Earth Parsing Through Satellite Images for Road and Street Network Identification

Riya Vardhan
Dept. of Electronics and Communication Engineering
MIT, Manipal
Udupi, India
riya.vardhan@learner.manipal.edu

Harshal Agarwal
Dept. of Electronics and Communication Engineering
MIT, Manipal
Udupi, India
harshal.agarwal@learner.manipal.edu

Mahima Mehta
Dept. of Electronics and Communication Engineering
MIT, Manipal
Udupi, India
mahima.mehta@learner.manipal.edu

***Abstract* – In the pursuit of enhancing road network detection from satellite imagery, we implemented semantic segmentation methodologies using the DeepGlobe Road Extraction dataset. Our approach entailed preprocessing the data with one-hot encoding for masks, followed by a training and validation split, and augmenting the training set to improve model robustness. We explored two prominent image segmentation architectures: DeepLabV3+ and U-Net. DeepLabV3+ with a ResNet50 encoder was utilized for its advanced encoder-decoder structure, adept at capturing significant features and delineating road boundaries. Alternatively, a U-Net model was constructed, emphasizing a symmetrical encoder-decoder design. Both models underwent training with suitable parameters, loss functions, and learning rate schedules. Model performance was evaluated on the validation set, leveraging metrics such as the IoU score and Dice loss. The IoU score was pivotal for assessing the accuracy of the predicted road network against the ground truth, and Dice loss was critical for fine-tuning the model's precision in overlaying road masks. The results, including normal accuracy metrics and binary cross entropy loss, were visualized to compare the original satellite imagery, ground truth masks, and predicted segmentation, highlighting the effectiveness of the applied methodologies in road network segmentation tasks.**

Keywords – road network, image segmentation, U-Net, DeepLabV3+, IoU score, Dice loss

I. Introduction

The proliferation of satellite imagery in recent years has opened up new avenues for large-scale, detailed mapping of the Earth's surface. One area of particular interest is the extraction of road networks, which is crucial for a variety of applications, including urban planning, traffic management, and autonomous navigation. The challenge, however, lies in accurately detecting and segmenting roads from highly complex and variable satellite images—a task that has traditionally required extensive manual effort.

Several previous papers and research works have attempted to address road extraction from satellite images. These approaches include traditional computer vision techniques, such as edge detection and thresholding, as well as machine learning methods, like random forests and support vector machines. In recent years, deep learning models have gained prominence for their ability to capture complex patterns in satellite imagery. Notable architectures include U-Net, DeepLab, and Mask R-CNN.

Despite their merits, many existing methodologies face limitations. Traditional methods often struggle to capture complex road layouts and are sensitive to imaging conditions. Some deep learning models may be computationally demanding and require large amounts of annotated data for training. Additionally, while these approaches perform well in controlled settings, they may not generalize effectively to diverse geographical regions or real-world conditions.

This work aims to overcome the limitations of existing methodologies by conducting a comprehensive comparative analysis of state-of-the-art deep learning models, U-Net and DeepLabV3+, utilizing the DeepGlobe Road Extraction Challenge Dataset. By leveraging these models, we aim to address the limitations through their specific strengths. U-Net's simplicity and efficiency make it a strong candidate for straightforward road extraction, while DeepLabV3+'s attention to contextual information may improve performance in complex scenarios. The use of advanced data augmentation and preprocessing techniques will also contribute to more robust and generalized models.

The primary objective of this work is to evaluate the performance of U-Net and DeepLabV3+ models and choose the best approach in the task of road extraction from satellite images using the DeepGlobe Road Extraction Challenge Dataset. This includes assessing their accuracy, efficiency, robustness to varying imaging conditions, and generalization capabilities. By addressing the limitations of existing methodologies, our goal is to contribute to the development of more accurate and practical road extraction solutions that can benefit a wide range of applications and stakeholders.

II. Dataset

DeepGlobe utilizes satellite imagery sourced from the DigitalGlobe +Vivid Images dataset. This collection features images taken over Thailand, Indonesia, and India. The granularity of these images stands at 50 cm per pixel, encompassing three channels: Red, Green, and Blue. The original geotiff images measure a vast 190584 × 190584 pixels. The annotation of these images begins by segmenting them into tiles and then importing them into QGIS. Based on this segmentation, specific regions of interest are identified from the mentioned countries. The process involves an even-handed selection between urban and rural landscapes. After these regions are chosen, the corresponding DigitalGlobe tiff images representing these areas are selected.

