Application of Semantic Segmentation

Student ID: 202018010119

Tutor: Dr Grace U. Nneji

Student: Noah

Contents

CHENGDU UNIVERSITY OF TECHNOLOGY
OXFORD BROOKES UNIVERSITY

Background

- Semantic segmentation technology has become key to self-driving by recognizing categories such as roads, vehicles and pedestrians. This provides accurate environment perception to self-driving vehicles, helping them to handle the complexity of urban roads, thus significantly reducing traffic congestion and accident rates.

Challenges

- Variable weather
- Light conditions
- High computational resources for existing models
- High real-time requirements

Solutions

- Building new semantic segmentation architectures
    - Convolution-based ASPP Module
    - Transformer Encoder Module
    - Edge Detection Moudle
    - Context Enhancement Module
    - ResNet 50
- Building GUIs to Enhance Interactability
    - Image segmentation
    - Video Segmentation
    - Real-time Segmentation
    - Other Scene Segmentation

Dataset

Cityscapes Dataset

What is cityscapes:

- Focuses on urban streetscape (roads, pedestrians, buildings, etc.)
- Images and their categories labeled
- 34 categories in total, 19 as semantic segmentation

How I use it:

- 2975 training data, 500 validation, 1525 testing
- Converted the dataset into 19, 15, and 11 categories for multiple dataset testing
- Overlay 3-channel images with single-channel annotations for random cropping
- Perform data enhancement (weather simulation, random saturation, random hue, etc.)
- Create dataset (Batch 24)

