Training Strategies for U-Net: A Performance Analysis

Ziliang Song, Pintzu Tseng, Dongni Li

Final Project Fall 2025 - DATA-645-001

01

Introduction

- Semantic segmentation is essential for autonomous driving
- Goal: Automatically segment car regions in urban scenes
- U-Net chosen for its strong pixel-level segmentation performance
- Focus: Analyze how training strategies affect accuracy & generalization

Background

Dataset: Carvana Image Masking Challenge (high-resolution car images)

01. Task: Pixel-level segmentation of vehicle contours

02. Challenges: varied lighting, reflections, car shapes

03. Importance:accurate car segmentation is critical for autonomous driving perception

02

Methodology: Baseline

- Classic encoder–decoder U-Net structure
- Skip connections to preserve spatial details
- Trained from scratch on Carvana dataset
- Serves as baseline for comparison with augmented & transfer learning models

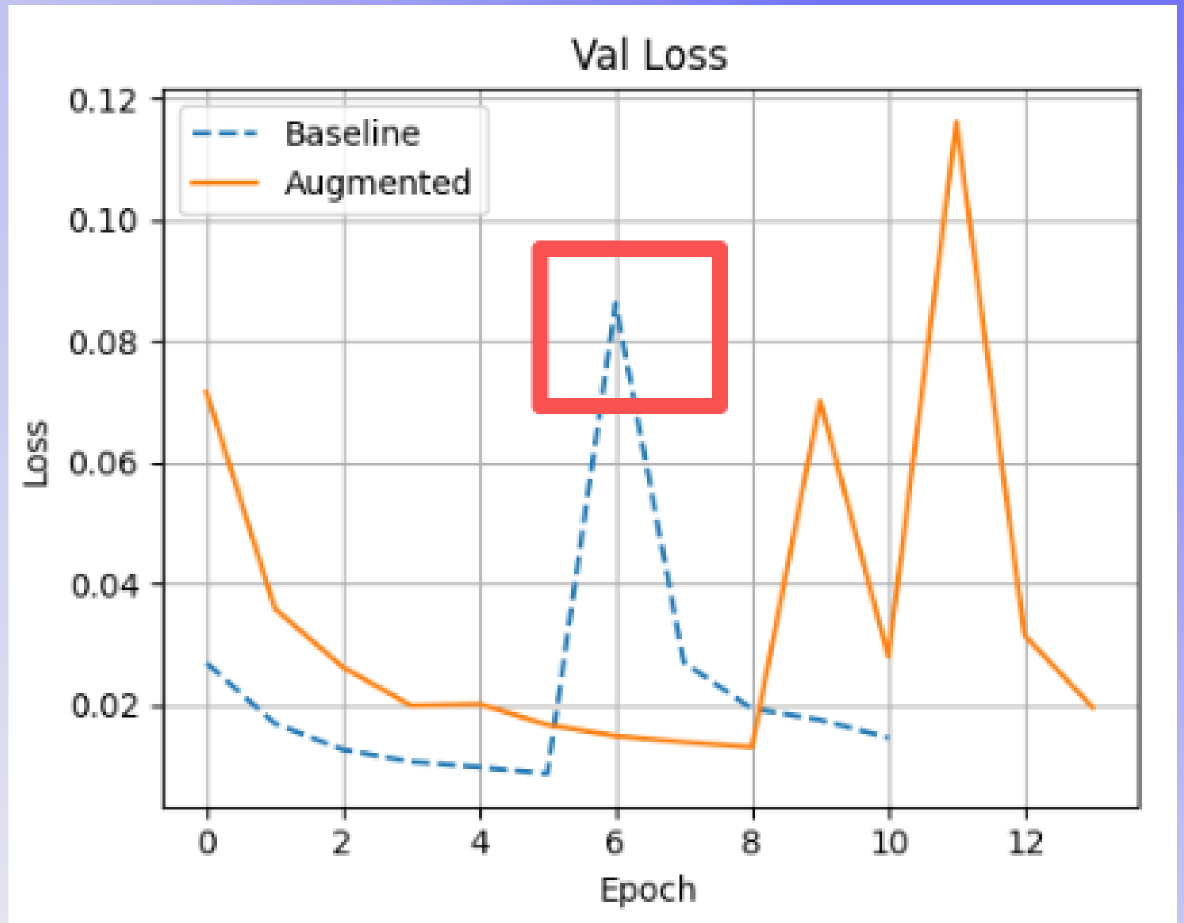
03

Methodology: Augmentation

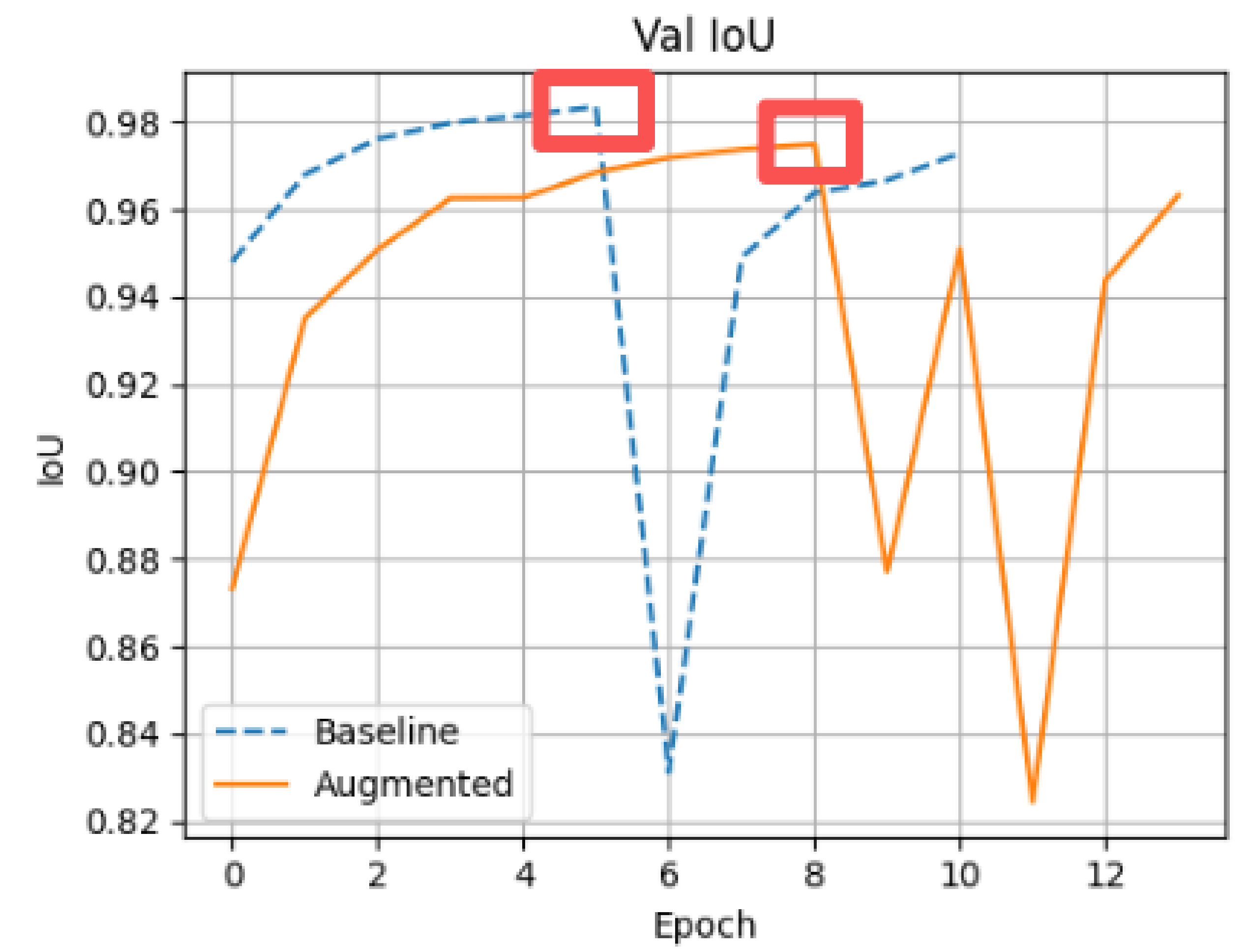
- Improve robustness & avoid loss spikes
- Increase training diversity
- Geometry: flips, 90°/180° rotations
- Color: brightness & contrast changes
- U-Net architecture unchanged

04

Results: Augmentation vs. Baseline



- Baseline shows a severe loss spike at Epoch 7
- Augmented model removes early loss spikes; late-epoch noise is controlled by Early Stopping
- Slight IoU drop: 0.9833 → 0.9746
- Robustness improved significantly

