Off-Road Multi-Class Semantic Segmentation

DeepLabV3+ with EfficientNet-B4 Backbone

Team HexTech (Ignitia Hackathon – Offroad Desert Segmentation Challenge)
Vedant Dusane
Arnav Kumar
Saqlain Abidi

Abstract

This project presents a deep learning-based semantic segmentation system for off-road terrain understanding. The objective is to classify each pixel of a 512×512 RGB image into one of ten terrain categories. The final solution uses DeepLabV3+ with an EfficientNet-B4 encoder and achieves strong validation performance across IoU, Dice Score, and Pixel Accuracy metrics.

1 Methodology

1.1 Problem Formulation

The task is formulated as a multi-class semantic segmentation problem with 10 terrain categories. Each input image is mapped to a dense pixel-wise mask.

1.2 Model Architecture

- Architecture: DeepLabV3+
- Encoder: EfficientNet-B4 (ImageNet Pretrained)
- Input Resolution: 512×512
- Number of Classes: 10
- Framework: PyTorch

DeepLabV3+ leverages Atrous Spatial Pyramid Pooling (ASPP) to capture multi-scale contextual information. EfficientNet-B4 provides a strong feature extraction backbone with high parameter efficiency.

1.3 Training Setup

- Training Samples: 2857
- Validation Samples: 317
- Optimizer: AdamW
- Learning Rate: 1.5e-4
- Epochs: 25
- Batch Size: 4
- Mixed Precision Training Enabled

Loss Function:

$$\mathcal{L}_{total} = 0.5 \cdot \mathcal{L}_{Focal} + 0.5 \cdot \mathcal{L}_{Dice}$$

This hybrid loss improves class balance handling and segmentation overlap quality.

2 Results and Performance Analysis

2.1 Loss Convergence

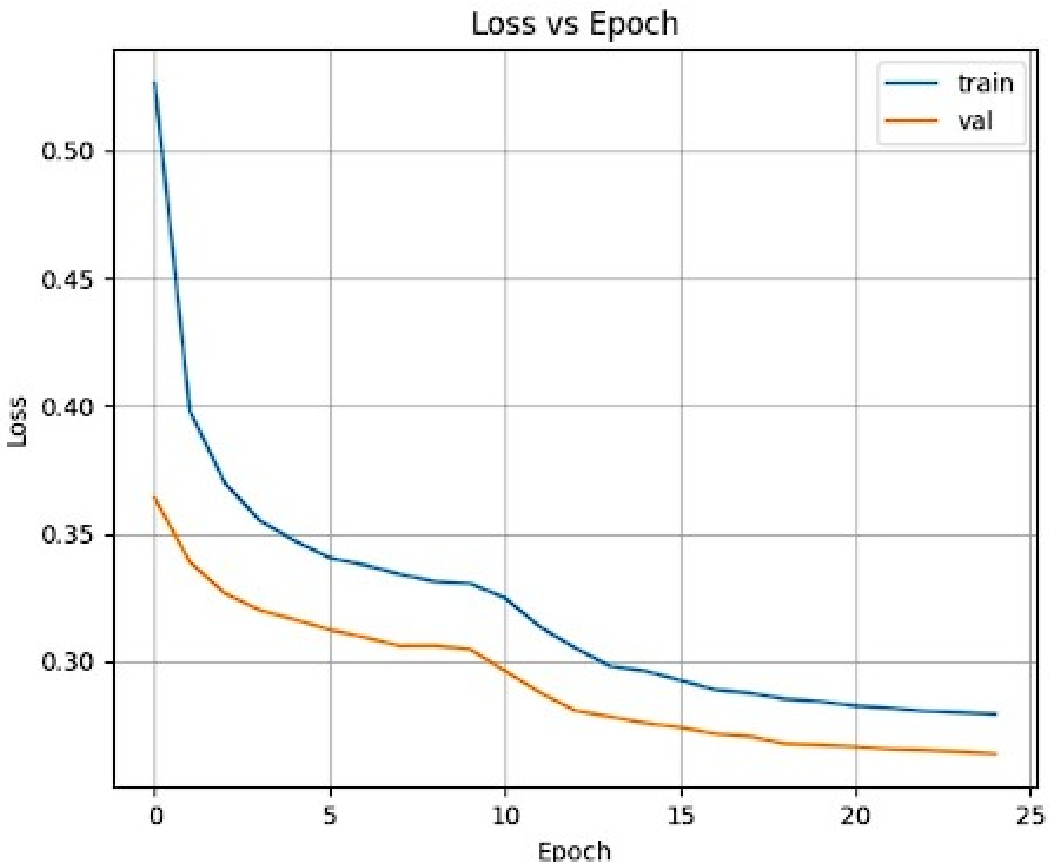

Figure 1: Training and Validation Loss vs Epoch

The loss curves show stable convergence with no divergence between training and validation, indicating good generalization and minimal overfitting.

