High-Accuracy Off-Road Semantic Segmentation using SegFormer-B4

Model Architecture: SegFormer-B4 (Transformer-based semantic segmentation)

Framework: PyTorch

Dataset: Off-Road Semantic Segmentation Dataset

TEAM : GEEKS

Team Members

1. Abhinav Raj 23MEI10016
2. Aryan Choubey 23MIM10173
3. Rathod Sagar 23MIM10166
4. Gokul Karthic 23MIM10098

PROBLEM STATEMENT

Autonomous vehicles operating in off-road environments require accurate understanding of terrain, obstacles, vegetation, water bodies, and drivable paths. Challenges include:

- Complex natural scenes
- Class imbalance
- Similar texture between terrain classes
- Variable lighting conditions

The goal was to build a semantic segmentation model capable of classifying each pixel into one of the dataset’s defined terrain classes.

APPROACH OVERVIEW

Our solution follows a deep learning pipeline consisting of:

Data preprocessing and label remapping
Strong but stable augmentation strategy
Transformer-based segmentation model (SegFormer-B4)
Combined loss function (Cross Entropy + Dice Loss)
Mixed precision training for efficiency
Validation using IoU and pixel accuracy metrics
Optimized inference pipeline for prediction generation

MODEL ARCHITECTURE

Selected Model – SegFormer-B4

The chosen architecture was:

SegFormer-B4 (Transformer-based semantic segmentation)

Reasons for selection:

- Strong performance on segmentation benchmarks
- Efficient memory usage
- Excellent balance between speed and accuracy
- Works well with medium-sized datasets through transfer learning

Architecture Highlights

- Hierarchical transformer encoder
- Multi-scale feature fusion
- Lightweight MLP decoder
- No heavy positional encodings

Training Configuraton

Parameter	Value
Model	SegFormer-B4
Image Size	512 × 512
Batch Size	2
Gradient Accumulation	2
Effective Batch Size	4
Epochs	12
Optimizer	AdamW
Learning Rate	6e-5
Scheduler	Cosine Annealing
Precision	Mixed (AMP)

RESULT

The model showed stable and consistent improvement across epochs.

Performance Summary

Epoch.	IoU	Accuracy
1	0.4431	0.8454
2	0.4928	0.8575
3	0.5210	0.8625
4	0.5363	0.8671
5	0.5439	0.8682
6	0.5537	0.8696
7	0.5557	0.8710
8	0.5595	0.8715
9	0.5608	0.8723
10	0.5622	0.8726
11	0.5642	0.8729
12	0.5639	0.8730

Final Best Model

Best IoU: 0.5642

Pixel Accuracy: 87.3%

Learning Behavior

Observed patterns:

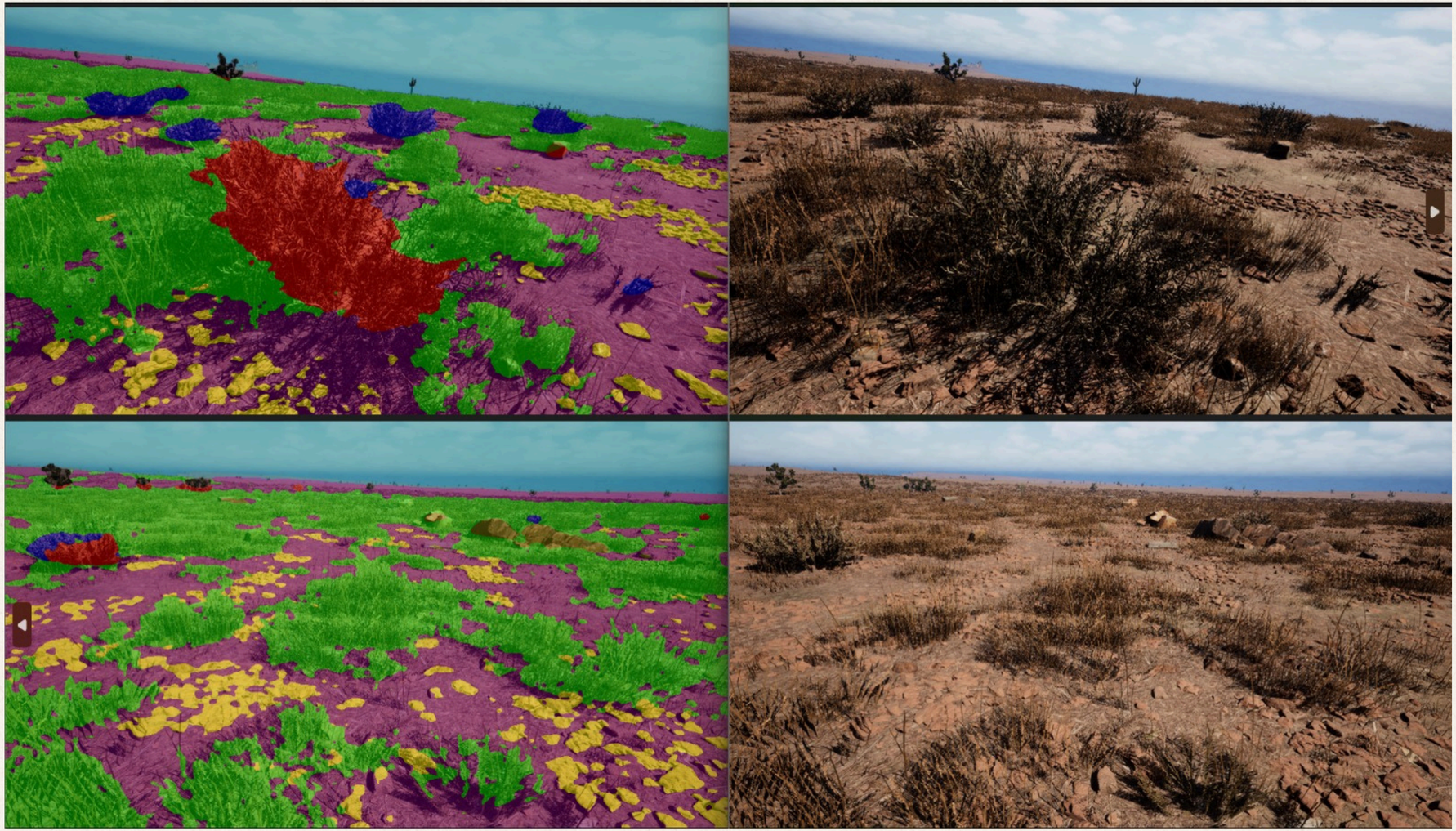
Rapid early learning (epochs 1–4)

