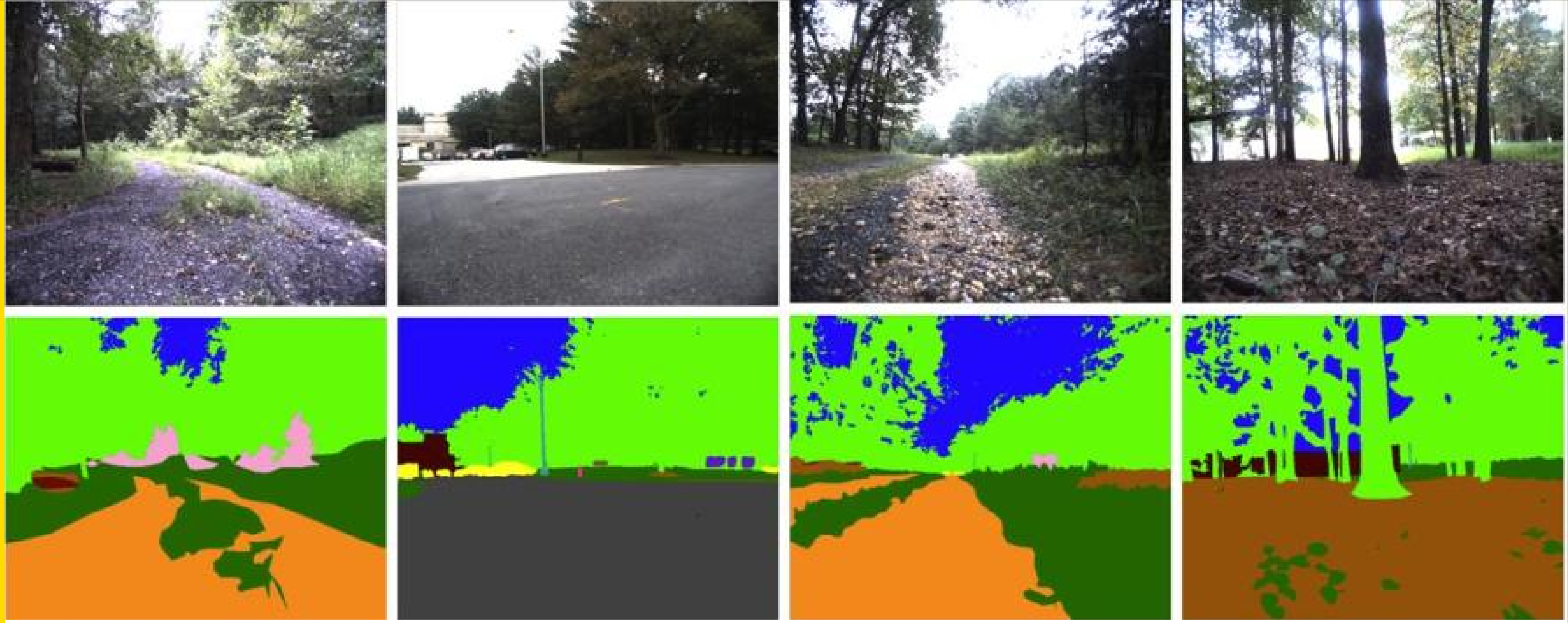
Project ID: 038

Semantic Segmentation in Unstructured Environments



AIMING FOR HD _GeeNeeTaiMay_Group

Adrian Zhang - z5242789

Xilang Zhang - z5511611

Jiehao Chen - z5507255

Tianyu Chen - z5543261

Ziyang Lu - z5459764

UNSW SYDNEY

Introduction

This project develops a semantic segmentation model for autonomous navigation in unstructured environments, addressing diverse terrains (e.g., forests, semi-urban areas) using the RUGD dataset. It aims to improve robot perception in complex settings for safer, more efficient navigation.

Input: RUGD dataset
(unstructured environment images)
Model: unet+attention unet+cbam unet+resnet unet

Motivation

Optimizing various models through literature review and selecting the best model based on data metrics comparison to enhance performance.

Presenter: Xilang Zhang - z5511611

Problem Statement

Model Selection: Determining the most suitable image segmentation model for unstructured environments is challenging due to varying strengths and weaknesses of different models across complex terrains.

Accuracy vs. Efficiency: Balancing high segmentation accuracy with computational efficiency is essential, particularly for real-time processing in autonomous systems.

Dataset Variability: The RUGD dataset contains diverse terrains and complex visual classes, making it difficult to create a model that generalizes well across different scenes.

Data Noise and Environmental Changes: Handling noisy data and adapting to environmental changes (e.g., lighting, weather) is necessary to ensure the model's robustness in real-world scenarios.

Metric Comparison for Optimization: Comparing different models based on performance metrics can be complex but is essential to determine the most effective model.

Literature Review

U-Net: Classic encoder-decoder structure that retains spatial details, suitable for fine-grained segmentation in unstructured environments.

Attention U-Net: Adds attention mechanisms to focus on key areas in complex backgrounds, improving segmentation accuracy.

CBAM U-Net: Enhances key region recognition with channel and spatial attention modules, ideal for noisy natural scenes.

Improved Model - ResUNet: Incorporates residual blocks in U-Net, boosting training stability and generalization, adapting well to diverse natural environments.