The images are segmented to isolate valuable sections, and GIS specialists select pertinent segments for further analysis. In choosing these segments, the aim is to evenly represent varied areas of interest, such as regions with diverse kinds of roadways (like unpaved, paved, and dirt roads) and settings that range from rural to urban.

Fig 1. Road labels are annotated on top of the satellite image patches, all taken from DeepGlobe Road Extraction Dataset.

III. Methodology

We have approached this problem through the process of semantic segmentation.

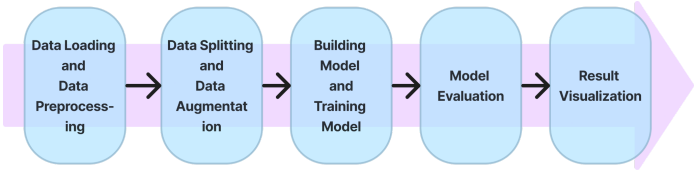


Fig 2. A block diagram to illustrate the methodology applied to approach the problem.

A. Data Loading and Data Preprocessing

The metadata of the DeepGlobe Road Extraction dataset, including the paths to the corresponding satellite images and their corresponding mask images. The data frame is created using this information and sampled randomly.

The data is then preprocessed by one-hot encoding the masks to convert them into a suitable format for semantic segmentation tasks.

B. Data Splitting and Data Augmentation

The dataset is split into training and validation split with a 90/10 split ratio.

Data augmentation is hence applied to the training dataset. Common augmentation techniques like horizontal flip and vertical flip are used to increase the diversity of the training data.

C. Data Building and Training Model

This task involves identifying and segmenting road networks within satellite images. Semantic segmentation is the right technique for this task, as it assigns a class label to each pixel in the image, effectively outlining the road network.

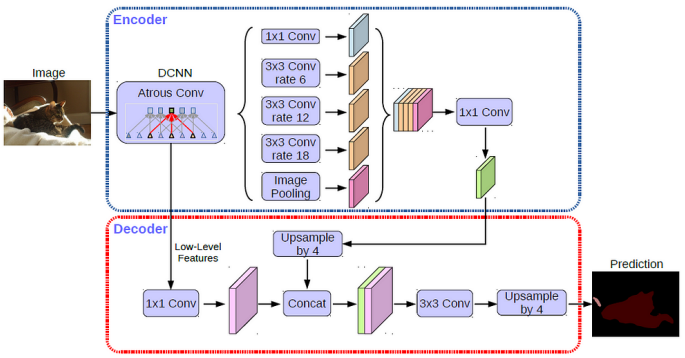


Fig 3. The architecture of the DeepLabV3+ model with an encoder and decoder block for image segmentation problems.

One of the ways of approaching this problem can be by harnessing the power of the existing pre-trained models for the image segmentation task. DeepLabV3+ is an extension of the DeepLab family of models, known for their effectiveness in semantic segmentation tasks. The "V3+" version incorporates several key improvements that will make it well-suited for this task. DeepLabV3+ employs an encoder-decoder architecture. The encoder, often based on pre-trained models like ResNet or MobileNet, captures high-level features from input images. The decoder refines the segmentation output and improves the localization of object boundaries. Our model leverages transfer learning by using pre-trained encoders on the DeepLabV3+ model initialized. The model uses ResNet50 as the encoder. ResNet models are well-known for their excellent feature extraction capabilities. The ResNet50 encoder helps the model in capturing meaningful features from satellite images, which are important for accurately identifying road networks.

Another way of approaching this problem is through a U-Net. We define the U-Net model consisting of an encoder path, which downsamples the input image, and a decoder path, which upsamples the feature maps to generate the segmentation mask. Each encoder block performs two 2D convolutions with batch normalization and ReLU activation. Max-pooling layers are used for downsampling, and dropout is applied after each max-pooling operation. Decoder blocks perform 2D transposed convolutions to upsample the feature maps and concatenate them with the corresponding encoder features. Then, two 2D convolutions are applied. The output of the model is a single-channel image with sigmoid activation, representing the predicted road mask.