Gradual refinement of boundaries

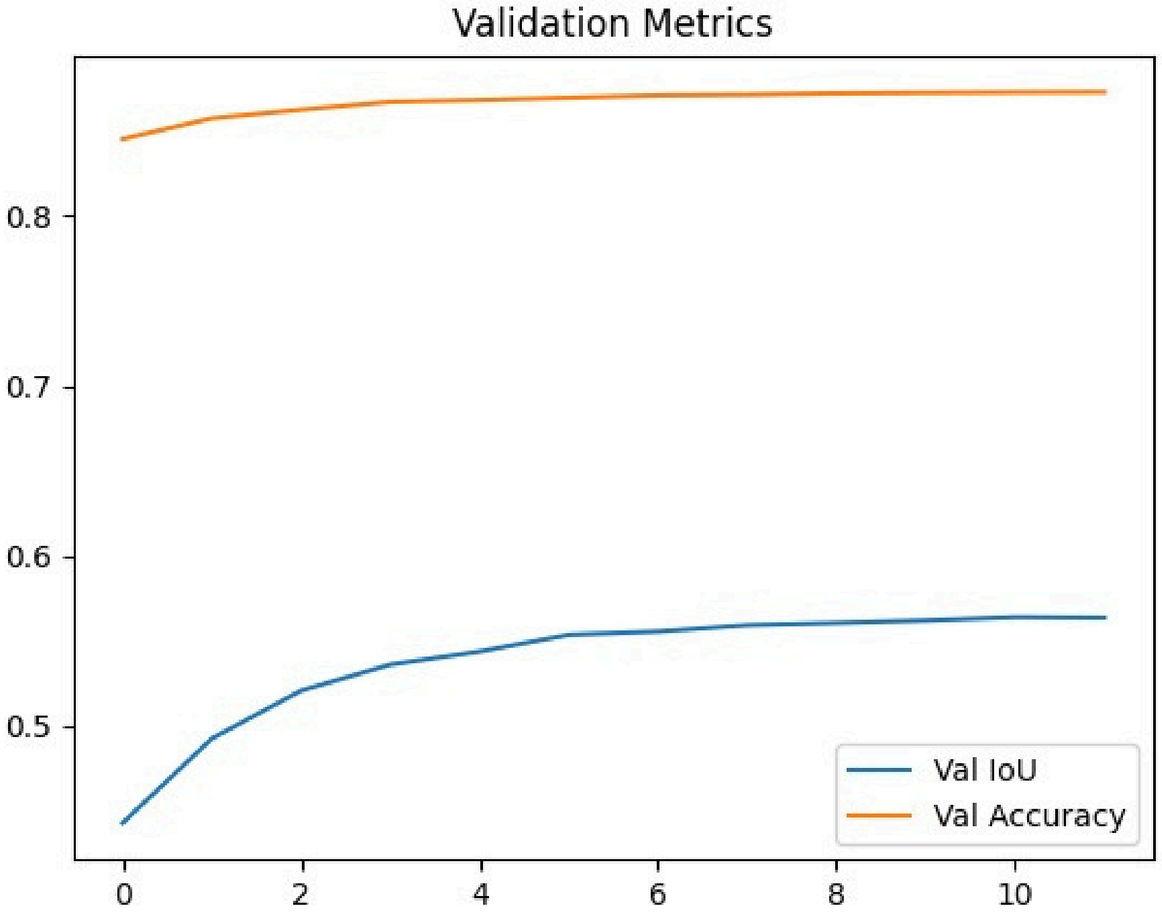
Plateau around epoch 10–12

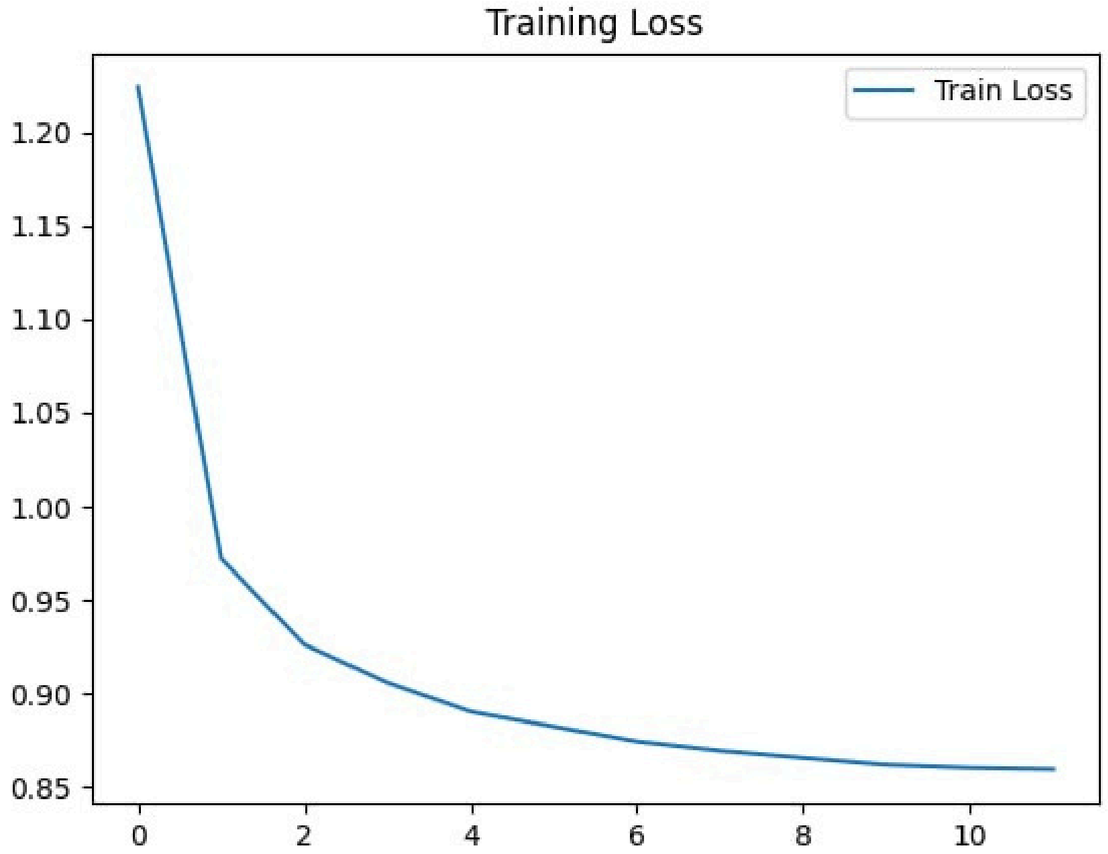
Training curves confirmed stable convergence without overfitting.