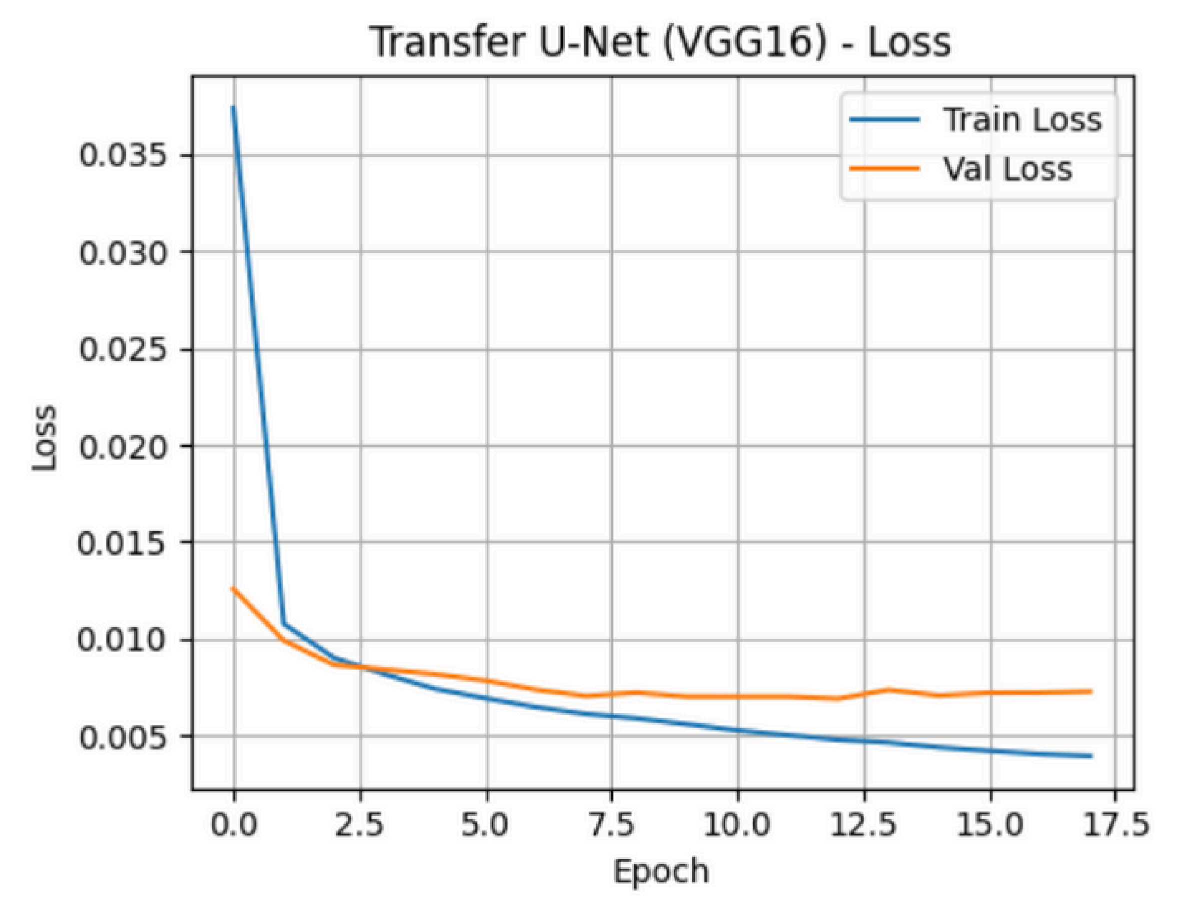


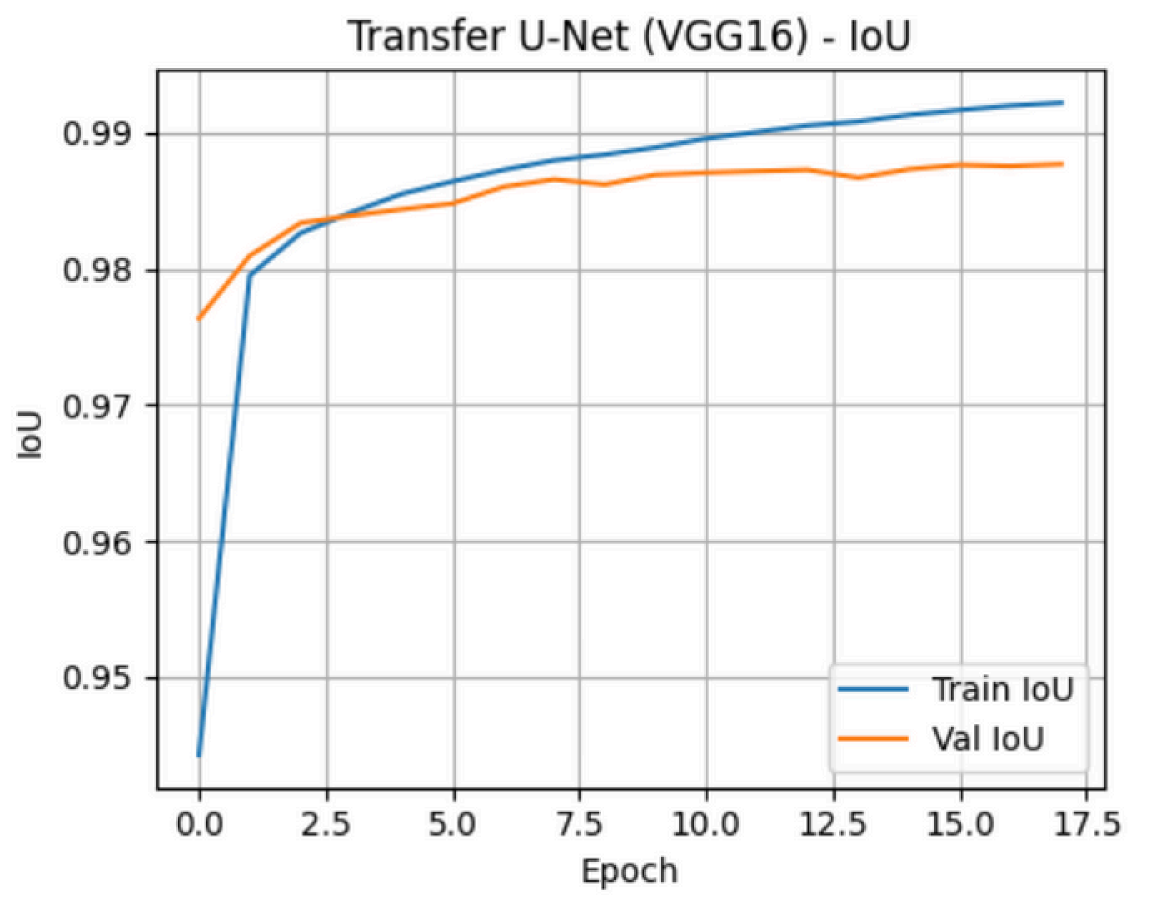
Small accuracy drop → major stability gain