2.2 Intersection over Union (IoU)

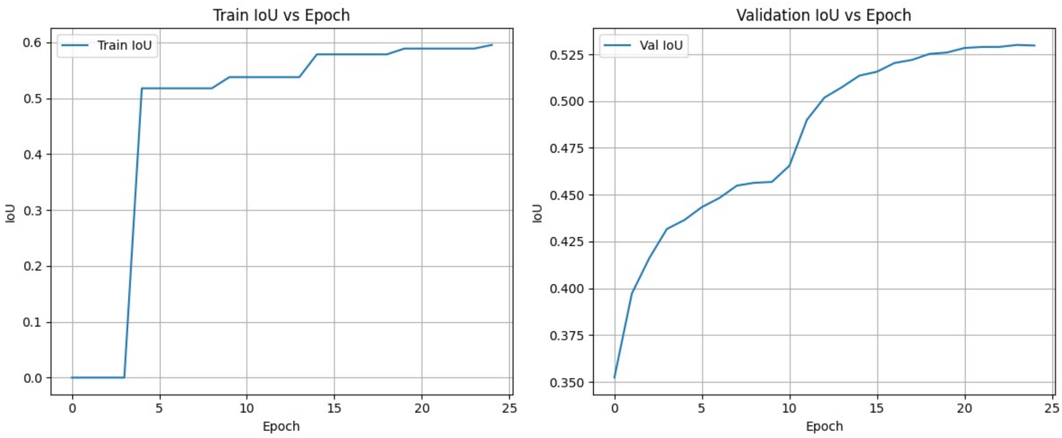

Figure 2: Training and Validation IoU vs Epoch

IoU steadily increases across epochs, reaching approximately 0.53 validation IoU. The small gap between training and validation curves indicates stable learning.

2.3 Dice Score

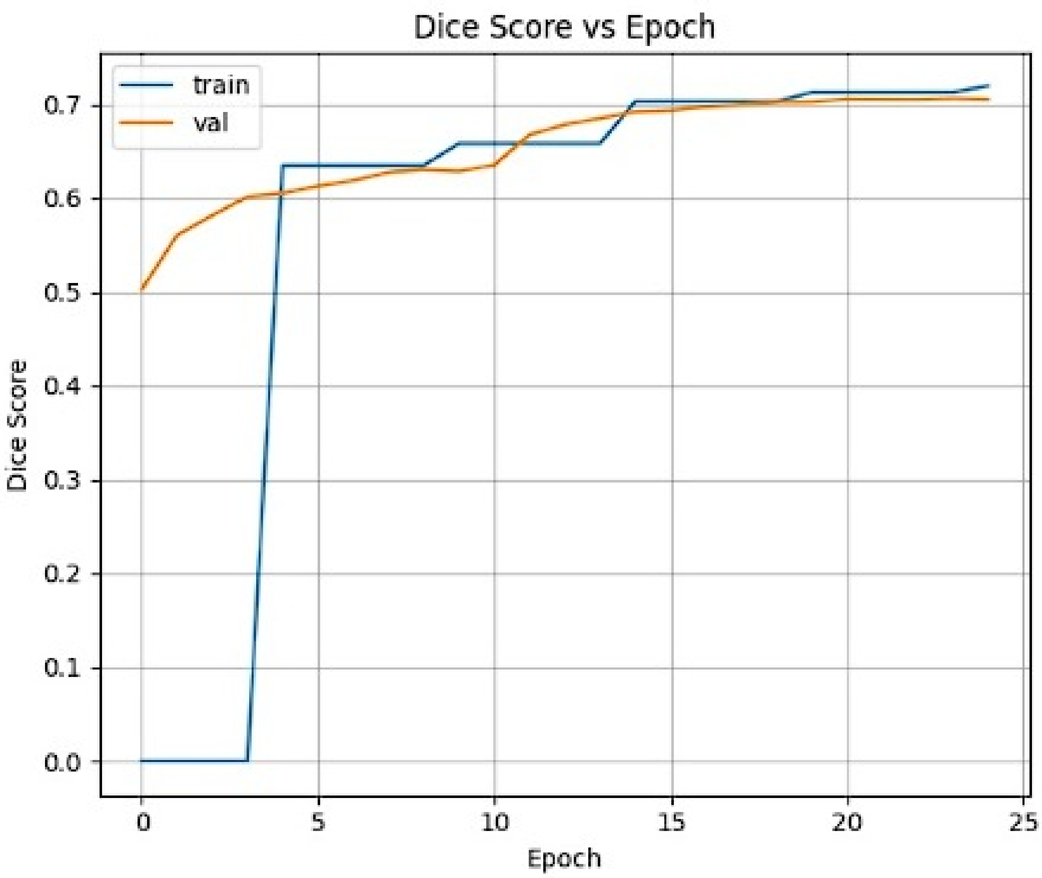

Figure 3: Training and Validation Dice Score vs Epoch

Dice score improves consistently, reaching approximately 0.70 on validation data, reflecting strong overlap performance across classes.

