OFF-ROAD DESERT

SEMANTIC SEGMENTATION

Pixel-wise terrain labelling for synthetic off-road desert scenes

using ResNet50 + Feature Pyramid Network Decoder

Team AIDHUNIK · Hack For Green Bharat · February 26, 2026

The Problem

Why terrain perception matters for autonomous off-road systems

Harsh Terrain Diversity

10 semantically distinct terrain classes that look visually similar in desert environments make classification hard.

Sim-to-Real Gap

Synthetic data has distribution shifts vs. real imagery. Models must generalise across lighting, texture, and geometry.

Class Imbalance

Sky and ground dominate pixel counts; rare classes like bushes/rocks are under-represented, biasing naïve models.

Dataset

Duality AI — Synthetic Desert Off-Road Scenes

2 857

Training
images

317

Validation
images

10

Semantic
classes

476×266

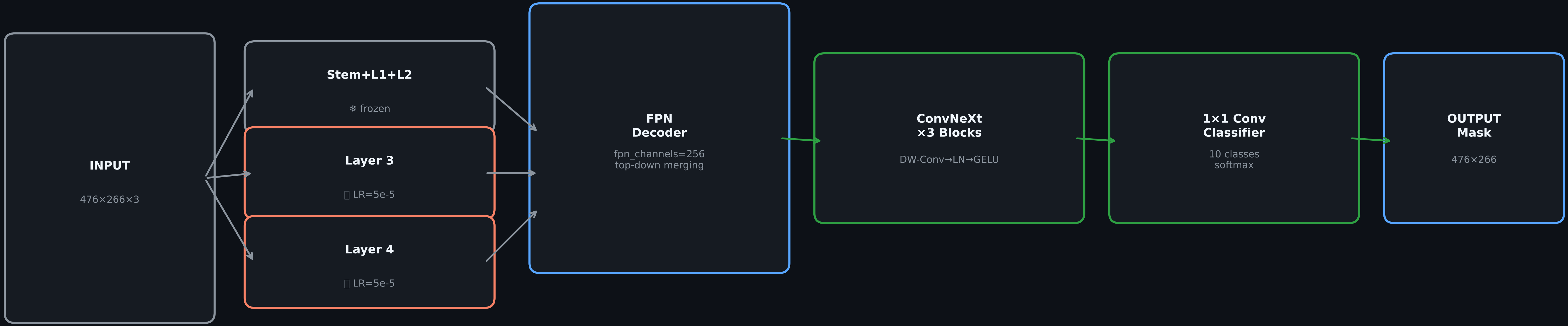
Image
resolution

Mask pixel label mapping:

● Background (raw=0)	● Trees (raw=100)	● Dry Grass (raw=200)	● Rocks (raw=300)	● Bushes (raw=500)
● Sky (raw=550)	● Ground Clutter (raw=700)	● Terrain (raw=800)	● Shrubs (raw=7100)	● Trail (raw=10000)