Relevance to the Task

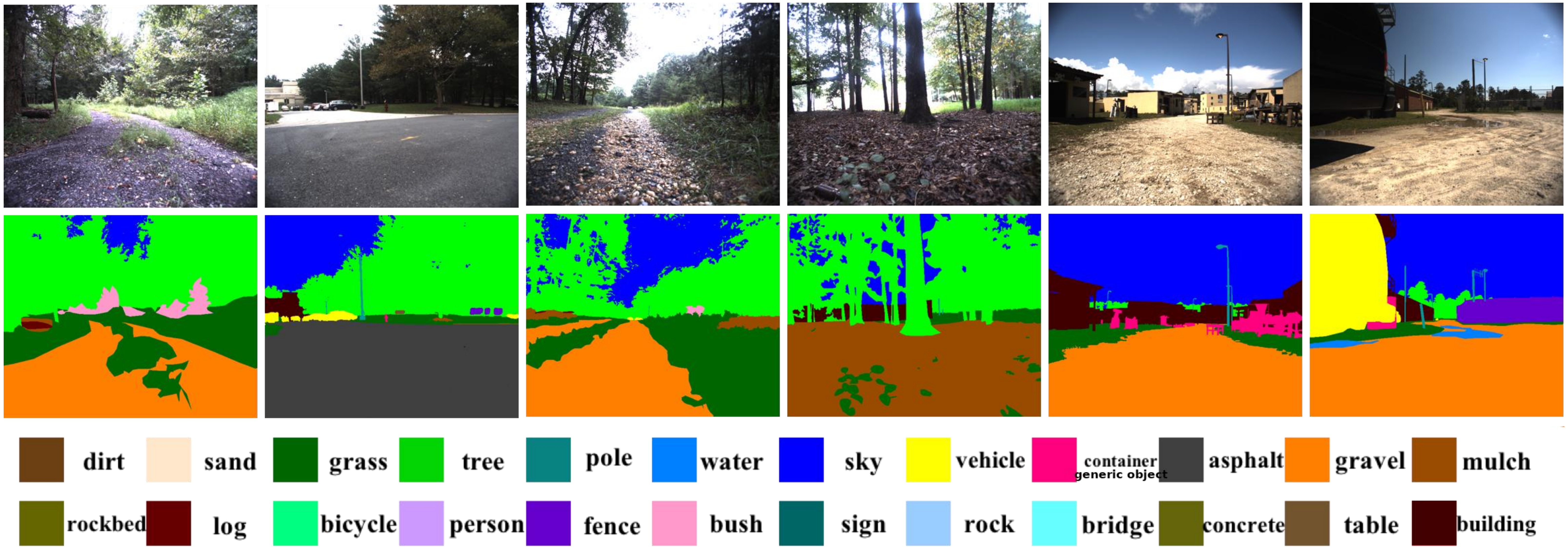
Complex Background Handling: Attention mechanisms and residual blocks help ignore irrelevant backgrounds and focus on essential features.

High-Resolution Detail Preservation: All models retain spatial details, meeting segmentation needs in unstructured environments.

Enhanced Generalization: ResUNet's residual blocks improve adaptability and stability in variable environments.

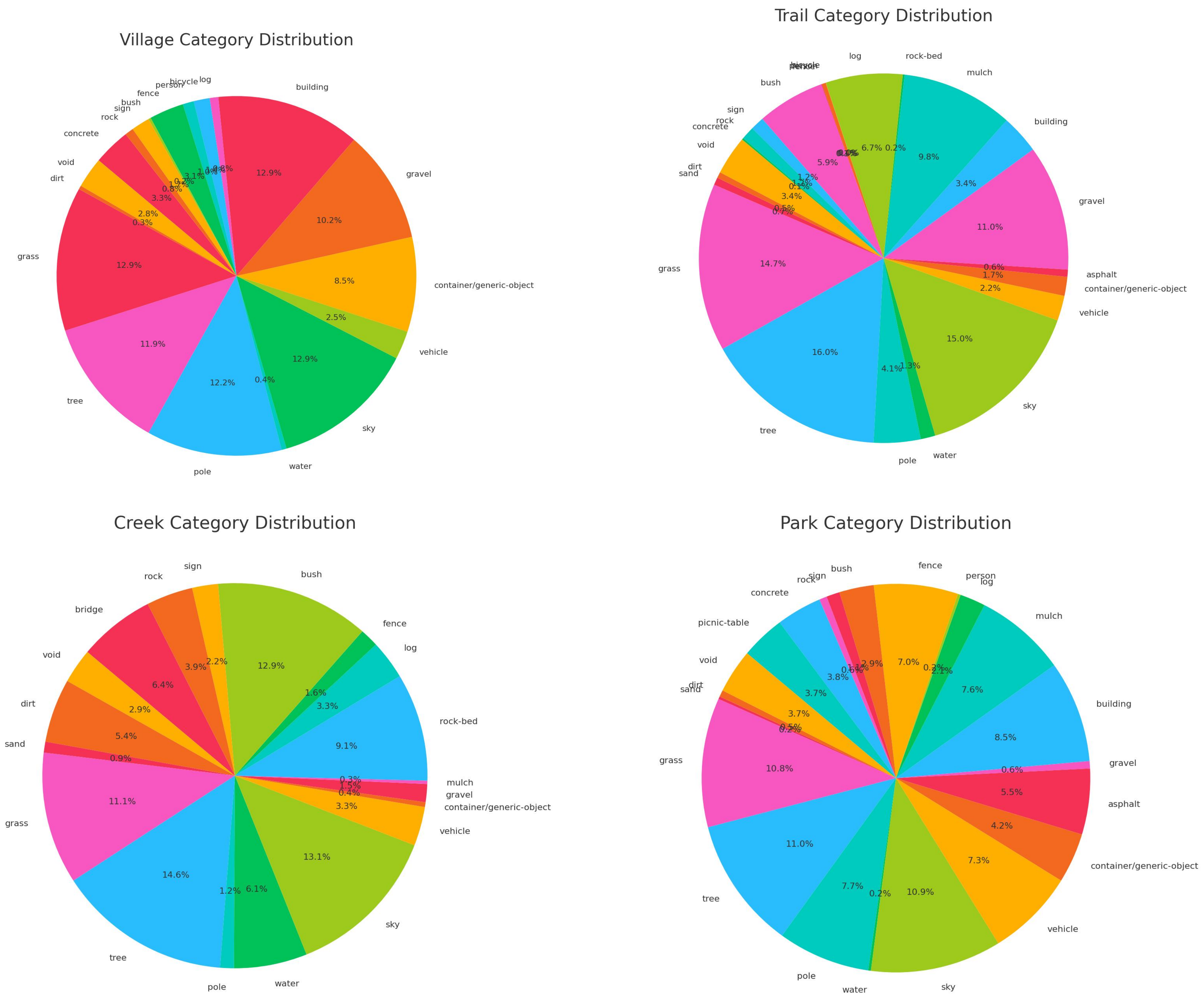
Dataset(s)