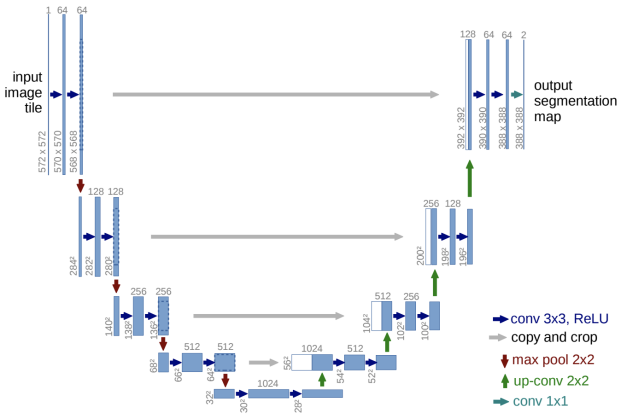


Fig 4. The U-Net architecture (example for 32x32 pixels in the lowest resolution) model consists of an encoder-decoder cascade structure for image segmentation problems.

Once the model is initialized, we set-up the apt training parameters including loss function, optimizers and learning rare scheduler. We compile and train the model on the training dataset and evaluate its performance on the validation dataset. The validation performance obtained during training, the corresponding loss and metrics and recorded for each epoch.

D. Model Evaluation

Once the training is complete, the best model is loaded and the evaluation is performed on the validation dataset. Evaluation metrics such as IoU score and Dice loss are calculated and reported.

The Intersection over Union (IoU) score is considered as an appropriate evaluation metric for the given model as it assesses how the predicted road mask aligns with the ground truth mask. A higher IoU indicates better agreement between the predicted and actual road regions. The IoU score is calculated as:

$$Intersection\ over\ Union\ (IoU) = \frac{|A \cap B|}{|A| \cup |B|} \quad (1)$$

Here, A represents the area of the ground truth mask and B represents the area of the predicted road mask.

Dice Loss is used as a loss metric for the given model as it encourages the model to produce accurate, well-overlapping road masks, which align with the goals of the segmentation task. It's particularly effective for capturing small road segments and is well-suited to tasks where the quality of the segmentation matters, as it quantifies the spatial agreement between the predicted and ground truth regions. Dice Loss can be computed as follows:

$$Dice\ Loss,\ DL\ (y, \hat{y}) = 1 - \frac{2y\hat{y} + 1}{y + \hat{y} + 1} \quad (2)$$

Here, $\hat{y}$ is the predicted road mask and y is the ground truth mask.

In the case of U-Net we use convolutional neural networks to build the model, hence, our evaluation metrics in the case of U-Net are in terms of normal model accuracy calculations and binary cross entropy loss including IoU score and Dice loss..

E. Result Visualization

Finally, the code randomly selects an image from the validation dataset and visualizes the original image, ground truth mask, and one-hot encoded mask for visualization and analysis.

IV. Post–Processing

In order to better visualise the predicted output masks, we post-process the images by converting them into heat maps. Heat maps are good visualisation tools in order to bring out the routes that have been faintly identified by the model in the predicted output mask. A ground truth mask does not bring out the faint road networks as appropriately as the heat maps.