2.4 Pixel Accuracy

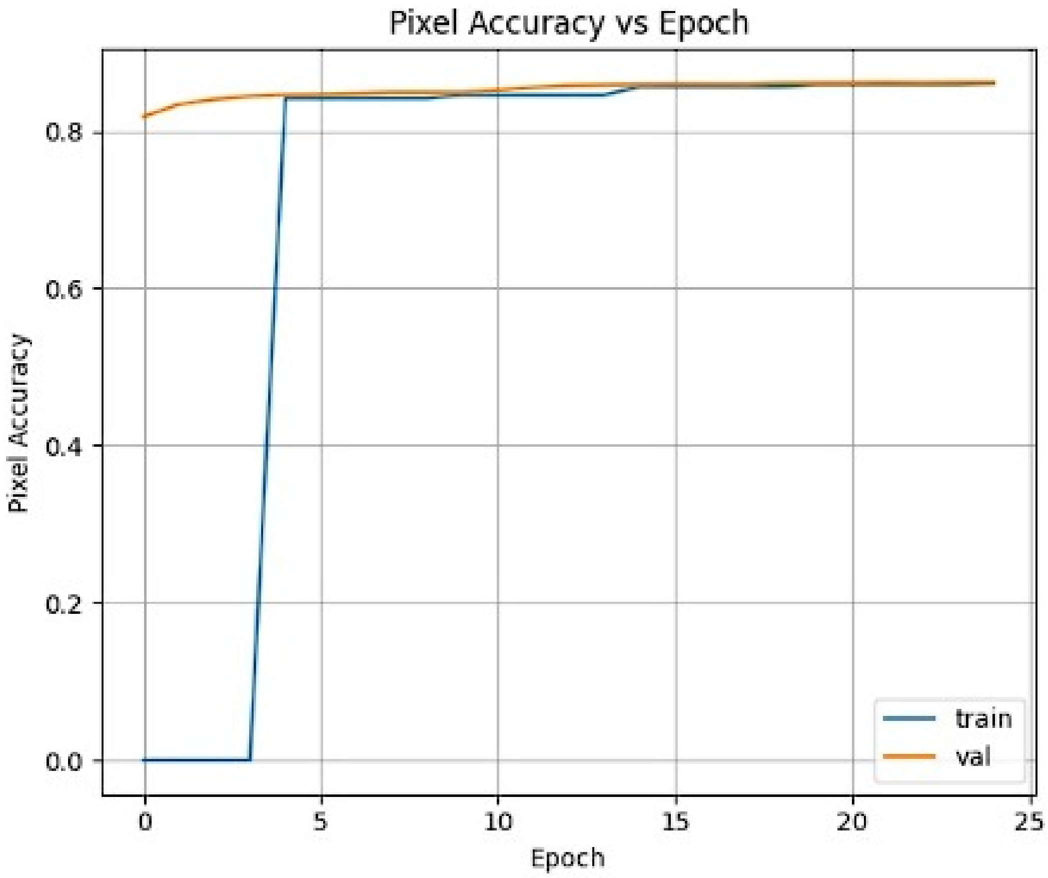

Figure 4: Training and Validation Pixel Accuracy vs Epoch

Pixel accuracy stabilizes above 86%, demonstrating reliable pixel-wise classification performance.

3 Challenges and Solutions

3.1 Class Imbalance

Certain terrain classes had fewer samples. This was addressed using Focal Loss combined with Dice Loss to improve minority class segmentation.

3.2 GPU Memory Constraints

EfficientNet-B4 required careful optimization. Mixed Precision Training and channels-last memory format were used to reduce memory footprint and improve throughput.

3.3 Checkpoint Management

Standardized checkpoint saving ensured reproducibility and stable inference loading.

4 Optimizations

- Automatic Mixed Precision (AMP)
- Test-Time Augmentation (Horizontal Flip TTA)
- Cosine Annealing Warm Restarts Scheduler
- Gradient Clipping
- Structured project directory for reproducibility

These optimizations improved convergence stability, training speed, and final validation metrics.

5 Conclusion

The proposed DeepLabV3+ EfficientNet-B4 segmentation system demonstrates strong performance on off-road desert terrain segmentation. The model achieves stable convergence, competitive IoU performance, and robust generalization.

The pipeline is modular, optimized, and reproducible, making it suitable for deployment in real-world terrain perception applications.