Boundaries-Informed Transformer for Terrain Segmentation with Enhanced Contour Detection: An exploration Study

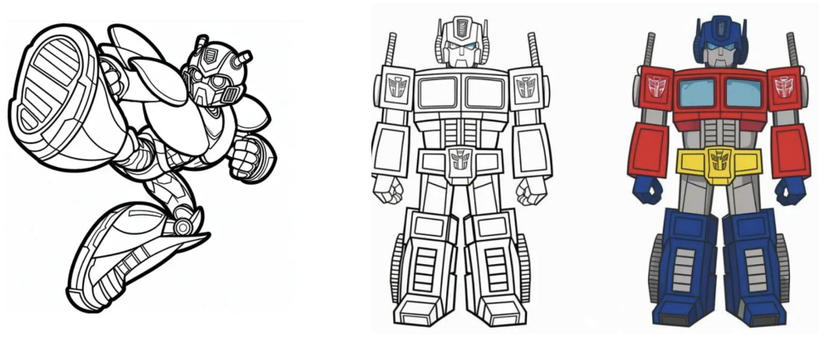
Yuchuan Dong, Zinan Ma

Chalmers University of Technology
Electrical Engineering Department

CHALMERS UNIVERSITY OF TECHNOLOGY

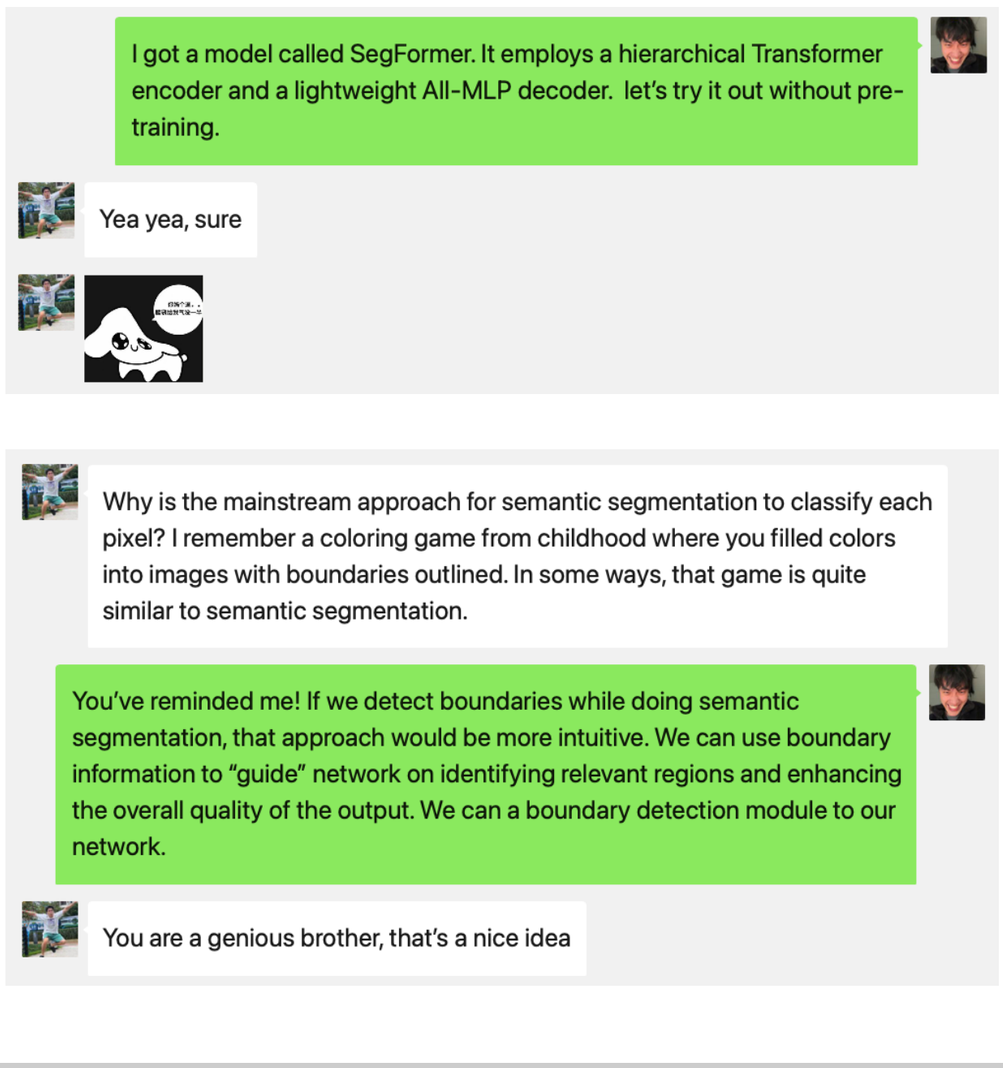
Overview

Off-road terrain detection is key for chassis tuning and autonomous navigation but is challenging due to the unstructured and varied terrain. In this study, we proposed Boundaries-Informed Transformer network.