Figure1.Weather simulation

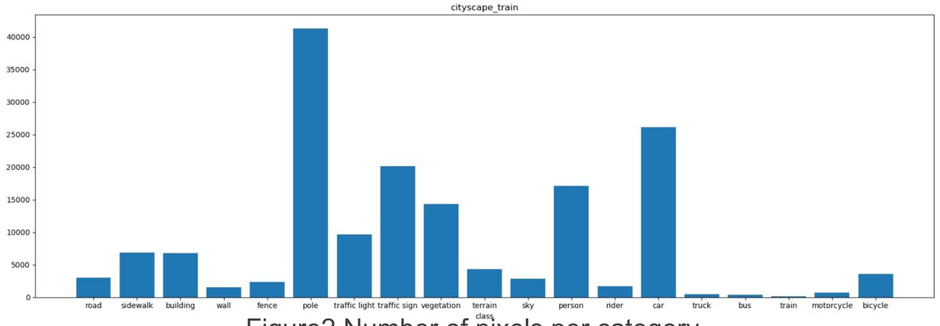

Figure2.Number of pixels per category

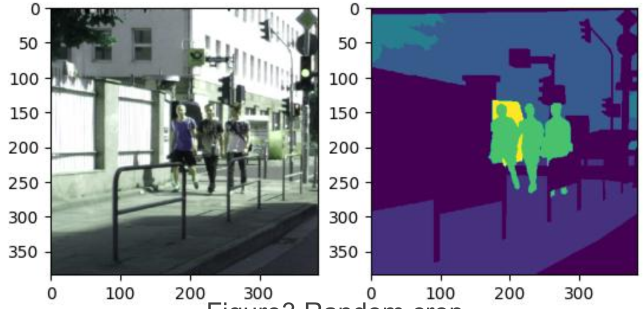

Figure3.Random crop

Result

Architecture of the proposed model

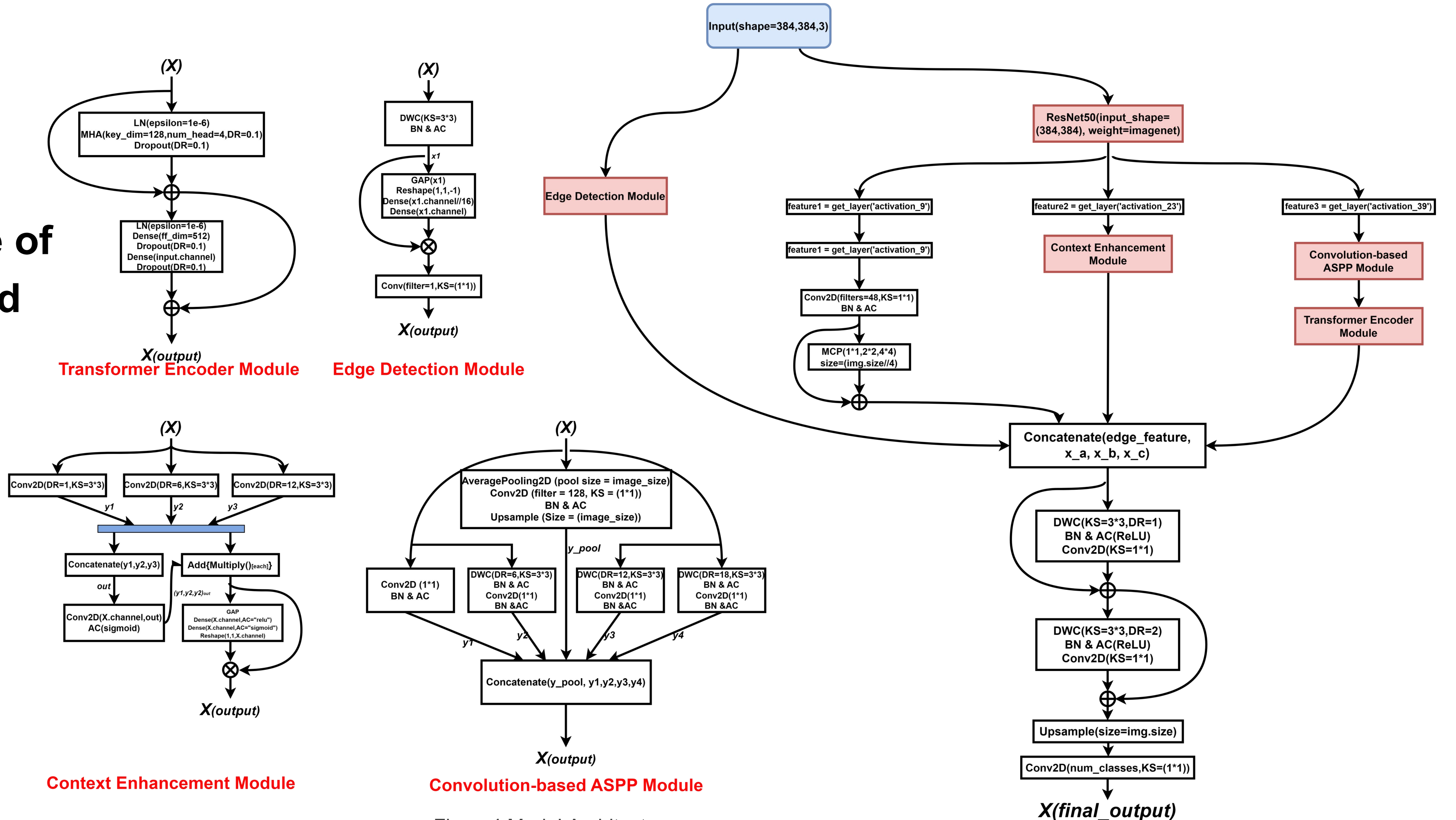

Figure4.Model Architecture

Result

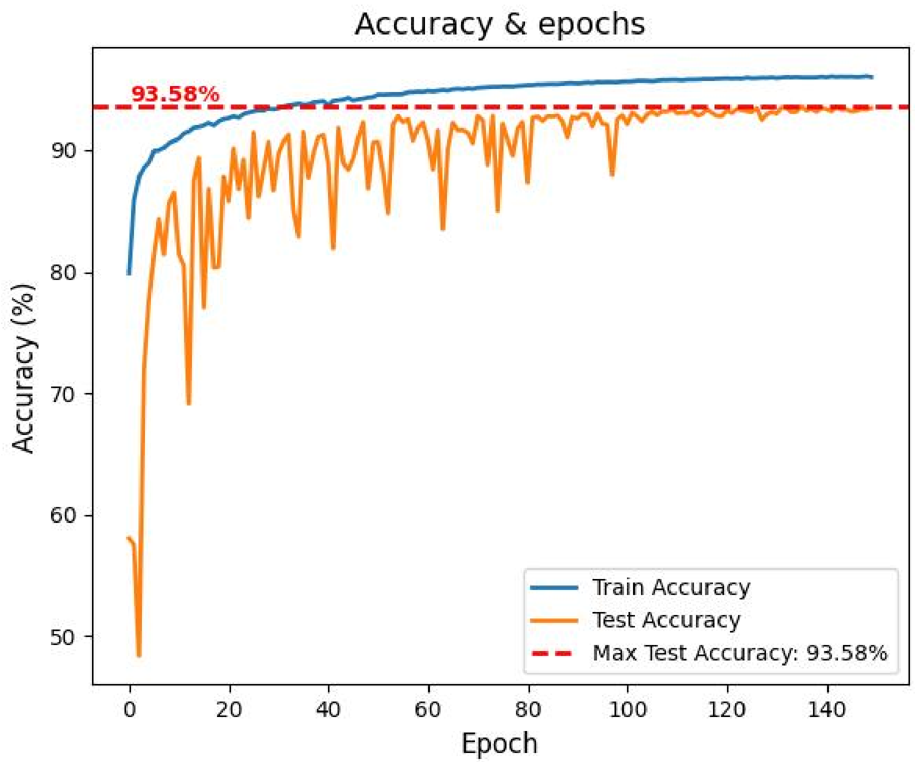

Figure5.Acc = 93.58%

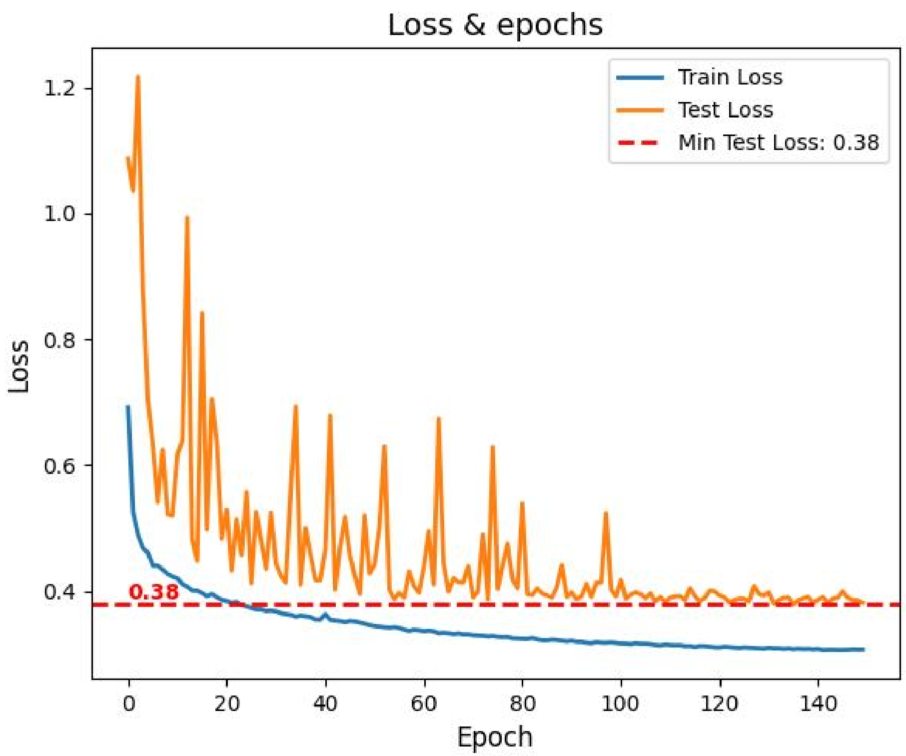

Figure6.Loss = 0.38

Result

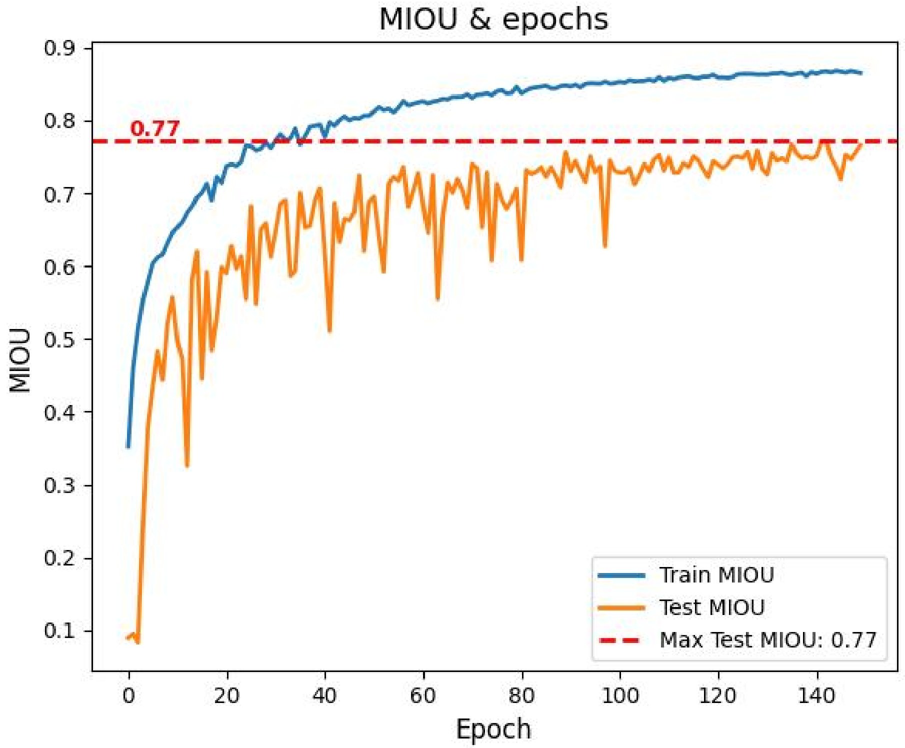

Figure7.MIOU = 0.77

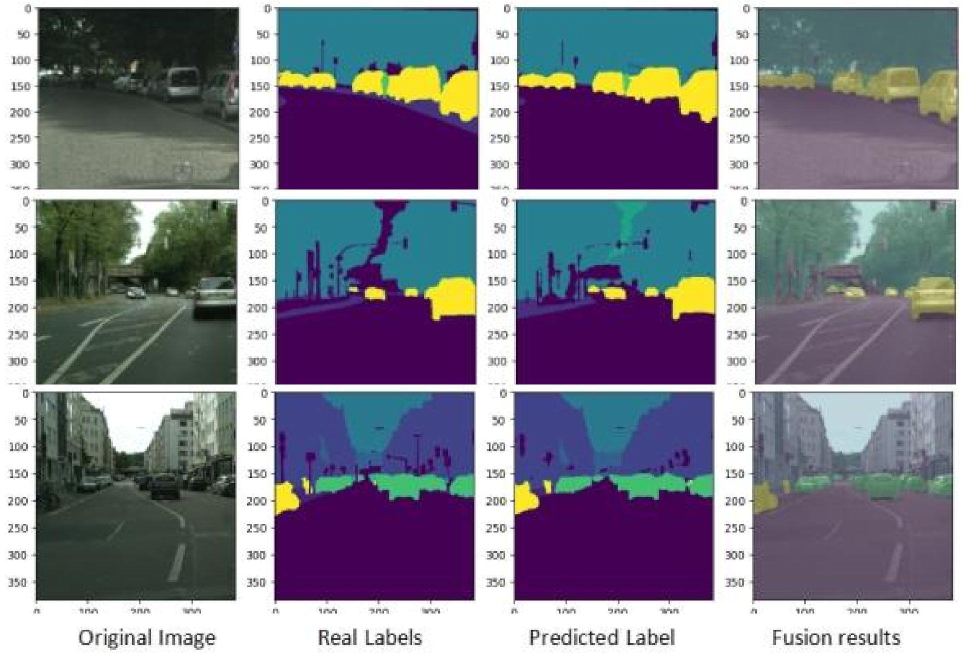

Figure8.Segmentation results

Result

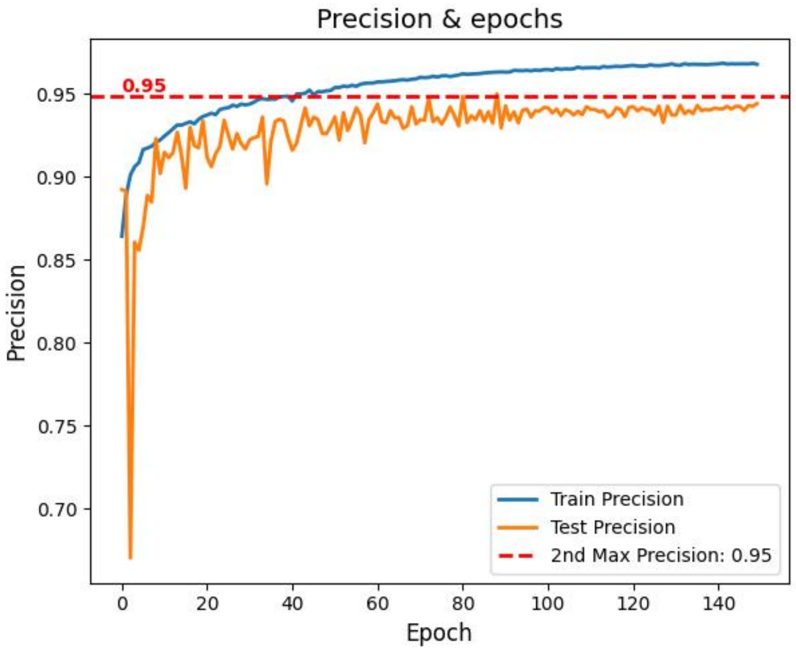

Figure9.Precision = 0.95

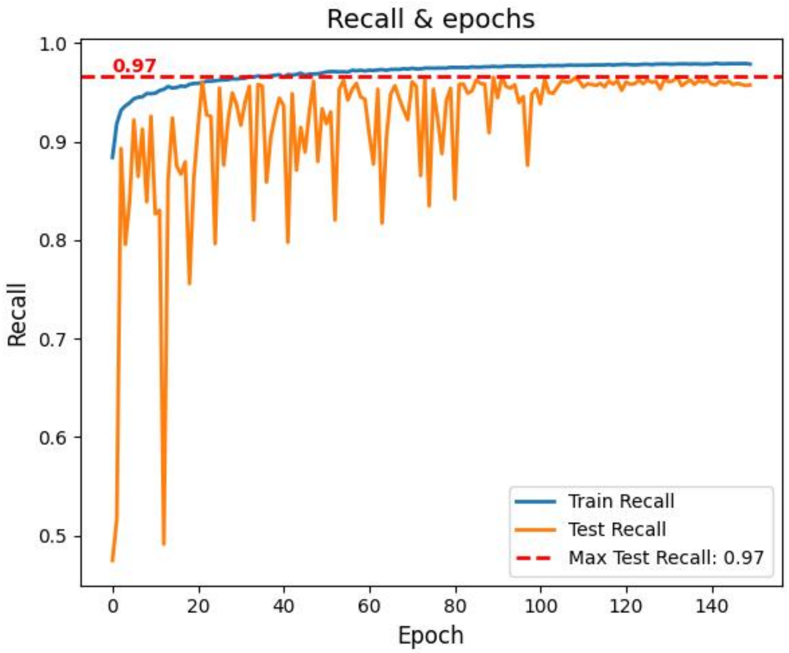

Figure10.Recall = 0.97

PA	CPA	Dice Coefficient	Kappa	FPS
0.9316	0.8510	0.8529	0.8972	63.28

Table1.Other Metrics

Comparison

No. of categories	Accuracy	Loss	Precision	Recall	MIOU	PA	CPA	Dice Coefficient	Kappa	FPS