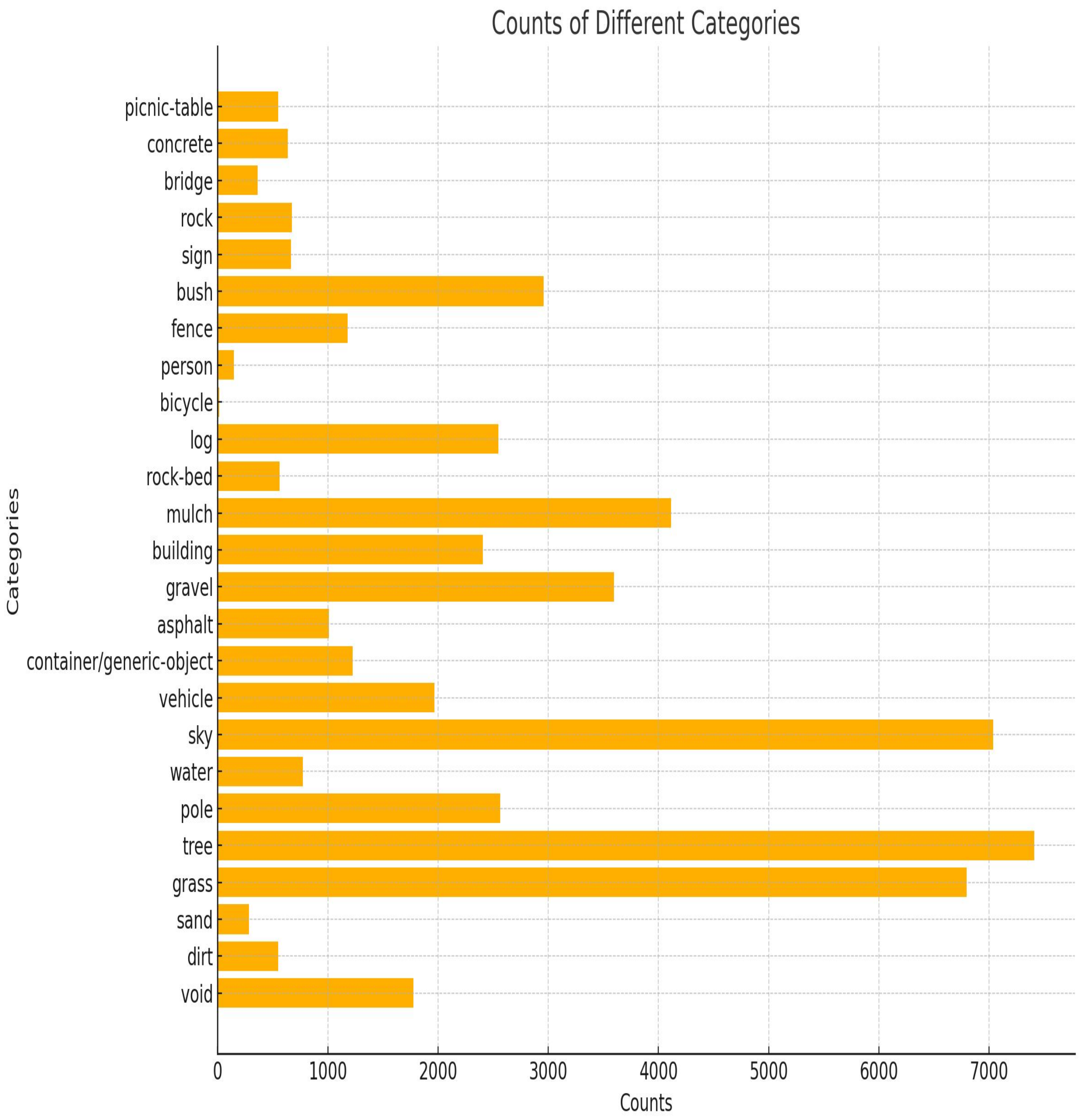
- **Dataset:** RUGD Dataset
- **Description:** Images from unstructured environments, including diverse terrains and vegetation.
- **Properties:** 7,436 images with various categories like grass, trees, vehicles, water, etc.



Data Analysis

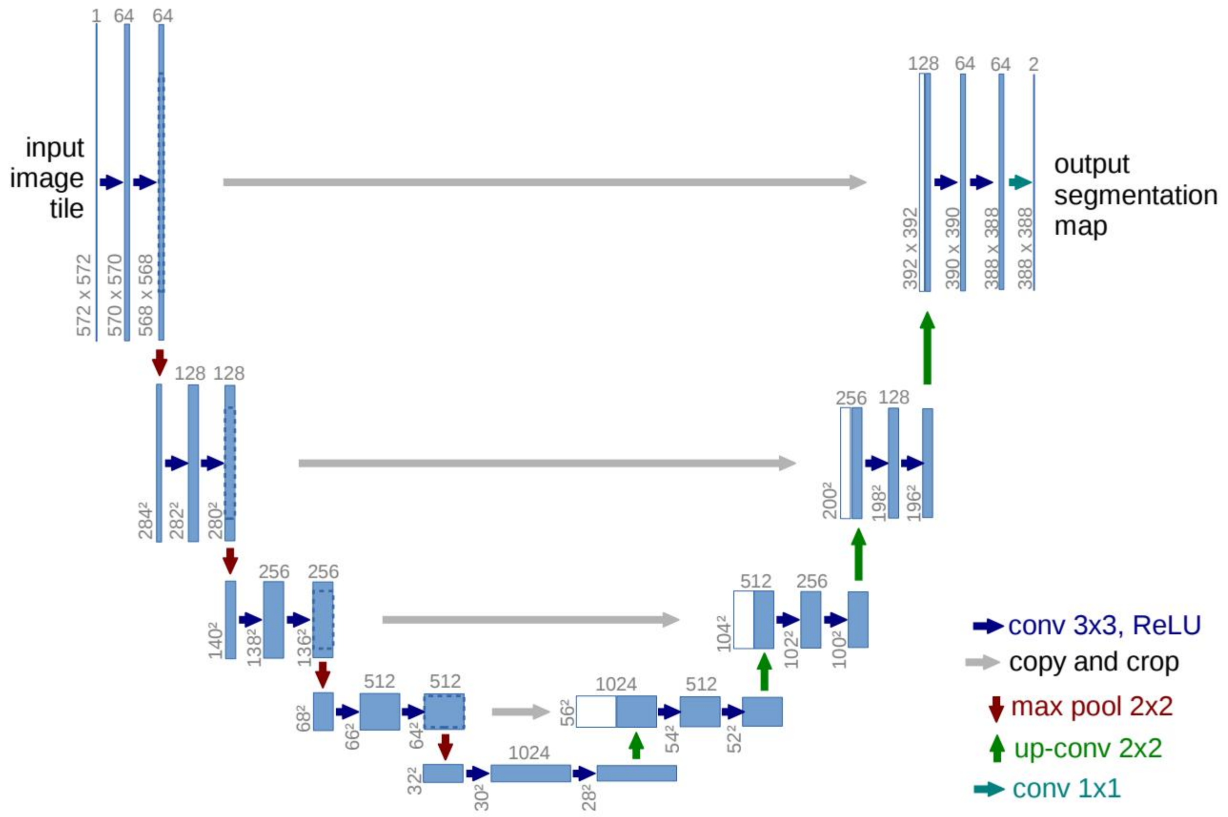
The distributions for four different categories



