KrackHack 3.0
AI/ML Track
Offroad Semantic Scene Segmentation

1. Team Details

Team Name: HyperBool

Project Repository: Github Repo Link

Member 1 (Team Leader) GitHub: Aryan Sisodiya

Member 2 GitHub: Daksh Rathi

Member 3 GitHub: Farhan Alam

Member 4 GitHub: Tanmay Pratap Singh

2. Problem Statement

Autonomous off-road navigation requires accurate semantic understanding of unstructured environments to identify terrain types, vegetation, and obstacles. Unlike urban environments, collecting and annotating real-world off-road data is expensive and logistically challenging.

The selected problem statement focuses on **semantic segmentation of off-road desert scenes** using **synthetic data**. The objective is to train a robust model that generalizes well to unseen environments and is evaluated using the mean Intersection over Union (mIoU) metric.

3. Dataset Overview

The dataset consists of synthetic desert images paired with pixel-wise semantic segmentation masks. Each pixel in the mask corresponds to a semantic class such as terrain, vegetation, obstacles, or sky.

3.1 Dataset Structure

```
dataset/
  train/
    Color_Images/
    segmentation/
  val/
    Color_Images/
    segmentation/
```

The segmentation masks contain integer class IDs that are remapped to contiguous indices during training. Since image and mask filenames do not directly correspond, image–mask pairs are aligned based on sorted order.

4. Proposed Approach

4.1 Model Architecture

The proposed solution uses **DeepLabV3+**, a semantic segmentation architecture designed for effective multi-scale context understanding. An **EfficientNet-B0** encoder pretrained on ImageNet is used for feature extraction, providing a balance between accuracy and computational efficiency.

4.2 Training Strategy

All images are resized to 256×256 prior to training. To address class imbalance, a composite loss function combining **Dice Loss** and **Focal Loss** is used. Optimization is performed using the Adam optimizer, and model performance is evaluated using the mIoU metric.

Mixed-precision training is enabled on GPU to improve training speed and reduce memory usage.

4.3 Data Augmentation

To improve generalization and reduce overfitting to synthetic textures, data augmentation techniques such as horizontal flipping, brightness and contrast adjustment, mild color jittering, and Gaussian noise are applied during training.

5. Tech Stack

Software:

- Python
- PyTorch
- segmentation-models-pytorch
- Albumentations
- OpenCV
- Jupyter Notebook

Hardware:

- NVIDIA RTX 4060 GPU (CUDA-enabled)

6. Workflow Overview

The project workflow is organized as follows:

1. Data loading and preprocessing

2. Class ID remapping for segmentation masks
3. Model training with data augmentation
4. Validation using mIoU and per-class IoU
5. Model checkpointing
6. Visualization and failure case analysis

7. Experiments and Results

7.1 Training Setup

The semantic segmentation model was trained on the synthetic desert dataset using mixed precision to accelerate training and reduce GPU memory usage. Input images were resized to a spatial resolution of 256×256, and a batch size of 8 was used throughout all experiments. Optimization was performed using the Adam optimizer with a fixed learning rate. The training objective combines region-overlap and class-balancing considerations through a composite loss function.

Training was conducted for 30 epochs. Model checkpoints were saved based on improvements in validation mean Intersection over Union (mIoU), and the best-performing checkpoint was retained for evaluation.

7.2 Quantitative Results

The model demonstrates stable optimization behavior, with validation performance improving consistently during the initial training phase. The validation mIoU increases from 0.5540 at the first epoch to a peak value of 0.5687 by epoch 13, after which performance saturates with marginal gains. This indicates effective learning from synthetic imagery and reasonable generalization to the validation split.

To better understand class-wise behavior, Intersection over Union (IoU) scores were analyzed per semantic category. Dominant terrain classes achieve relatively high IoU, while visually ambiguous or smaller obstacle classes exhibit lower performance, reflecting the inherent challenges of off-road perception under appearance variation and class imbalance.

Table 1: Class-wise IoU on the validation set (selected classes).

Class ID	**IoU**
100	0.77
200	0.62
300	0.65
500	0.44
550	0.25
700	0.37
10000	0.97
Mean IoU	**0.5687**

7.3 Qualitative Observations

Qualitative inspection of predicted segmentation maps shows that the model successfully captures large contiguous regions such as terrain and sky, while finer structures and visually similar classes (e.g., small obstacles against sandy backgrounds) remain challenging. These results highlight both the promise and the current limitations of training purely on synthetic data for off-road semantic segmentation.

7.4 Discussion

The presented results serve as an initial baseline demonstrating the feasibility of learning meaningful pixel-level representations from synthetic desert imagery. While the achieved performance indicates stable training and reasonable class separation, certain categories remain difficult to segment accurately. This motivates future work on improved data augmentation, domain adaptation techniques, and architectural enhancements to further bridge the synthetic-to-real domain gap.