34 categories	91.77%	0.55	1	1	0.54	0.9516	0.6315	0.6510	0.8902	150(A6000)
19categories	92.60%	0.45	0.97	0.99	0.71	0.9307	0.7188	0.7413	0.9037	149(A6000)
15 categories	91.91%	0.44	0.97	0.98	0.70	0.921	0.8114	0.8193	0.8931	148(A6000)
11 categories	93.58%	0.38	0.95	0.97	0.77	0.9316	0.8510	0.8529	0.8972	63.28(3090Ti)

Table2.Comparison of the performance of different categories

Comparison

Model	Accuracy	Loss	Precision	Recall	MIOU	PA	CPA	Dice Coefficient	Kappa	FPS
LinkNet	78.82	0.70	0.91	0.95	0.36	0.68	0.30	0.3084	0.49	61.8
UNet	89.21	0.50	0.91	0.96	0.60	0.89	0.67	0.6732	0.83	33.5
proposed model	93.58	0.38	0.95	0.97	0.76	0.9316	0.851	0.8529	0.8972	63

Table3.Comparison of Common Models

Comparison

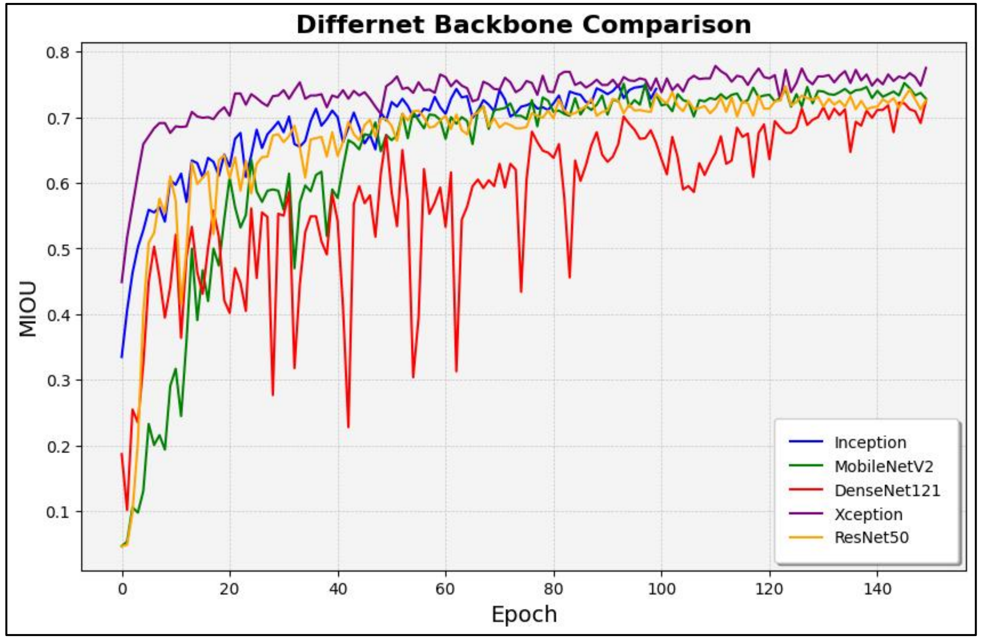

Figure11.Different backbone comparison

Backbone	Accuracy	Loss	Precision	Recall	MIOU	PA	CPA	Dice Coefficient	Kappa	FPS	Weight
InceptionV3	92.91%	0.38	0.94	0.96	0.73	0.9276	0.8386	0.8422	0.8934	62.67	65.4MB
