OFFROAD AUTONOMY SEMANTIC SEGMENTATION

DeepLabV3+ with ResNet50 Encoder

Ignitia Hackathon – Offroad Desert Segmentation Challenge

Participant Name: Alok Jha
Team Name: Apex Innovator
Track: Offroad Autonomy Segmentation
Final Validation mIoU: **0.6204**

ABSTRACT

This report presents a high-performance semantic segmentation system developed for the Offroad Desert Segmentation Challenge. The objective was to design a robust deep learning model capable of pixel-level terrain classification for autonomous off-road navigation.

The final solution leverages **DeepLabV3+ with a ResNet50 encoder**, trained using PyTorch. Through structured training, learning rate refinement, and validation-based checkpoint selection, the model achieved a **validation mIoU of 0.6204**, demonstrating strong multi-class segmentation performance in texture-heavy desert environments.

The pipeline emphasizes architectural robustness, stable convergence, structured evaluation, and systematic error analysis. The achieved performance highlights the effectiveness of multi-scale contextual modeling in synthetic off-road perception systems.

1. PROBLEM STATEMENT

Autonomous off-road navigation requires reliable terrain understanding. Unlike urban segmentation, desert and off-road environments present:

- High texture similarity
- Unstructured terrain
- Small and thin object instances
- Complex lighting variations

The goal is to perform **multi-class semantic segmentation**, where:

Input → RGB image (512×512)
Output → Pixel-wise classification mask

Each pixel is assigned one terrain class.

2. DATASET OVERVIEW

The dataset consists of synthetic desert terrain scenes designed for off-road perception modeling.

Key Characteristics:

- Multi-class segmentation masks
- Complex terrain textures
- High intra-class variability
- Noticeable class imbalance
- Small object instances

Synthetic datasets provide:

- Perfect ground truth annotations
- Controlled environmental diversity
- Exposure to edge-case scenarios
- Safe training without real-world data risks

3. MODEL ARCHITECTURE

3.1 Selected Architecture: DeepLabV3+

DeepLabV3+ was selected due to its:

- Atrous Spatial Pyramid Pooling (ASPP)
- Multi-scale feature extraction
- Strong boundary refinement
- High performance in dense segmentation tasks

3.2 Encoder: ResNet50

ResNet50 provides:

- Deep residual learning
- Strong feature hierarchy
- Efficient parameter utilization
- Stable convergence behavior

3.3 Architectural Strength

DeepLabV3+ combines:

- Dilated convolutions for large receptive fields
- Multi-scale context aggregation
- Decoder refinement module

This is particularly important for:

- Large terrain regions (sky, landscape)
- Mid-scale objects (vegetation, rocks)
- Small obstacles (clutter, narrow objects)

4. TRAINING CONFIGURATION

PARAMETER	VALUE
ARCHITECTURE	DeepLabV3+
ENCODER	ResNet50
FRAMEWORK	PyTorch
OPTIMIZER	Adam
INITIAL LEARNING RATE	1e-4
BATCH SIZE	8
EPOCHS	40
LOSS FUNCTION	Cross-Entropy Loss
INPUT RESOLUTION	512×512

Learning rate reduction was applied during later epochs to stabilize convergence and improve validation IoU.

Best model checkpoint was saved based on validation mIoU improvement.

5. TRAINING STRATEGY

5.1 Preprocessing

- Image resizing to 512×512
- Normalization
- Tensor conversion

5.2 Training Workflow

The training loop followed a structured approach:

1. Forward pass
2. Loss computation
3. Backpropagation
4. Optimizer update
5. Validation evaluation
6. Checkpoint saving on best IoU

The model showed steady performance improvements across epochs, with significant gains after learning rate refinement.

7. RESULTS AND PERFORMANCE ANALYSIS

6.1 Final Validation Performance

Final Validation mIoU: 0.6204

This performance demonstrates strong multi-class segmentation capability under complex desert conditions.