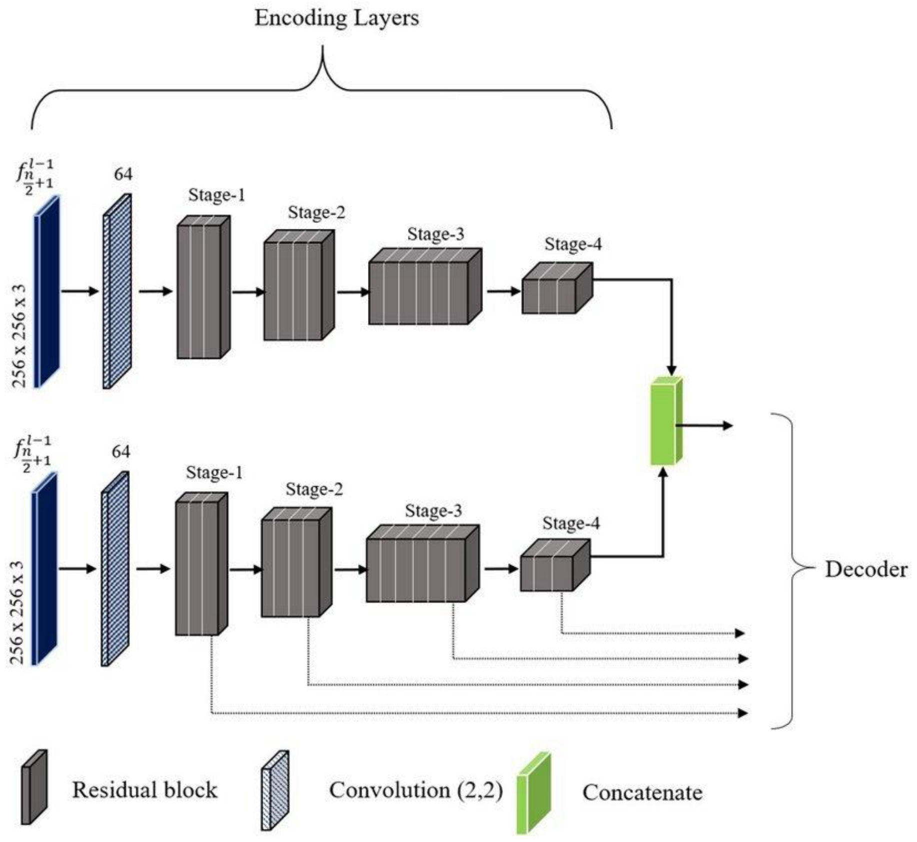


Method(s)