Xception	93.81%	0.37	0.95	0.97	0.73	0.9352	0.8592	0.8567	0.9078	63.28	122.4MB
DenseNet121	92.62%	0.33	0.95	0.97	0.72	0.9233	0.8377	0.8342	0.8902	62.23	44.1MB
MobileNetV2	93.23%	0.38	0.96	0.97	0.72	0.9292	0.8278	0.8294	0.9875	63.96	30.8MB
Proposed model(ResNet50)	93.58%	0.38	0.95	0.97	0.77	0.9316	0.8510	0.8529	0.8972	62.14	60.1MB

Table4.Comparison of Other backbones models

Comparison

Author	Model	Mlou	Accuracy	Recall	FPS	Para(M)
Badrinarayananet.al.**[1]**	SegNet	56.1%	*	*	*	29.46
Abdigapporovet.al.**[2]**	BiFPN	56.4%	89.6	79.8	65.7	*
Paszkeet.al.**[3]**	ENet	58.3%	*	*	46.8	0.4
Poudelet.al.**[4]**	Fast-SCNN	68%	83.5	*	123.5	1.11
Yu et al **[5]**	BiSeNet	69%	65.5	*	65.5	14.1
Fourure et al**[6]**	GridNet	69.5%	*	*	*	*
Chen et al**[7]**	Deep-Lab CRF	70.4%	*	*	*	15.2
Lin et al **[8]**	RefineNet	73.6%	80.6	*	*	*
Li et.al.**[9]**	BiAttnNet	74.7%	*	*	89.2	2.2
My Model	proposed model	76%	93.58	0.97	63	10