Architecture

ResNet50 Multi-Scale Backbone + FPN Decoder + ConvNeXt Head



INPUT

476×266×3

Stem+L1+L2

❄ frozen

Layer 3

□ LR=5e-5

Layer 4

□ LR=5e-5

FPN Decoder

fpn_channels=256
top-down merging

ConvNeXt ×3 Blocks

DW-Conv→LN→GELU

1×1 Conv Classifier

10 classes
softmax

OUTPUT Mask

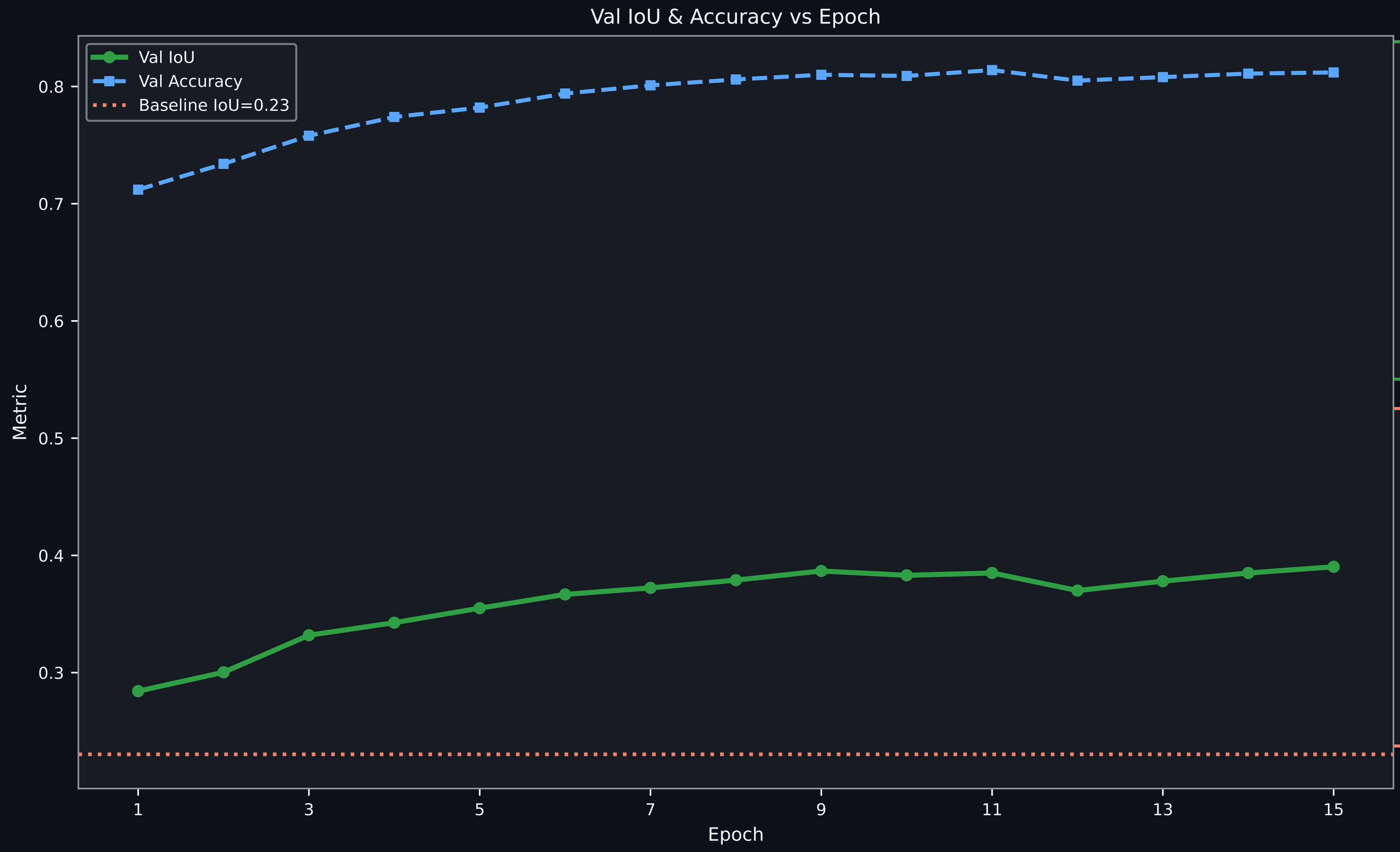
476×266

Loss = CrossEntropy(class-weighted, label-smooth=0.05) + 0.5 × SoftDice

Optimiser: AdamW · Scheduler: CosineAnnealingLR T_0=10 · Head LR=5e-4 · Backbone LR=5e-5

Training Results

V3 FPN — 15 epochs on single T4 GPU



0.3903

Best Val IoU

0.5615

Best Val Dice

81.2%

Val Accuracy

+70%

IoU vs Baseline

Model Comparison

Baseline v1 → V3 FPN

Model	Epochs	Val IoU	Val Dice	Val Accuracy	Val Loss
Baseline v1 (frozen backbone)	10	0.2305	0.3633	66.08%	0.96
V3 FPN ★ (partial fine-tune)	15	0.3903	0.5615	81.20%	1.79
Δ improvement	—	+69.6%	+54.5%	+22.9 pp	—

Key Innovations

What made V3 outperform the baseline by 70%

1 Multi-Scale FPN

Fuse features from layer2, layer3, layer4
Captures both fine detail and coarse semantics
Top-down pathway with lateral skip connections

2 Partial Backbone Fine-Tuning

Layer3 + Layer4 unfrozen at LR=5e-5
Adapts ImageNet weights to desert textures
10× lower LR than head to avoid forgetting