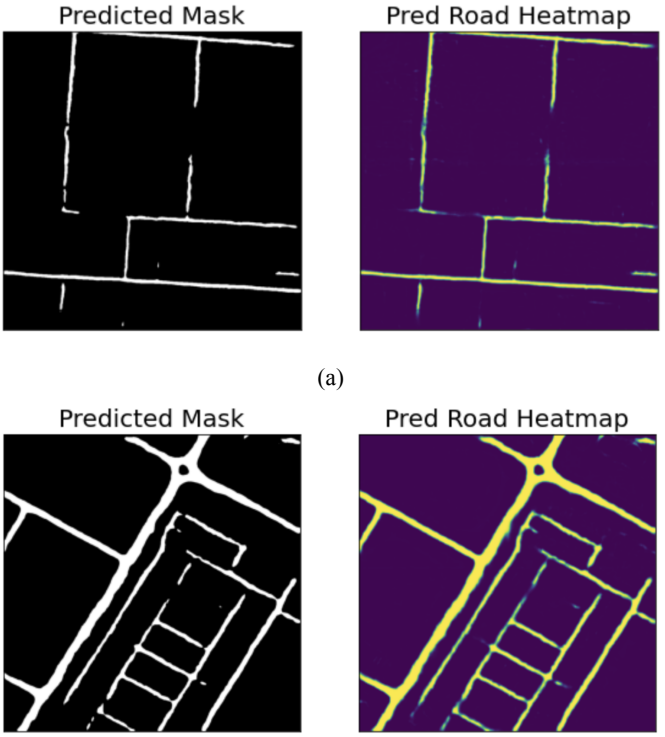


Fig 5. Post-processing the predicted ground truth masks to heat-maps to bring out the faintly marked road networks in the predicted image mask in (a) and (b).

V. Results

The output results of both the DeepLabV3+ model and the U-Net model were visualised using the original image, the ground truth mask and the predicted output mask and heatmaps.

Figure 6 shows the predicted output masks of 3 images in the road network identification dataset using the DeepLabV3+ model.

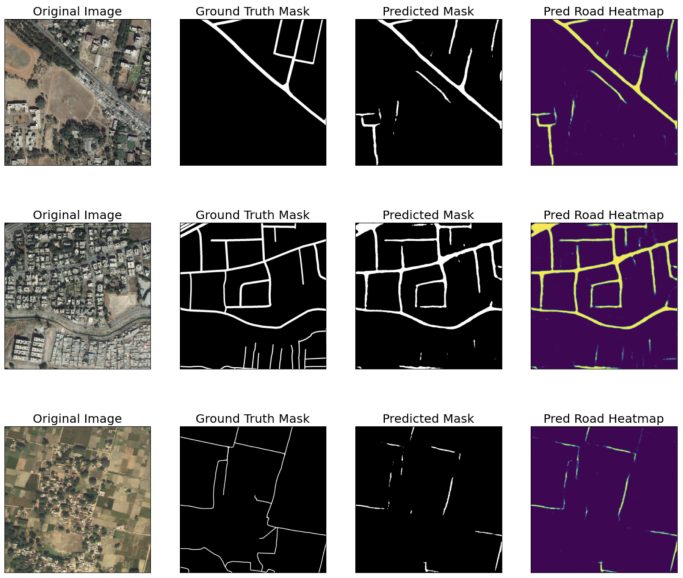


Fig 6. Predicted output masks by the DeepLabV3+ model

Figure 7 shows the predicted output masks of 3 images in the road network identification dataset using the U-Net model.

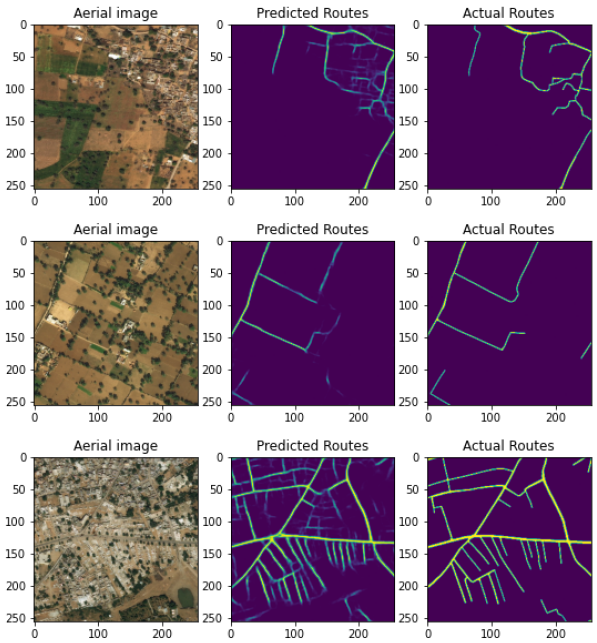


Fig 7. Predicted output masks by the U-Net model.

Table 1 is a comparative study, where we delve into the efficiency and effectiveness of both models. One of the key aspects evaluated in Table 1 is the rate of convergence, represented by the number of training epochs needed for each model to achieve the desired accuracy. This metric provides insights into the training efficiency of the models, which is a crucial factor for practical deployment.

Furthermore, the table presents a detailed examination of training and validation accuracies. Specifically, it employs the Intersection over Union (IoU) score. The IoU score quantifies the degree of overlap between the predicted road masks and the ground truth annotations, providing a robust measure of the model's performance.