Table5.Comparison of Other models in literature

Comparison

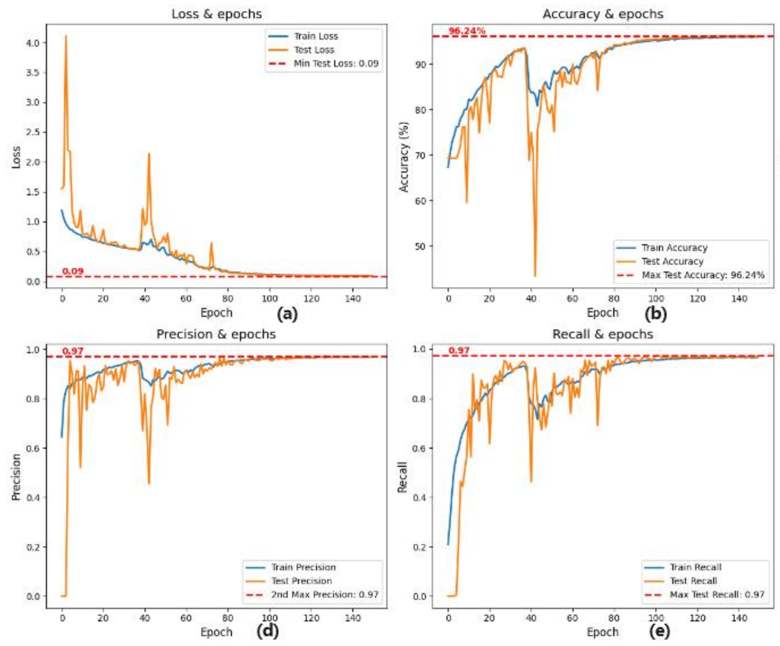

Table12.(a)Loss=0.09;(b)Acc=96.24;(c)MIOU=0.88;(d)Precision=0.97;(e)Recall=0.97

Other dataset train

CLASS	Person	Aeroplane	Bicycle	Bird	Boat	Bottle	Bus	Car	Cat	Chair	Cow	Dinning_Table
MIOU	0.97	0.89	0.53	0.88	0.88	0.89	0.96	0.94	0.94	0.81	0.94	0.95

CLASS	Dog	Horse	Motorbike	Potted_plant	Sheep	Sofa	Train	TV/Monitor	Void/Edges	Unlabeled/BG	Overall
MIOU	0.94	0.91	0.87	0.88	0.41	0.94	0.95	0.95	0.96	0.48	0.86

Table6.Each class IOU with VOC 2012 Dataset

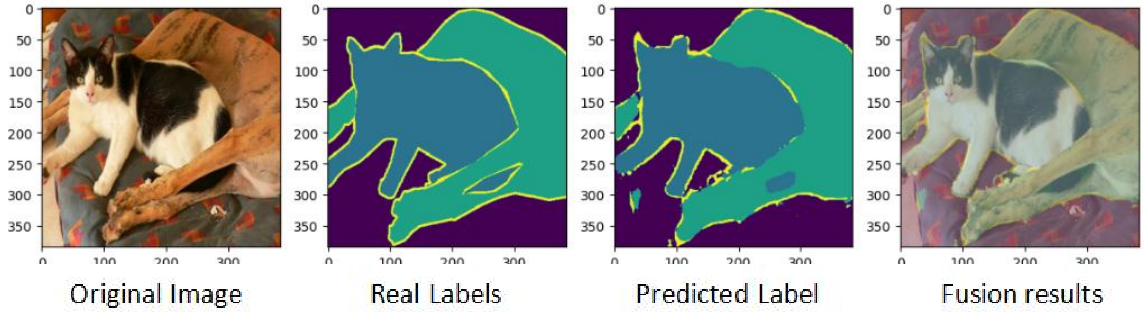

Figure12.Other Dataset Segmentation results

GUI

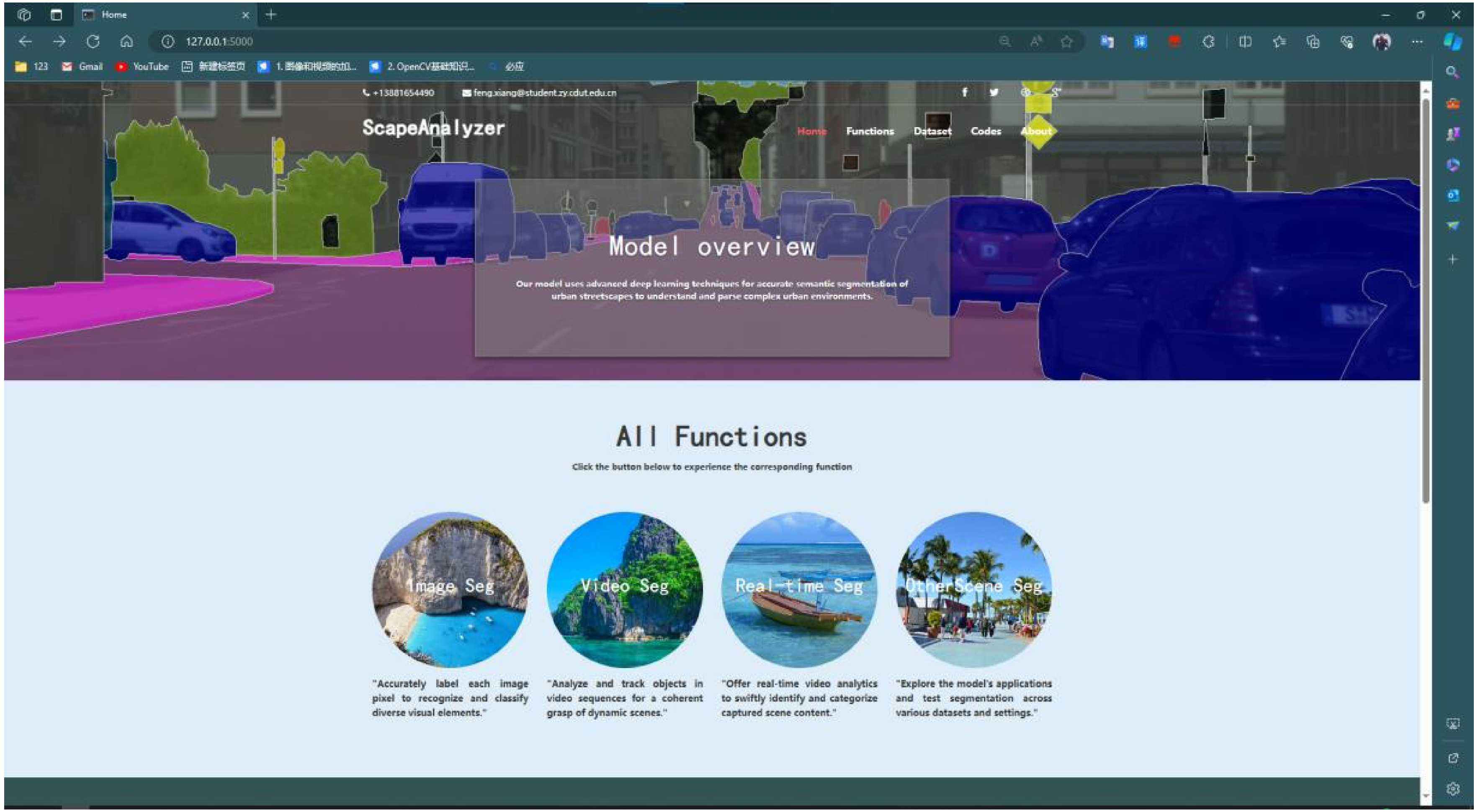

Figure13.Web Homepage(4 Functions)