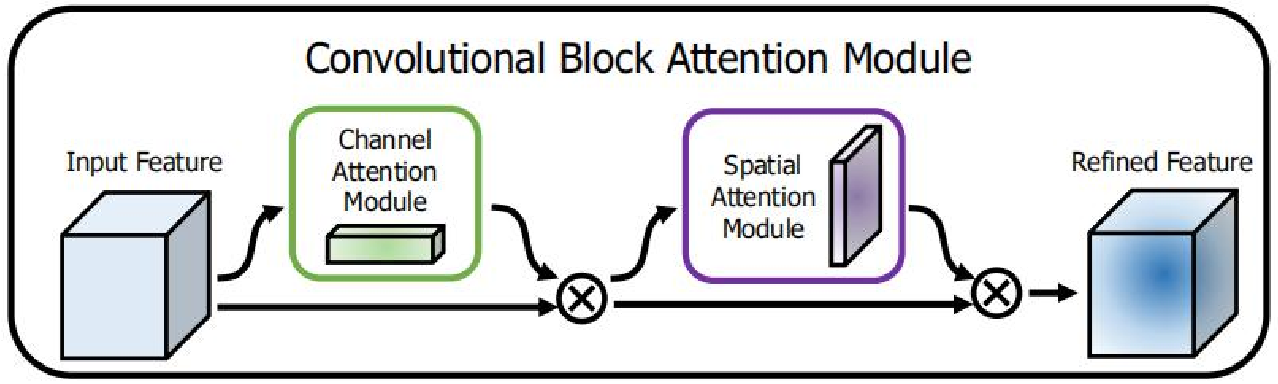
Base Unet



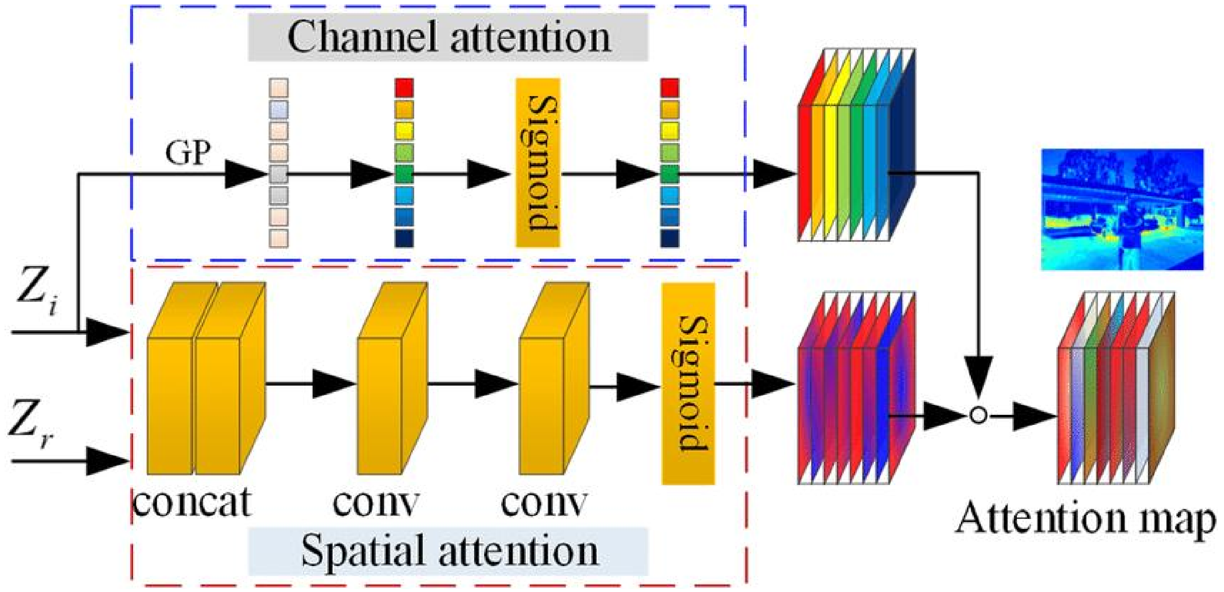
Resnet



CBAM



Attention



basic model - U-Net

1.Encoder-Decoder Structure:

This structure is particularly well suited for semantic segmentation tasks and is able to recover the feature details of the input image well into the output segmented image.

2.Jump Connections:

Useful when working with complex backgrounds or border details

```
Per-class accuracy:
void                       : 0.0000
dirt                       : 0.0008
sand                       : 0.0000
grass                      : 0.5191
tree                       : 0.8459
pole                       : 0.1364
water                      : 0.2756
sky                        : 0.7392
vehicle                    : 0.0035
container/generic-object   : 0.0000
asphalt                    : 0.0000
gravel                     : 0.3284
building                   : 0.0000
mulch                      : 0.0000
rock-bed                   : 0.1195
log                        : 0.0000
bicycle                    : 0.0000
person                     : 0.0000
fence                      : 0.0000
bush                       : 0.2591
sign                       : 0.0000
rock                       : 0.0000
bridge                     : 0.0512
concrete                   : 0.0000
picnic-table               : 0.0000
Test results - Loss: 8.3551, Acc: 0.5018, mIoU: 0.6071, MPA: 0.7329
```

basic model - attention unet

```
Per-class accuracy:
void                        : 0.0000
dirt                        : 0.0008
sand                        : 0.0000
grass                       : 0.5104
tree                        : 0.8589
pole                        : 0.1058
water                       : 0.1649
sky                         : 0.7826
vehicle                     : 0.0166
container/generic-object : 0.0000
asphalt                     : 0.0000
gravel                      : 0.2065
building                    : 0.0000
mulch                       : 0.0000
rock-bed                    : 0.0763
log                         : 0.0000
bicycle                     : 0.0000
person                      : 0.0000
fence                       : 0.0000
bush                        : 0.1536
sign                        : 0.0000
rock                        : 0.0000
bridge                      : 0.0642
concrete                    : 0.0000
picnic-table                : 0.0000
Test results - Loss: 8.2751, Acc: 0.5385, mIoU: 0.6589, MPA: 0.7524
```

Attention Mechanism:

allowing the model to selectively focus on parts that contribute more to a particular task.

basic model - CBAM

Channel Attention Module:

Each channel is assigned a different weight so as to enhance the characteristics of the important channels.

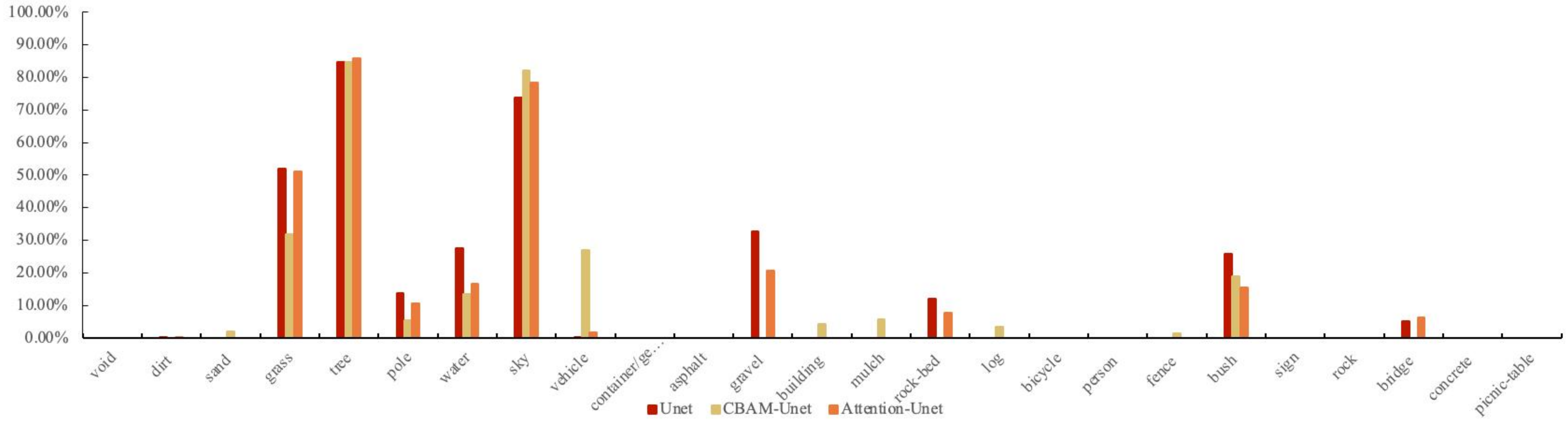
Spatial Attention Module:

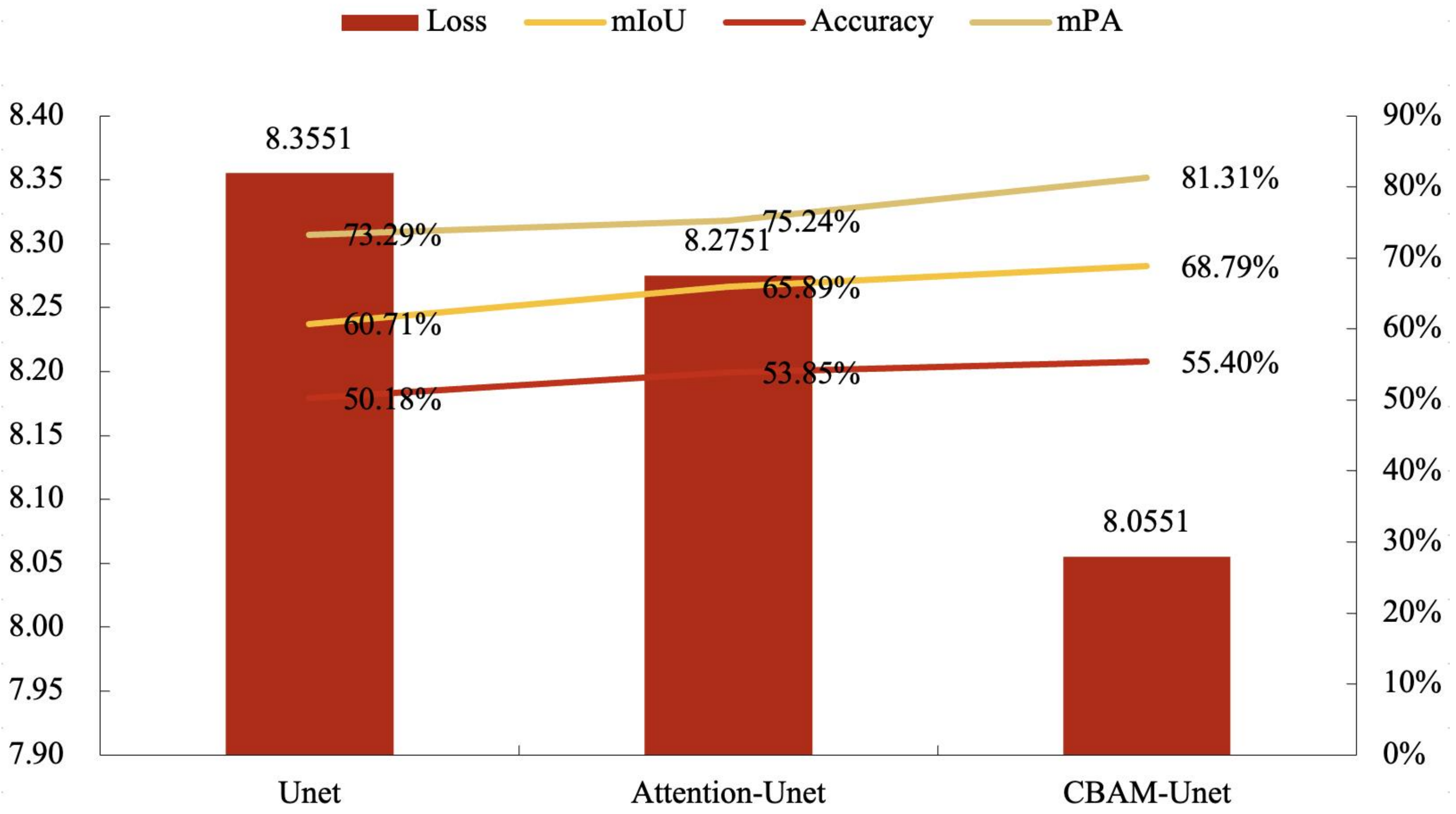
By calculating the importance of each location, the entire feature map is weighted to highlight important locations in space.

```
Per-class accuracy:
void                     : 0.0000
dirt                     : 0.0000
sand                     : 0.0193
grass                    : 0.3191
tree                     : 0.8459
pole                     : 0.0531
water                    : 0.1359
sky                      : 0.8204
vehicle                  : 0.2684
container/generic-object : 0.0000
asphalt                  : 0.0000
gravel                   : 0.0000
building                 : 0.0415
mulch                    : 0.0569
rock-bed                 : 0.0000
log                      : 0.0328
bicycle                  : 0.0000
person                   : 0.0000
fence                    : 0.0136
bush                     : 0.1896
sign                     : 0.0000
rock                     : 0.0000
bridge                   : 0.0000
concrete                 : 0.0000
picnic-table             : 0.0000
Test results - Loss: 8.0551, Acc: 0.5540, mIoU: 0.6879, MPA: 0.8131
```