RESULT



GRAPHS


Validation Metrics
0.8
0.7
0.6
0.5
0
2
4
6
8
10
Val IoU
Val Accuracy


Training Loss
Train Loss
1.20
1.15
1.10
1.05
1.00
0.95
0.90
0.85
0
2
4
6
8
10

COMPARISON

We trained and evaluated three advanced semantic segmentation models on the Offroad Segmentation Dataset:

- DeepLabV3+ (ResNet38)
- DeepLabV3+ (ResNet50)
- SegFormer-B4

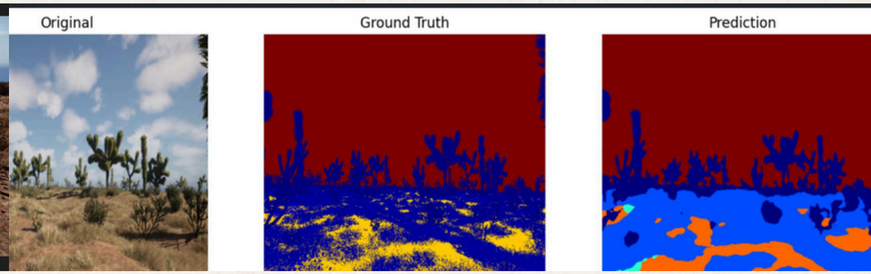
Quantitatively, the DeepLabV3+ variants achieved slightly higher IoU scores compared to SegFormer-B4. However, during qualitative evaluation, we observed that SegFormer produced more semantically consistent and visually accurate predictions, particularly in complex terrain and boundary regions.

While IoU is an important metric, our application prioritizes real-world reliability, structural consistency, and accurate terrain understanding. SegFormer demonstrated better generalization and more stable predictions in challenging scenarios.

Therefore, despite a marginally lower IoU score, we selected SegFormer-B4 as our final model because it solved the core problem more effectively and aligned better with the project’s objective.

COMPARISON

SegFormer-B4 DeepLabV3+ ResNet38 DeepLabV3+ ResNet50



backbone	**Transformer**	**ResNet38**	**ResNet50**
IoU	**Best IoU Achieved: 0.564**	**Best IoU Achieved: 0.857**	**Best IoU Achieved: 0.658**
strength	**Better semantic consistency, strong global context modeling, stable predictions in complex scenes**	**Highest quantitative IoU, strong pixel-wise accuracy, effective multi-scale feature extraction**	**Deeper encoder improves feature richness over smaller backbones**
weakness	**Less consistent boundaries in complex terrain , sensitive to class similarity**	**Higher complexity with limited performance gain, unstable in boundary regions**	**Lower IoU compared to CNN-based models**

CHALLENGES, SOLUTIONS AND RESOLUTION

1 .Class Imbalance

Off-road datasets contain dominant classes (soil, vegetation) and sparse classes (stones, obstacles), which can bias training.

Solution:

A hybrid loss combining Cross Entropy and Dice Loss was used to improve learning for small regions.

2 .Visual Similarity Between Terrain Classes

Classes such as grass, shrubs, and soil show similar textures, causing boundary confusion.

Solution:

A transformer-based SegFormer-B4 model was adopted to capture global context, supported by Dice Loss for better boundary learning.

3 .Lighting and Environmental Variations

Different lighting and weather conditions affected model generalization.

Solution:

Controlled appearance-based augmentations (brightness, contrast, hue, blur) were applied during training.

4 . Training Stability and Augmentation Overuse

Initial experiments with aggressive augmentations resulted in unstable learning and sudden IoU degradation, particularly during early epochs.

Solution

Augmentation intensity was carefully tuned to strike a balance between diversity and stability. Only transformations that preserved semantic meaning were retained, and extreme distortions were removed.

5 . Computational Constraints and Training Efficiency

Transformer-based models are computationally intensive, and training at high resolutions significantly increases memory usage and training time.

Solution

Several optimization techniques were implemented:

- Mixed Precision Training (AMP)
- Gradient Accumulation to achieve effective batch size of 4
- AdamW optimizer with cosine annealing scheduler
- CuDNN benchmarking and non-blocking GPU transfers

Image resolution was fixed at 512 × 512, providing a balance between spatial detail and computational feasibility.

6 . Windows DataLoader Multiprocessing Issues

Training on a Windows environment initially caused DataLoader crashes due to multiprocessing worker spawning issues.

Solution

The training script was restructured using a safe main entry point (if __name__ == "__main__":) and worker settings were adjusted to ensure compatibility.

CONCLUSION

This work demonstrates that transformer-based segmentation with transfer learning can effectively address off-road terrain segmentation challenges. The final SegFormer-B4 model achieved a best IoU of 0.5642 and 87.3% pixel accuracy, showing stable convergence and strong qualitative results. The developed pipeline is efficient, reproducible, and suitable for autonomous navigation research.

FUTURE WORK

- Future improvements include:
- Multi-scale training and inference
- Test-time augmentation
- Class-balanced and boundary-aware losses
- Larger transformer models
- Exploration of Mask2Former for enhanced global context