GUI

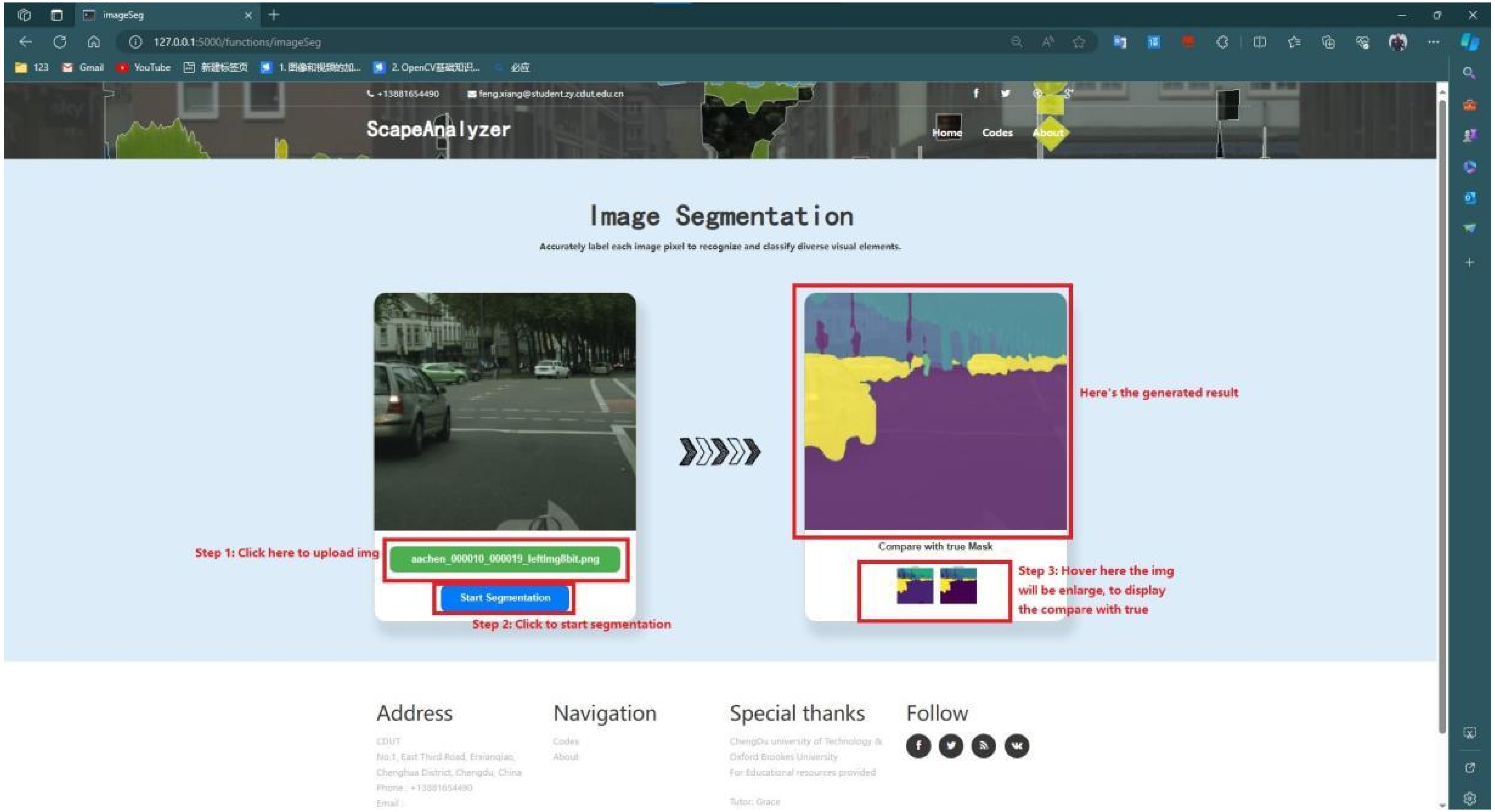

Figure14.Image Segmentation

GUI

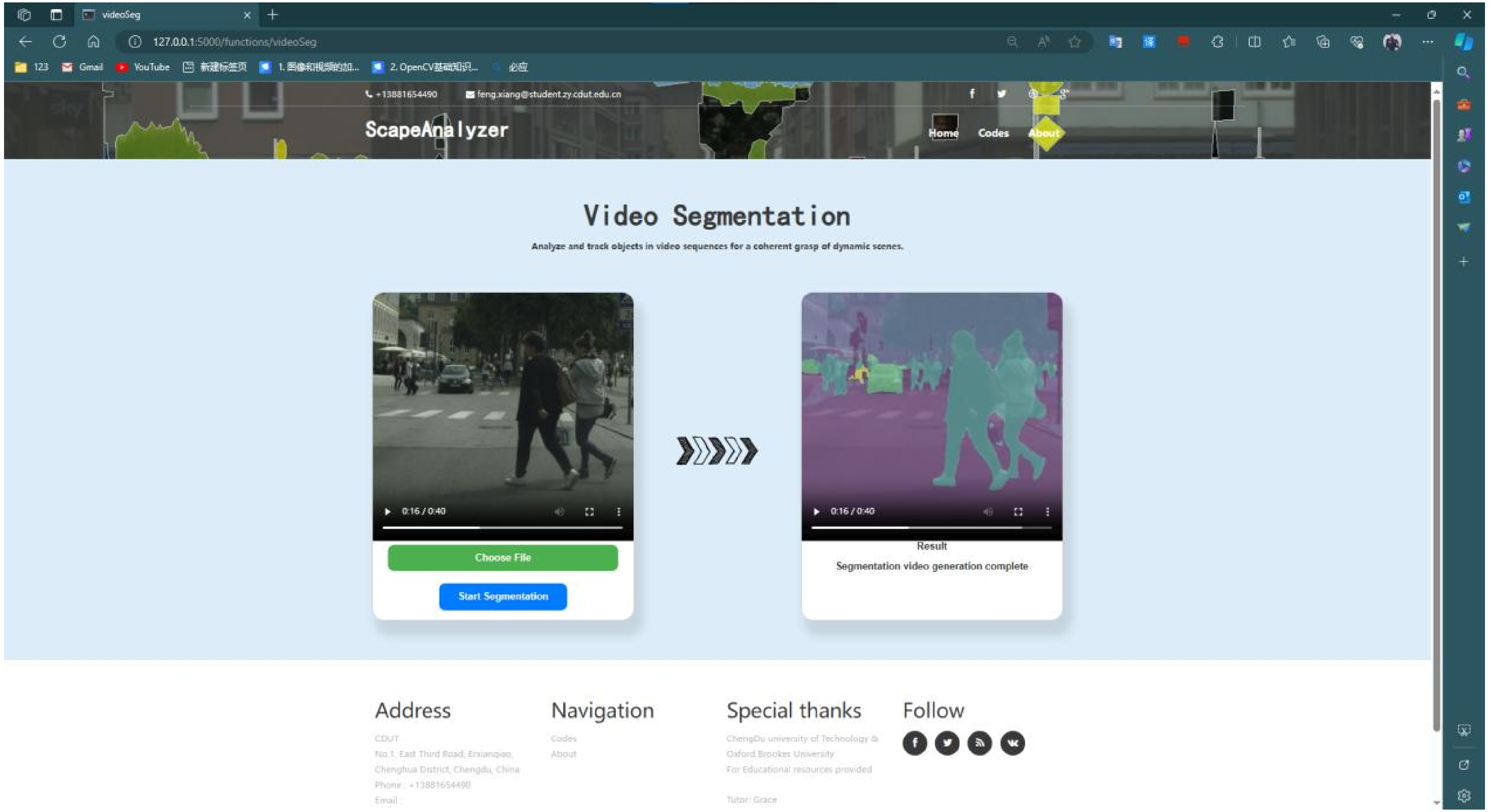

Figure15.Video Segmentation

GUI

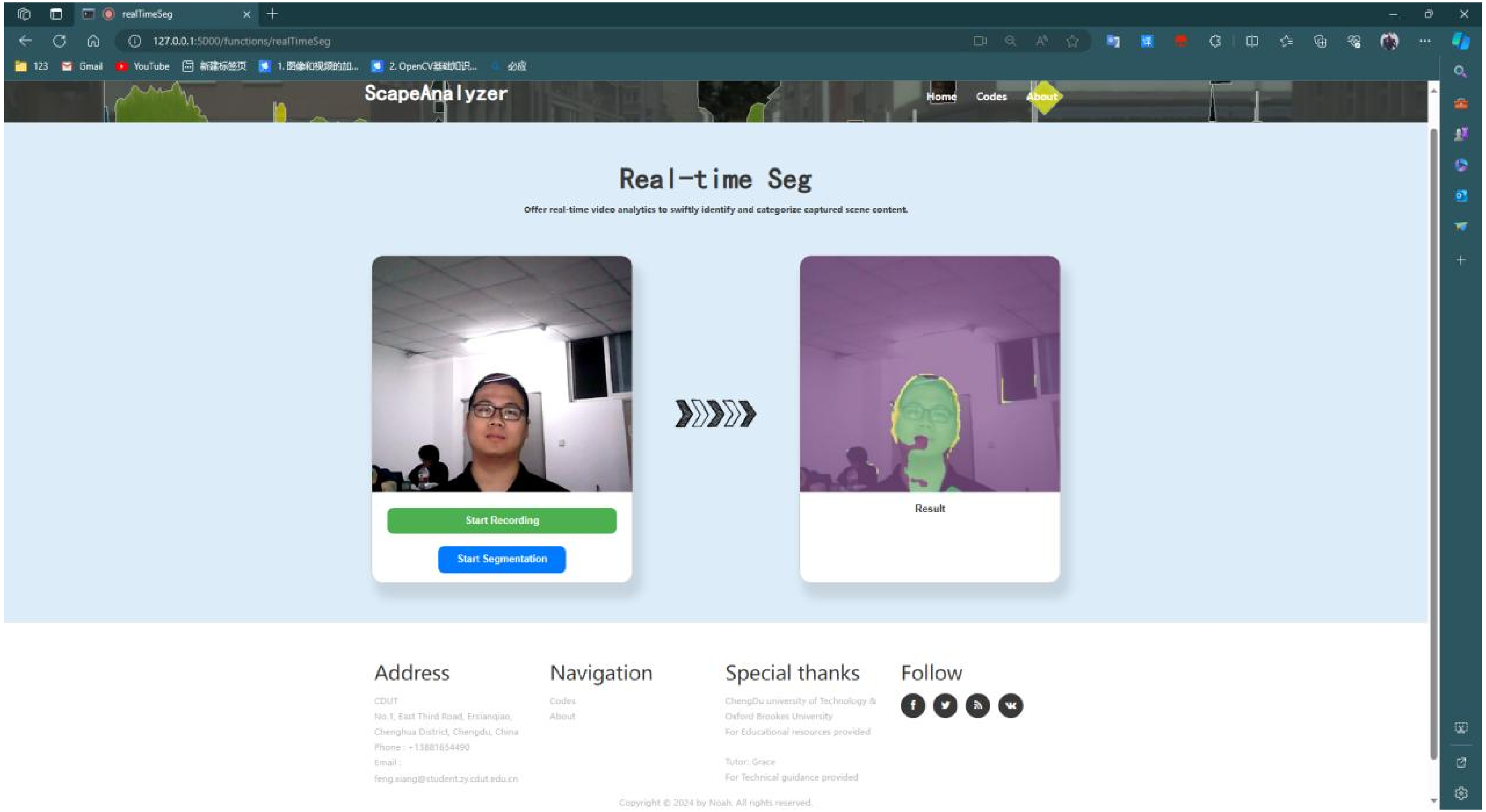

Figure16.Real-time Segmentation

GUI

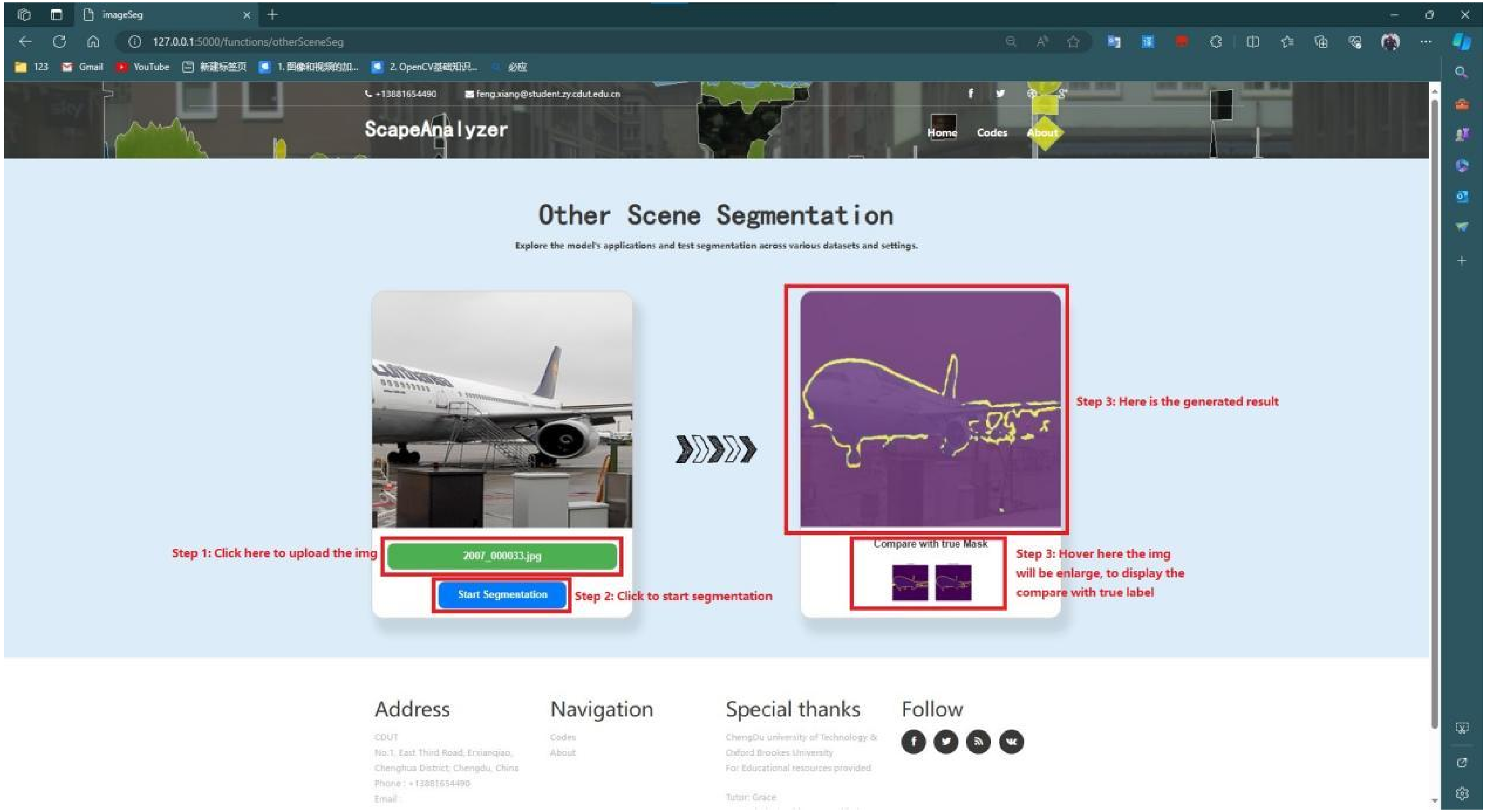

Figure17.Other Dataset Segmentation

Limitation & Challenge

Limitation

- Global feature loss due to cropping
- Semantic segmentation models large training costs
- Limitations of model's detail handling
- Geographical limitations

Challenge

- Real-time computing requirements (cross-platform)
- Pixel numbers in each category are not balanced in the dataset
- The scale of segmented objects varies greatly
- Dynamic scene processing

Future work

- Improved loss function (more focus on fewer categories)
- Investigate small target object detection techniques (high fines feature fusion or attention mechanisms)
- Explore lightweight model architectures (model pruning and knowledge distillation)
- Enhance the model's ability analyze dynamic scenes (integrated LSTM)

Conclusion

This research designs and implements a high-performance semantic segmentation model, which combines multiple deep learning modules in order to achieve better complex scene processing, multi-scale target recognition and environment adaptation.