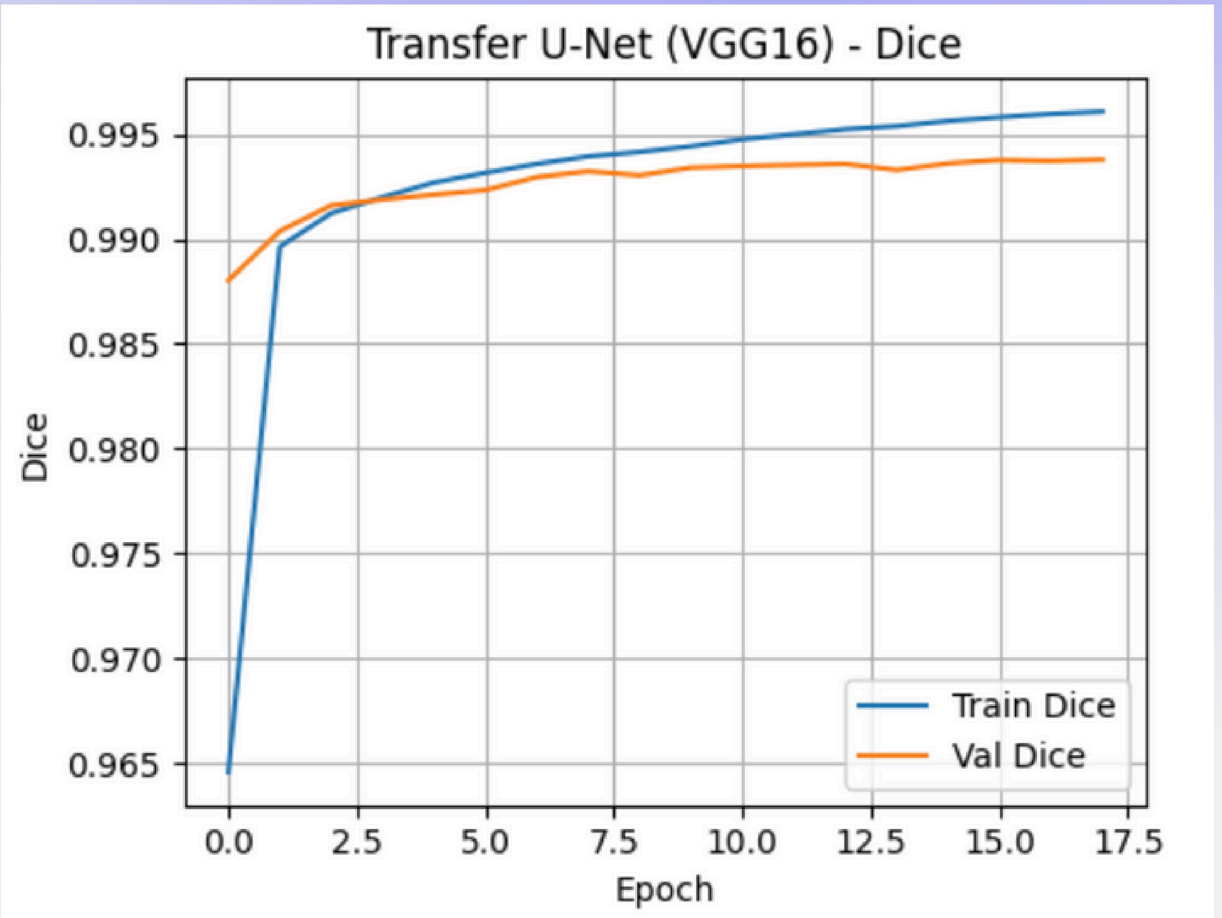
Transition: Can we achieve BOTH? → Transfer Learning