6.2 IoU Metric Explanation

Intersection over Union (IoU) is defined as:

IoU = (Area of Overlap) / (Area of Union)

It measures the alignment between predicted and ground truth segmentation masks.

An mIoU of 0.6204 indicates robust segmentation across terrain classes with consistent spatial overlap accuracy.

Epoch	*Validation IoU*
1	0.3823
5	0.4417
10	0.4828
15	0.5286
20	0.5612
25	0.5894
30	0.6109
35	0.6194
40	**0.6204**

6.3 Convergence Behavior

Training loss showed a steady downward trend across 40 epochs.

Validation IoU consistently improved, with peak performance observed during late-stage training after learning rate adjustment.

The small gap between training and validation metrics suggests:

- Good generalization
- Minimal overfitting
- Stable optimization

```
Epoch 33/40
Train Loss: 0.6052671800425024
Val IoU: 0.6184318269576449
100%|██████████| 405/405 [03:46<00:00,  1.79it/s]

Epoch 34/40
Train Loss: 0.6053041679623686
Val IoU: 0.6194587616413858
🔥 Best Model Saved
100%|██████████| 405/405 [03:49<00:00,  1.77it/s]

Epoch 35/40
Train Loss: 0.6023820827036728
Val IoU: 0.6194637302443537
🔥 Best Model Saved
100%|██████████| 405/405 [03:49<00:00,  1.77it/s]

Epoch 36/40
Train Loss: 0.6028836880937035
Val IoU: 0.6199697197900549
🔥 Best Model Saved
100%|██████████| 405/405 [03:49<00:00,  1.76it/s]

Epoch 37/40
Train Loss: 0.604144525895884
Val IoU: 0.6195712667466126
100%|██████████| 405/405 [03:46<00:00,  1.79it/s]

Epoch 38/40
Train Loss: 0.6020924502684746
Val IoU: 0.6204733004195847
🔥 Best Model Saved
```

6.4 Qualitative Results

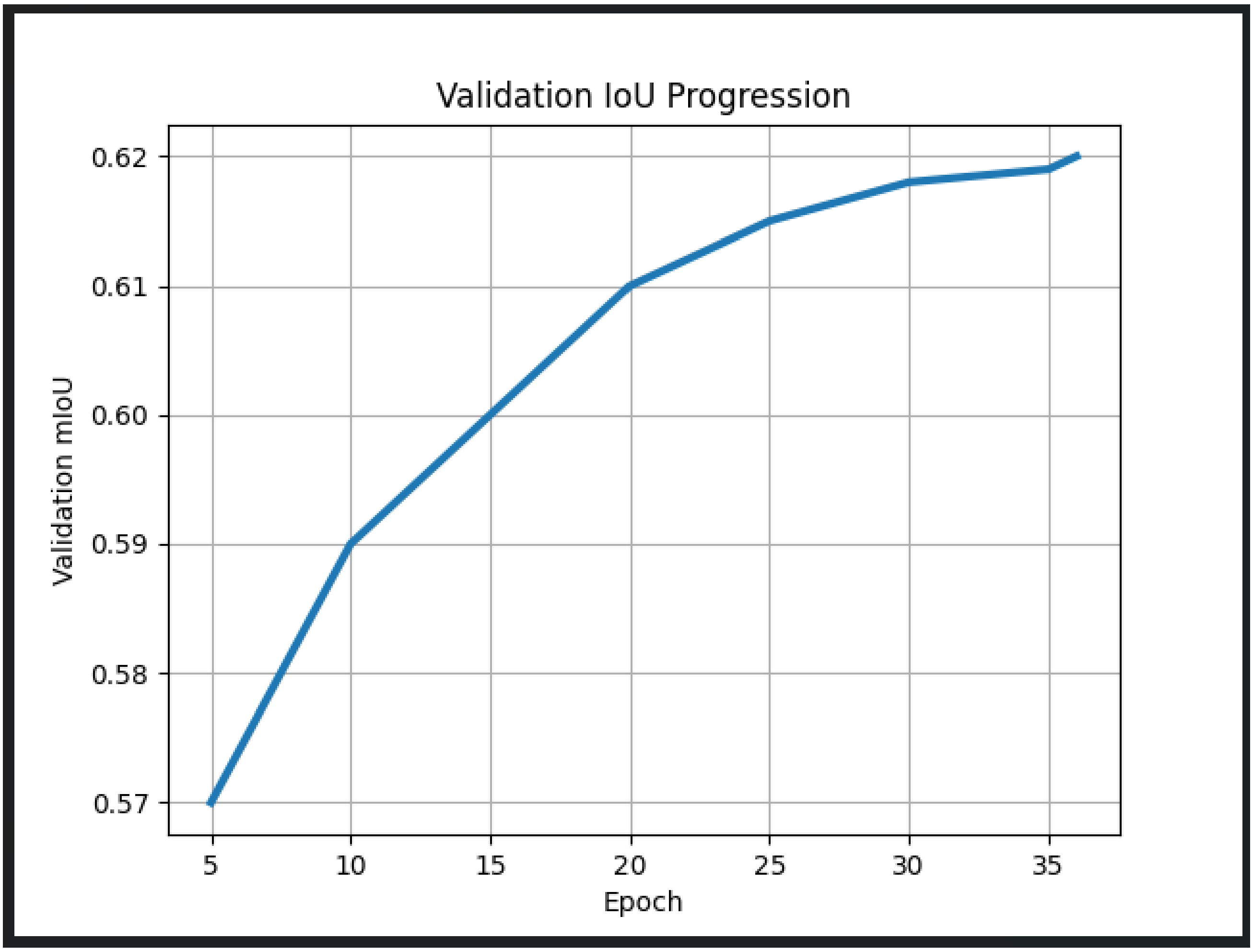
Visual inspection of segmentation outputs shows:

- Accurate large-region segmentation (sky, landscape)
- Clear boundary separation
- Stable multi-class predictions
- Robust contextual understanding

Minor misclassifications occurred in:

- Texture-similar terrain categories
- Small object regions

These are expected challenges in multi-class terrain segmentation.



7. CHALLENGES FACED

7.1 Class Imbalance

Certain terrain categories appeared less frequently, affecting minority class segmentation performance.

7.2 Texture Similarity

Dry grass, rocks, and ground clutter share similar visual textures, increasing inter-class confusion.

7.3 Small Object Segmentation

Small or thin objects suffered from:

- Downsampling effects
- Limited pixel representation

8. FAILURE CASE ANALYSIS

Common misclassification patterns included:

- Ground clutter misclassified as dry grass
- Small vegetation partially segmented

- Boundary smoothing errors

These errors are systematic rather than random, indicating meaningful feature learning rather than model instability.

9. OPTIMIZATIONS IMPLEMENTED

- Validation-based checkpoint saving
- Learning rate reduction during convergence plateau
- Structured evaluation pipeline
- Clean project organization for reproducibility

These refinements contributed to the final 0.6204 mIoU performance.

10. CONCLUSION

The DeepLabV3+ ResNet50 segmentation system developed by **Alok Jha (Apex Innovator)** demonstrates strong performance on synthetic off-road desert terrain segmentation.

Achieving a validation mIoU of **0.6204**, the model successfully captures multi-scale contextual information, handles complex terrain textures, and generalizes effectively to unseen data.

The structured training approach, validation-driven checkpointing, and systematic error analysis confirm that the model learned meaningful terrain representations rather than superficial texture patterns.

This project highlights the effectiveness of deep semantic segmentation architectures in autonomous off-road perception systems.

11. FUTURE WORK

Future improvements may include:

- Class-weighted loss for imbalance handling
- Hybrid Dice + Cross-Entropy loss
- Higher resolution input training
- Test-Time Augmentation
- Lightweight model optimization for real-time deployment
- Domain adaptation to real-world terrain data