The following works were completed in this project:

1. Semantic segmentation framework design
2. Model performance evaluation
3. Model comparison experiments
    - Performance on different categories of datasets
    - Comparison with common semantic segmentation models
    - Performance comparison of different backbone models
4. Complete user GUI interface

In conclusion, the project shows the great potential of deep learning techniques in semantic segmentation.

Reference

1. X. Wang, L. Yan, and Q. Zhang, "Research on the Application of Gradient Descent Algorithm in Machine Learning," in Proc. 2021 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, China, 2021, pp. 11-15, doi: 10.1109/ICCNEA53019.2021.00014.
2. A. Mao, M. Mohri, and Y. Zhong, "Cross-entropy loss functions: Theoretical analysis and applications," in Proc. ICML 2023. [Online]. Available: https://arxiv.org/pdf/2304.07288.pdf
3. S. Kato and K. Hotta, "Adaptive t-vMF dice loss: An effective expansion of dice loss for medical image segmentation," Computers in Biology and Medicine, vol. 168, p. 107695, 2024, doi: https://doi.org/10.1016/j.compbiomed.2023.107695.
4. F. Lateef and Y. Ruichek, "Survey on semantic segmentation using deep learning techniques," Neurocomputing, vol. 338, pp. 321–348, 2019. Available: https://doi.org/10.1016/j.neucom.2019.02.003.
5. A. Chaurasia and E. Culurciello, "LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation," CoRR, vol. abs/1707.03718, 2017. [Online]. Available: http://arxiv.org/abs/1707.03718
6. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," CoRR, vol. abs/1505.04597, 2015. [Online]. Available: http://arxiv.org/abs/1505.04597
7. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," CoRR, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385
8. F. Yu and V. Koltun, "Multi-Scale Context Aggregation by Dilated Convolutions," in 2016 International Conference on Learning Representations (ICLR), 2016. [Online]. Available: https://doi.org/10.48550/arXiv.1511.07122
9. C.-F. (Richard) Chen, Q. Fan, and R. Panda, "CrossViT: Cross-Attention Multi-Scale Vision Transformer for Image Classification," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2021, pp. 357-366.

Thank you all for your listening!