Results: Transfer Learning

- Highest accuracy: Pixel Accuracy = 0.9973,
- Dice = 0.9936, IoU = 0.9872
- Stable training: smooth loss curves with no spikes
- More efficient: 26% faster per training step than Baseline

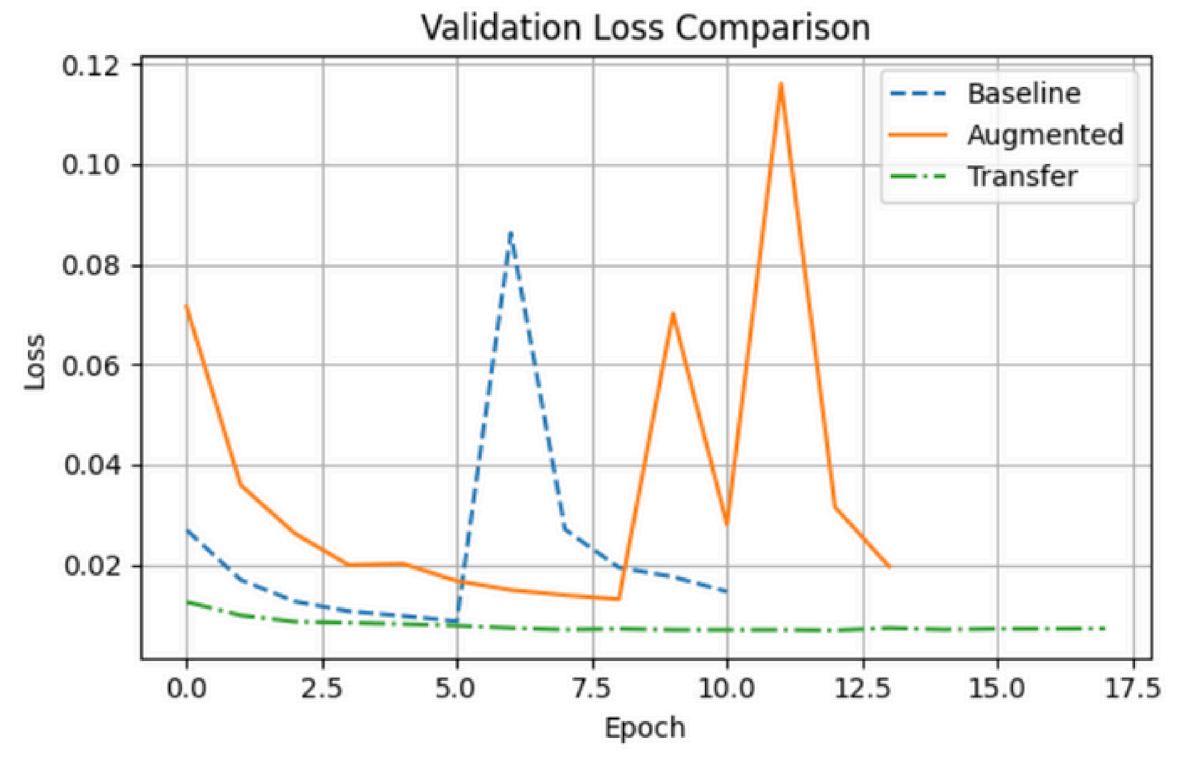


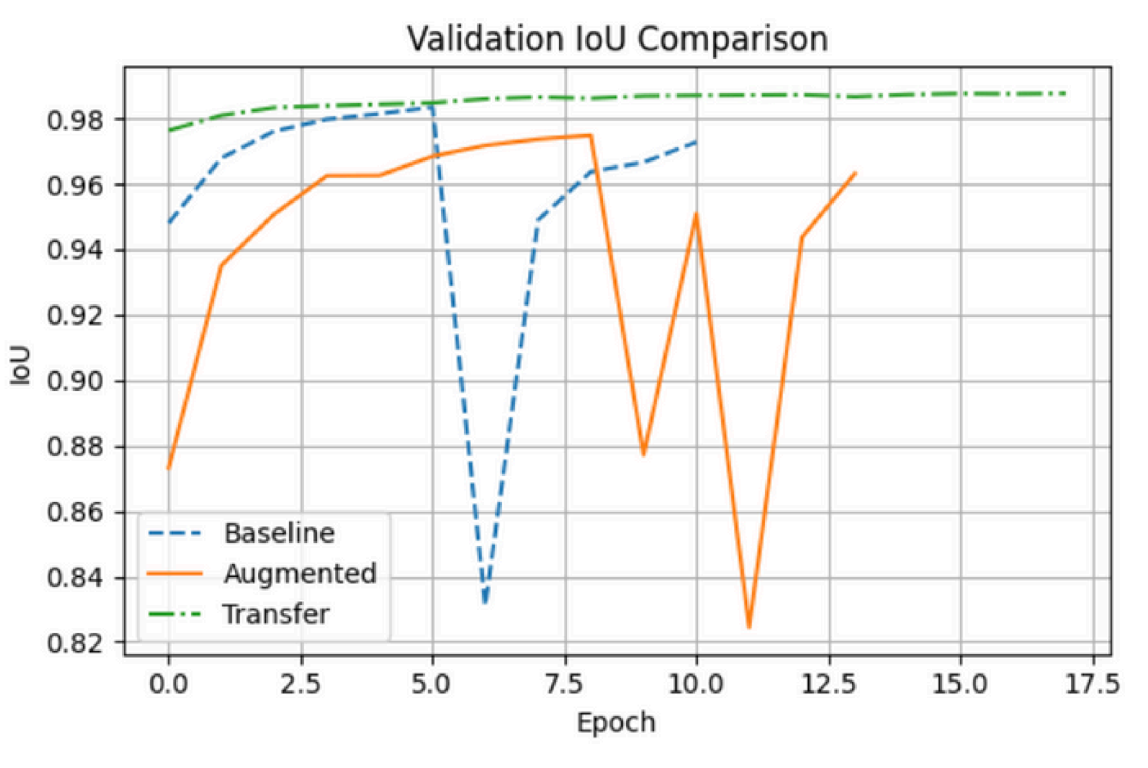


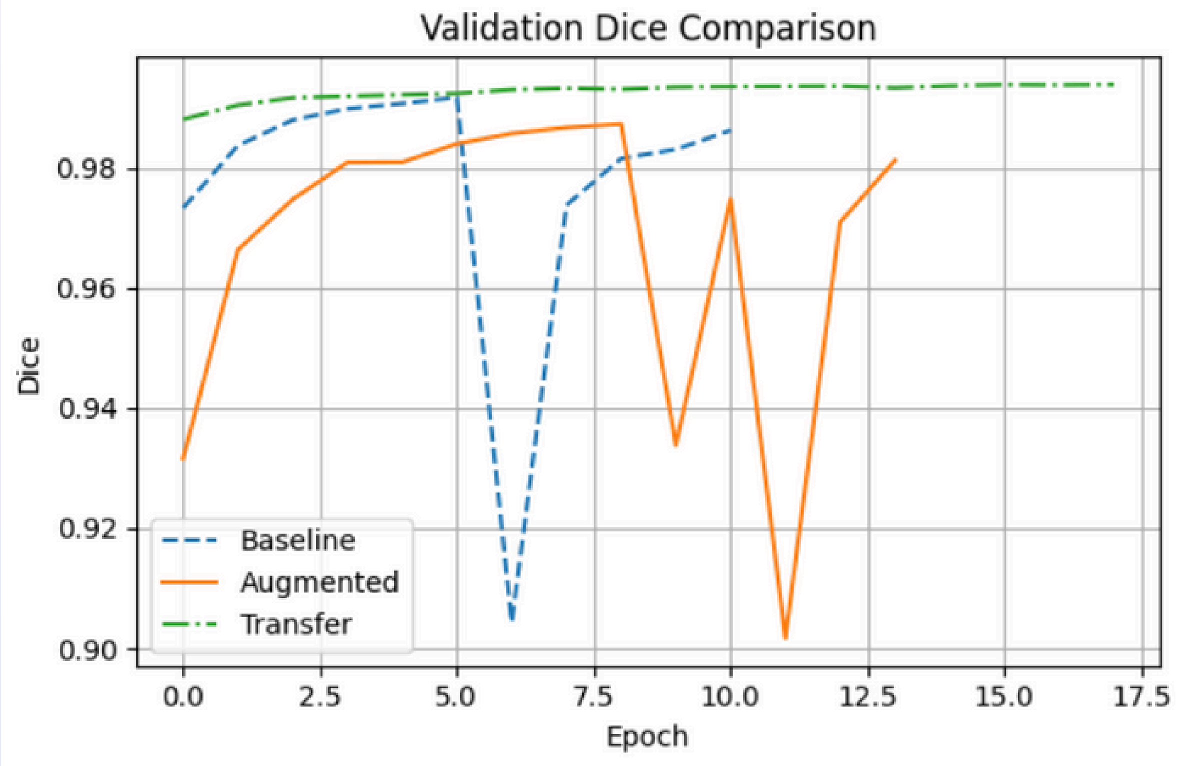


Discussion: Comparison

- Baseline model shows unstable training with large spikes in loss and sudden drops in IoU and Dice.
- Augmented model improves early stability but still experiences oscillations, especially near convergence.
- Transfer Learning model achieves the most stable curves, with smooth loss reduction and consistently high IoU and Dice.
- Overall, Transfer Learning provides the best accuracy and robustness of the three strategies.