3 Combined Loss + Class Weights

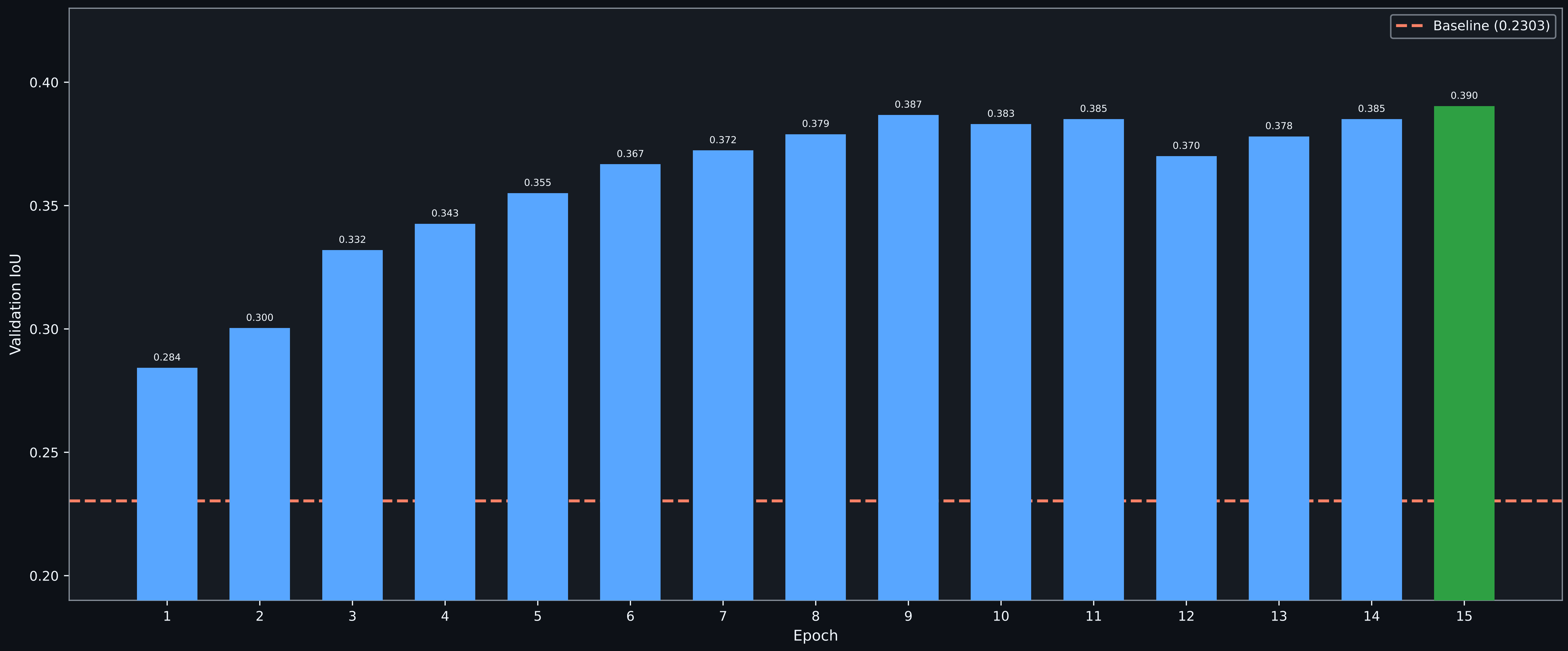
CE (label-smooth=0.05) + 0.5 × SoftDice
Inverse-frequency class weights tackle imbalance
Dice loss directly optimises the IoU metric

4 ConvNeXt-Style Head

3 × ConvNeXt blocks: DW-Conv → LN → GELU
More expressive than plain conv stacks
CosineAnnealing LR restarts every T_0=10 epochs

Per-Epoch Val IoU

Training progression over 15 epochs



Best: Epoch 15 → Val IoU = 0.3903 (+69.6% over baseline)

Submission Package

What we're delivering

SUBMISSION/src/	train_segmentation.py · test_segmentation.py · visualize.py
models/	segmentation_head_best.pth (120 MB — best checkpoint, epoch 15)
train_stats/v3/	evaluation_metrics_v3.txt · 4 training curve plots (PNG)
README.md & REPORT.md	Architecture docs, results table, setup & usage instructions
requirements.txt	torch · torchvision · numpy · opencv-python · albumentations · tqdm
GitHub Repo	github.com/Akum030/duality-desert-segmentation

Thank You!

Team AIDHUNIK

0.3903	0.5615	81.2%	+70%
Val IoU	Val Dice	Val Acc	Δ IoU

github.com/Akum030/duality-desert-segmentation

Hack For Green Bharat · February 26, 2026