It increased accuracy by 5% comparing to a pure SegFormer model. This work has not reached a satisfying result yet. However, it inspired us further exploration ideas for this newborn model.

A Theater of How This Idea Comes Up (Draw Coloring Books?)



Method

- SegFormer architecture: a **hierarchical** Transformer encoder with a **lightweight** All-MLP decoder.
- **Boundary Detection Module**
  - Boundary MLP

$$\text{boundary\_out} = W_3 * \text{ReLU}\left(W_2 * \text{ReLU}\left(W_1 * F + b_1\right) + b_2\right) + b_3$$

  - Fusion Layer

$$C = \text{Concat}(\text{seg\_out}, \text{boundary\_out})$$

$$F_{\text{final}} = \text{ReLU}(W_f * C + b_f)$$

$$\text{final\_out} = W_{\text{seg}} * F_{\text{final}} + b_{\text{seg}}$$

- **Boundary-Informed Segmentation Loss**
  - Segmentation Loss

$$L_{\text{seg}} = \text{WeightedCrossEntropy}\left(P_{\text{seg}}, Y_{\text{seg}}\right)$$

  - Boundary Loss

$$L_{\text{boundary}} = \text{WeightedBinaryCrossEntropy}\left(P_{\text{boundary}}, Y_{\text{boundary}}\right)$$

  - Boundary-Informed Segmentation Loss

$$L_{\text{total}} = L_{\text{seg}} + \lambda * L_{\text{boundary}}$$

Structure of the network

- The network supports different MiT backbones(B0-B4) across encode stages(C1-C4) with varying parameter in figure 1.
- Proposed network fuse SegFormer and boundary-informed layer by contacting the two outputs and further processing with a MLP layer.

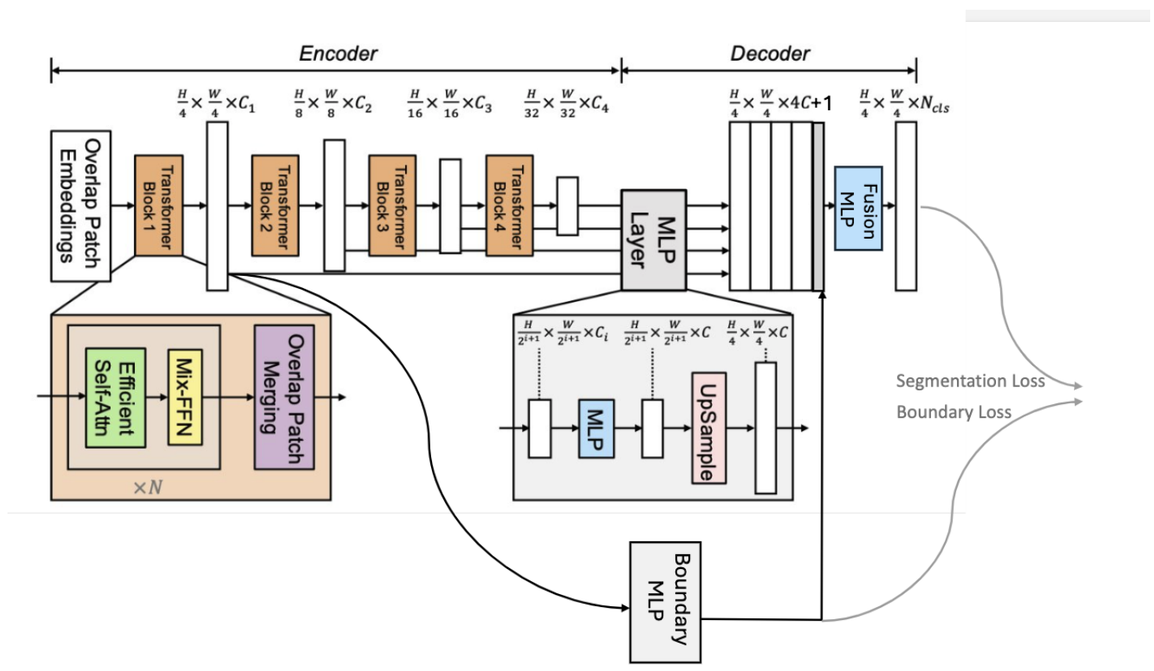


Figure 1: Structure of the proposed network improved from SegFormer [1]

Evaluation

Figure 2 presents three test cases, each illustrating the original image, ground truth, boundary detection, SegFormer segmentation result with boundary information (using model Mit_B0BI), and SegFormer segmentation result (using model Mit_B0). Two models are generated using identical input resolution, optimizer, loss functions, and iterations.