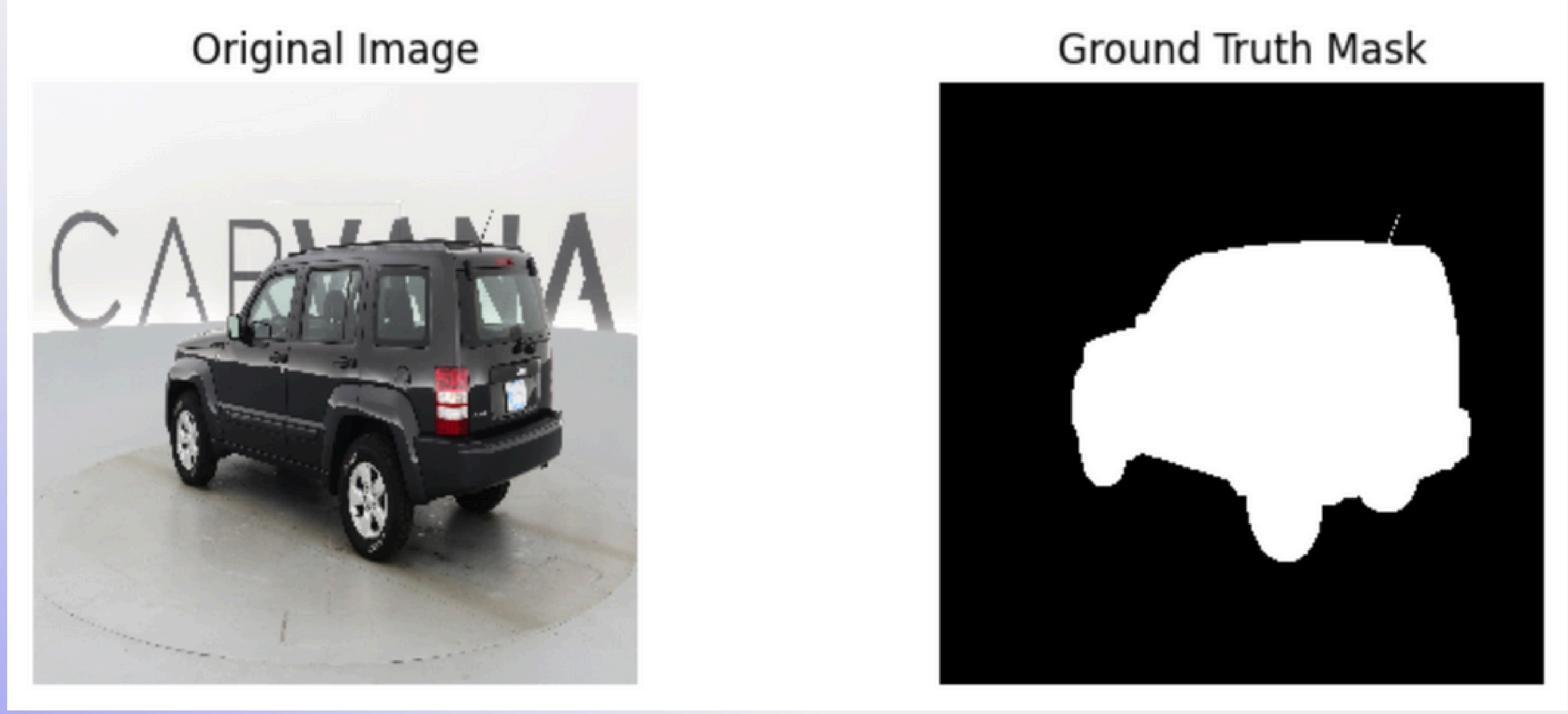


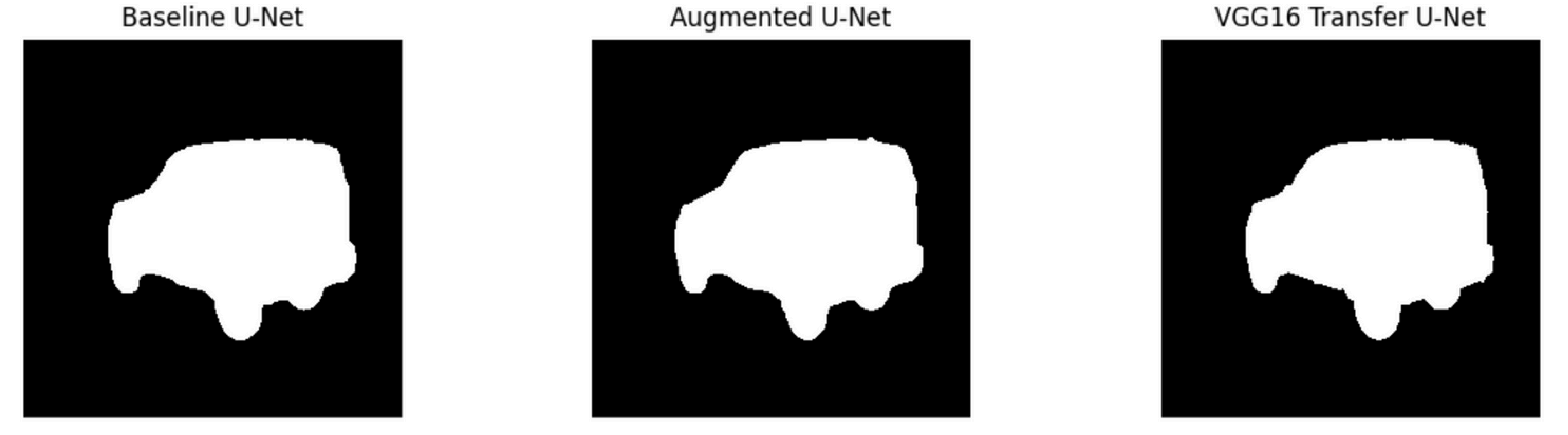




Discussion: Qualitative Analysis

- Baseline: correct shape but noisy/uneven edges
- Augmented: smoother masks, but loses fine details
- Transfer Learning: sharpest boundaries and closest to ground truth
- Qualitative results match the quantitative trends





9

Conclusion

Key Contribution

- U-Net + VGG16 Transfer Learning provides the best overall performance
- Achieves highest accuracy (IoU ≈ 0.9872)
- Fastest training efficiency (~235 ms / step)
- Resolves Baseline instability and Augmentation accuracy trade-off

Why It Matters

- Most stable training behavior among all models
- Strong generalization on a clean segmentation benchmark
- Practical choice for real-world deployment

Future Work

- Test VGG16 U-Net on complex real-world datasets
  - Cityscapes
  - BDD100K
- Evaluate generalization under diverse lighting, weather, and occlusion
- Explore other pretrained backbones
  - ResNet50, EfficientNet, ConvNeXt
- Experiment with more advanced augmentations
  - CutMix, MixUp, random occlusions
- Deploy in real-time inference setups

Thank You