020. 2
- [15] X. Hu, F. Li, D. Samaras, and C. Chen, "Topology-preserving deep image segmentation," *Advances in neural information processing systems*, vol. 32, 2019. 2
- [16] S. Jadon, "A survey of loss functions for semantic segmentation," in *IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. IEEE, 2020, pp. 1–7. 2, 4
- [17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III*. Springer, 2015, pp. 234–241. 4, 6
- [18] H. Ghandorh, W. Boulila, S. Masood, A. Koubaa, F. Ahmed, and J. Ahmad, "Semantic segmentation and edge detection—approach to road detection in very high resolution satellite images," *Remote Sensing*, vol. 14, no. 3, p. 613, 2022. 4, 6
- [19] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raskar, "Deepglobe 2018: A challenge to parse the earth through satellite images," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 172–181. 5
- [20] A. Van Etten, D. Lindenbaum, and T. M. Bacastow, "Spacenet: A remote sensing dataset and challenge series," *arXiv preprint arXiv:1807.01232*, 2018. 5
- [21] C. Wiedemann, C. Heipke, H. Mayer, and O. Jamet, "Empirical evaluation of automatically extracted road axes," *Empirical evaluation techniques in computer vision*, vol. 12, pp. 172–187, 1998. 5, 6
- [22] P. K. Diederik, "Adam: A method for stochastic optimization," (*No Title*), 2014. 6
- [23] L. Zhou, C. Zhang, and M. Wu, "D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 182–186. 8
- [24] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890. 8
- [25] H. Yan, C. Zhang, J. Yang, M. Wu, and J. Chen, "Did-linknet: Polishing d-block with dense connection and iterative fusion for road extraction," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. IEEE, 2021, pp. 2186–2189. 8