Table 1. IoU scores and Dice Losses for the two models for a certain number of epochs on the training and validation set

Models Used			DeepLabV3+	U-Net
Result after 3 epochs	Training Set	IoU Score	0.9473	0.3032
		Dice Loss	0.03212	0.8260
	Validation Set	IoU Score	0.9471	0.3030
		Dice Loss	0.03206	0.8353
Result after 100 epochs	Training Set	IoU Score	–	0.7173
		Dice Loss	–	0.3410
	Validation Set	IoU Score	–	0.6953
		Dice Loss	–	0.4667

The DeepLabV3+ performance is superior to U-Net in multiple ways. The number of epochs to reach the high level of accuracy on both training and validation data is much lesser in the DeepLabV3+ model approach compared to U-Net. DeepLabV3+ gives higher accuracy in fewer epochs, on the other hand, U-Net takes more number epochs to reach a decent level of accuracy on the model. In terms of IoU scores and Dice loss, DeepLabV3+ is the better model.

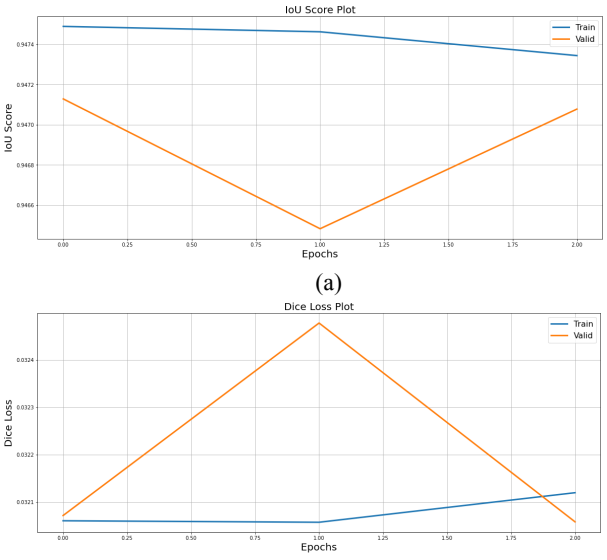
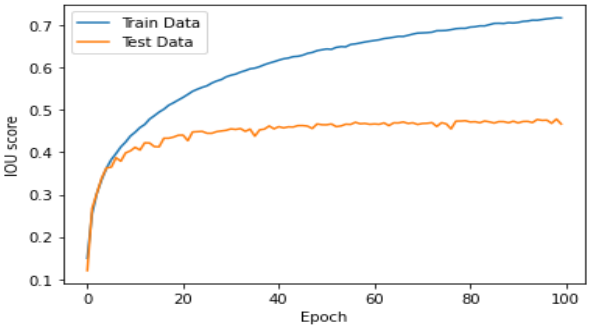
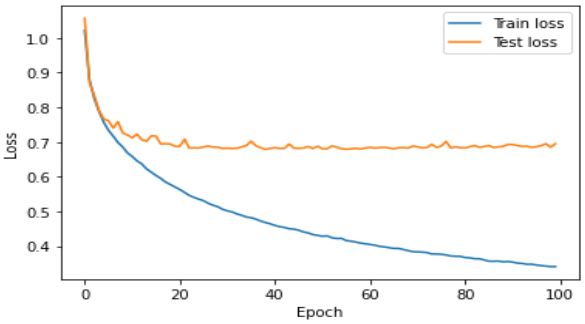


Fig 8. The training and validation plots of (a) IoU scores of DeepLabV3+ model and (b) Dice Losses of DeepLabV3+ model.



(a)



(b)

Fig 9. The training and validation plots of (a) IoU scores of U-Net model and (b) Dice Losses of U-Net model.

VI. Conclusions

The project addresses the challenge of road network segmentation in satellite images through semantic segmentation. Two prominent architectures, DeepLabV3+ and U-Net, were employed to tackle the task. DeepLabV3+ utilizes an encoder-decoder structure with a ResNet50 backbone, capitalizing on transfer learning to enhance feature extraction for precise road identification. Meanwhile, U-Net's architecture consists of symmetric encoder and decoder paths with convolutional and max-pooling layers, renowned for its efficiency in medical image segmentation tasks.