Discussion

Three BaseLine model Per-Class Accuracy



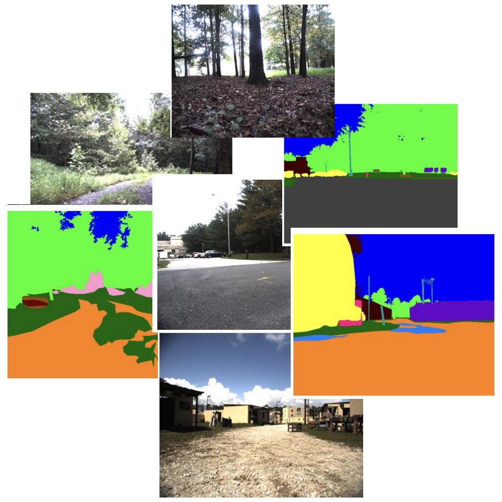


Three BaseLine model comparison of metrics

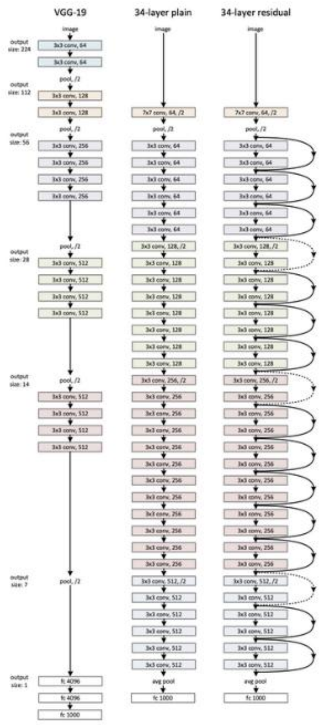
Resnet34 - Unet

Model Architecture

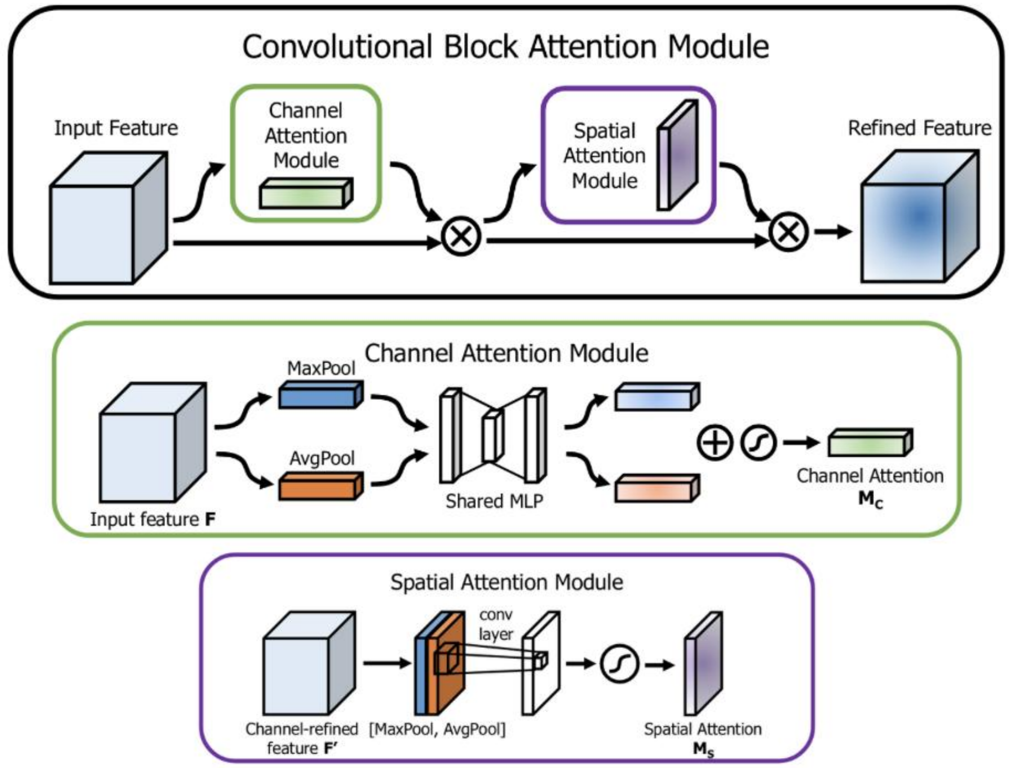
RUGD data set



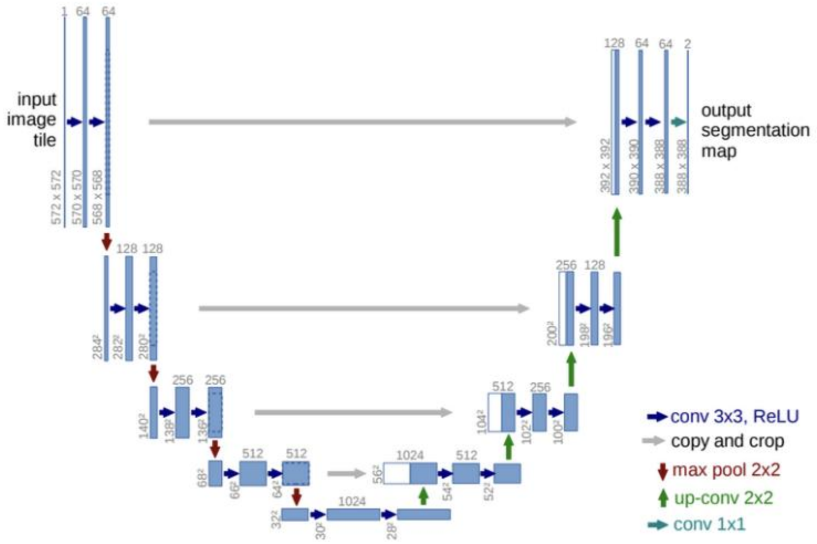
Resnet34
BackBone



CBAM

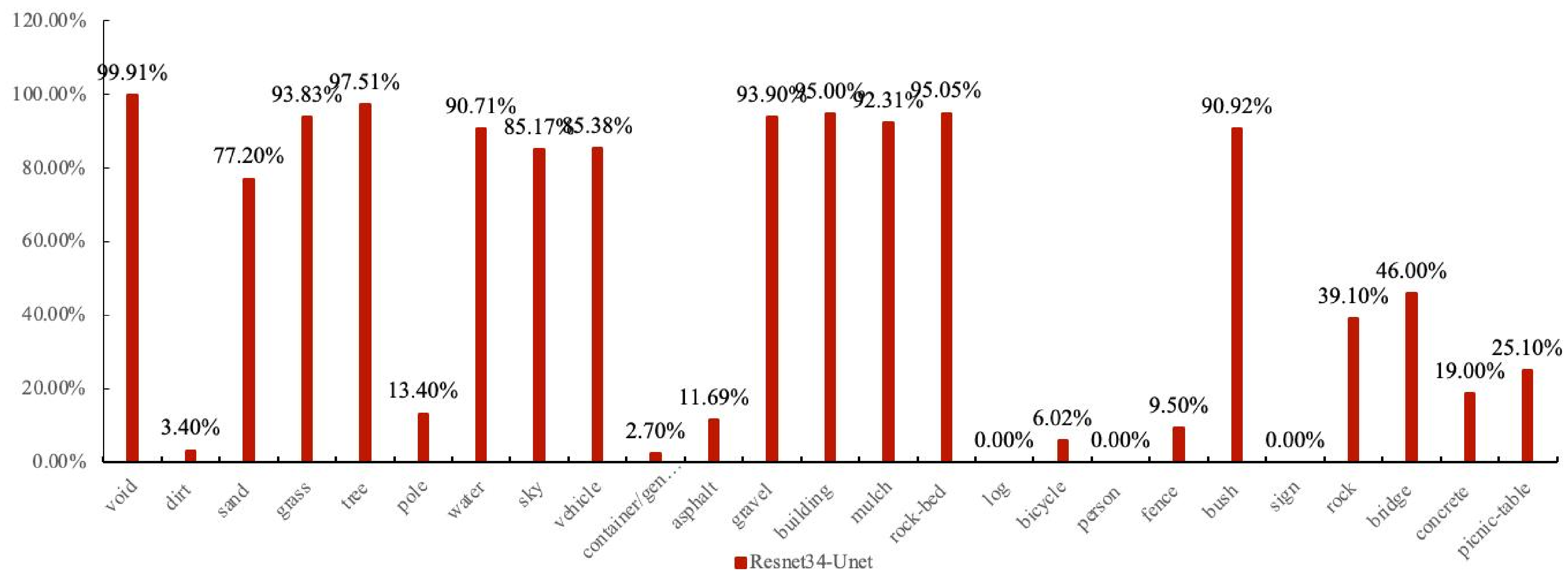


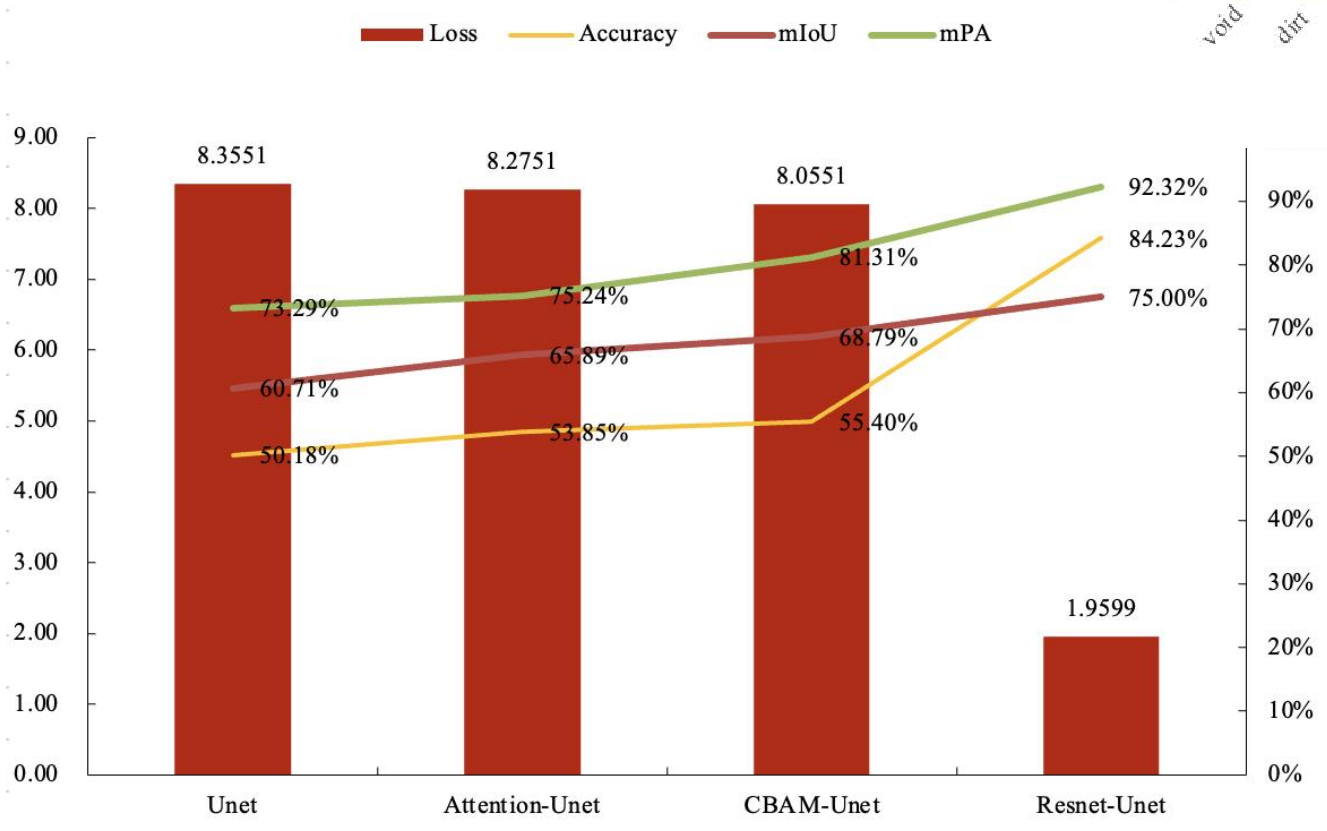
Unet



Resnet - Unet

Per-Class Accuracy of Resnet-Unet model





Models Performance Comparison

Results

- The performance of the models (including advanced model) was evaluated using three primary metrics: Mean Pixel Accuracy (MPA), Intersection over Union (mIoU), and Accuracy. Each of these metrics provides a distinct perspective on the quality of semantic segmentation and the effectiveness of different model variations. Here is the chart:

	ACC	MIOU	MPA
UNet	0.5018	0.6071	0.7329
Attention-UNet	0.5385	0.6589	0.7524
CBAM-UNet	0.5540	0.6879	0.8131
Res-UNet	0.8423	0.7508	0.9232

Conclusion

In this project, we complete those key steps for semantic segmentation in unstructed environments:

- **Data analysis and preprocessing**
- **Baseline model implement**
- **Evaluation metric compareity**
- **modle improvment**

And these steps support applications like auto-driving segments different items in complex environments, also will be improved in future researchs.

References

1. Ronneberger, O., Fischer, P., & Brox, T. (2015, May 18). U-Net: Convolutional Networks for Biomedical Image Segmentation. ArXiv.org. https://arxiv.org/abs/1505.04597
2. Oktay, O., Schlemper, J., Folgoc, Loic Le, Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, Nils Y, Kainz, B., Glocker, B., & Rueckert, D. (2018). Attention U-Net: Learning Where to Look for the Pancreas. ArXiv.org. https://arxiv.org/abs/1804.03999
3. Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). CBAM: Convolutional Block Attention Module. ArXiv:1807.06521 [Cs]. https://arxiv.org/abs/1807.06521
4. Diakogiannis, F. I., Waldner, F., Caccetta, P., & Wu, C. (2020). ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS Journal of Photogrammetry and Remote Sensing, 162, 94–114. https://doi.org/10.1016/j.isprsjprs.2020.01.013


Questions