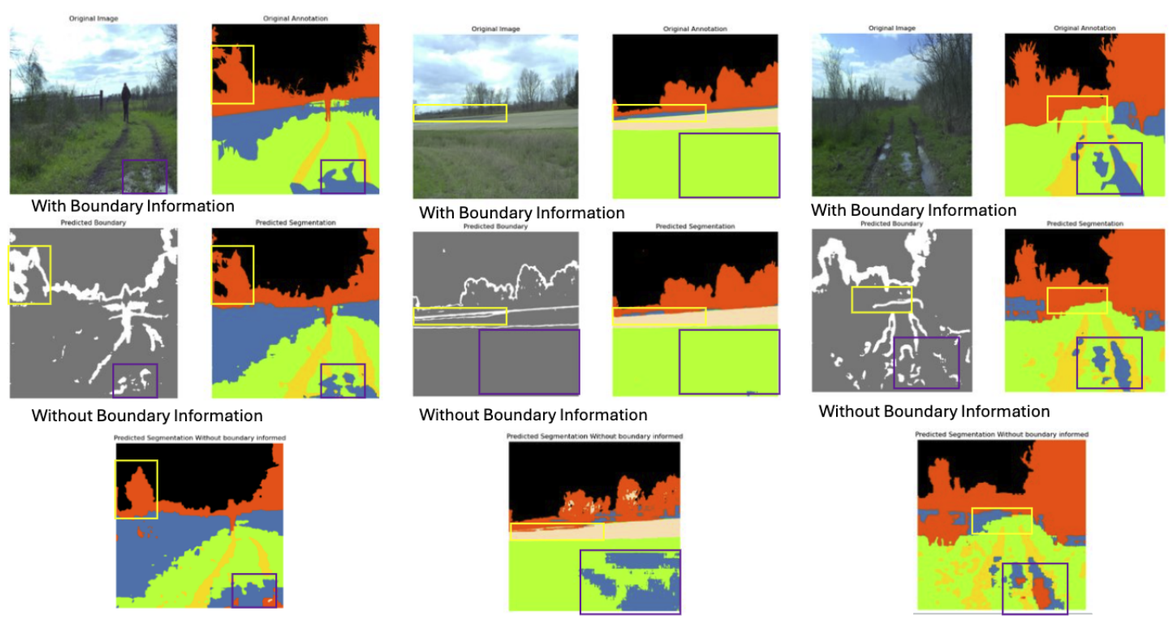


Figure 2: The model is tested on three test dataset instances.

- Adding boundary-informed improves MIoU across all models.
- Boundary-informed model use less parameters to achieve better performance.

Model	Params	BG	Smooth	Rough	Bumpy	Forbidden	Obstacle	MIoU
Mit_B0	7.7M	92.7	57.6	72.6	10.5	47.9	51.5	55.4
Mit_B0BI	8.4M	93.1	57.8	78.9	18.0	54.8	59.7	60.3
Mit_B1	30.7M	93.3	63.2	81.4	16.1	52.8	49.7	59.5
Mit_B1BI	33.4M	94.2	62.4	82.0	20.7	60.8	62.7	63.8

Table 1: Model performance comparison.(Notice that Mit_B*BI stand for Mit_B*Boundary-Informed model, BG stand for background.)

Conclusion and inspiration

Preliminary testing shows that our method has improved SegForm's MIoU performance. The fusion of boundary detection with SegForm, along with its initial implementation, has given us insights and posed future challenges, including:

- How to effectively fuse the SegForm network with boundary detection.
- Suitable boundary detection-based semantic segmentation schemes for fusion.
- In-depth research from the perspectives of effectiveness and interpretability.

References

[1] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo.
Segformer: Simple and efficient design for semantic segmentation with transformers.
Advances in neural information processing systems, 34:12077–12090, 2021.

yuchuan@student.chalmers.se zinanm@student.chalmers.se