The models were trained and model performance was gauged using metrics such as the Intersection over Union (IoU) score and Dice loss, ensuring accurate and well-overlapping segmentation of road networks. Upon training completion, the best-performing model was selected based on these metrics. Finally, for a comprehensive evaluation, a random image from the validation set was chosen to demonstrate the original image alongside its ground truth and predicted segmentation masks, offering a clear visual assessment of the model's performance.

In conclusion, the project effectively demonstrates the application of advanced neural network architectures for the semantic segmentation of roads in satellite imagery, with promising results that could be significant for geographical information systems and urban planning.
The methodology contributes to the United Nations Sustainable Development Goal 11: Sustainable Cities and Communities. By enhancing the accuracy and precision of road network segmentation in satellite imagery, this research aids in the better planning and management of urban infrastructure, leading to more sustainable and resilient cities. Accurate maps are crucial for urban planning, traffic management, and emergency response services, all of which are essential components of sustainable urban ecosystems.

To expand this work further and align it with the UN's sustainable development agenda, future efforts could focus on developing interfaces or plugins that allow this segmentation method to be integrated directly into urban planning and geographic information system (GIS) software, improving the models to work with real-time data, supporting dynamic urban management and immediate responses to changes in road conditions, traffic patterns, and urban growth, combining satellite imagery with other data sources such as traffic flow sensors, social media data, and demographic information to create comprehensive models for urban development and ensuring that the technology is accessible to cities in developing countries, which might benefit the most from advanced planning tools due to their rapid urbanization.

Team Member Contributions

Riya Vardhan – Investigation, Models and Methodologies, Code, Report writing

Mahima Mehta – Investigation, Methodology, Result Analysis, Report Writing

Harshal Agarwal – Investigation, Result Analysis, Report Writing

References

[1] *Competition*. CodaLab. (n.d.). https://competitions.codalab.org/competitions/18467

[2] Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., & Raskar, R. (2018, May 17). *DeepGlobe 2018: A Challenge to parse the Earth through satellite images*. arXiv.org. https://arxiv.org/abs/1805.06561

[3] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018, August 22). *Encoder-decoder with atrous separable convolution for Semantic Image segmentation*. arXiv.org. https://arxiv.org/abs/1802.02611

[4] Wang, H., Yu, F., Xie, J., & Zheng, H. (2022, October 27). *Road extraction based on improved Deeplabv3 plus in remote sensing image*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. https://isprs-archives.copernicus.org/articles/XLVIII-3-W2-2022/67/2022/

[5] S. Das, A. A. Fime, N. Siddique and M. M. A. Hashem, "Estimation of Road Boundary for Intelligent Vehicles Based on DeepLabV3+ Architecture," in IEEE Access, vol. 9, pp. 121060-121075, 2021, doi: 10.1109/ACCESS.2021.3107353.

[6] Y. SU, Y. Lin, X. Fang and L. Zhong, "Improved DeepLabV3+ Network Segmentation Method for Urban Road Scenes," 2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 2022, pp. 1274-1280, doi: 10.1109/ITAIC54216.2022.9836922.

[7] N. Y. Q. Abderrahim, S. Abderrahim and A. Rida, "Road Segmentation using U-Net architecture," 2020 IEEE International conference of Moroccan Geomatics (Morgeo), Casablanca, Morocco, 2020, pp. 1-4, doi: 10.1109/Morgeo49228.2020.9121887.

[8] OZTURK, O., SARITÜRK, B., & SEKER, D. Z. (2020, December 6). *Comparison of fully convolutional networks (FCN) and U-Net for road segmentation from high resolution imageries*. International Journal of Environment and Geoinformatics. https://dergipark.org.tr/en/pub/ijegeo/issue/56780/737993#article_cite

[9] Hofesmann, E. (2021, March 1). *IOU A Better Detection Evaluation Metric*. Medium. https://towardsdatascience.com/iou-a-better-detection-evaluation-metric-45a511185be1

[10] *Papers with code - dice loss explained*. Explained | Papers With Code. (n.d.). https://paperswithcode.com/method/dice-loss#:~:text=Introduced%20by%20Sudre%20et%20al,y%20%2B%20p%20%C2%AF%20%2B%201%20)