Artificial Intelligence for Medical Image Segmentation

Improving Robustness with Sharpness-Aware Optimization and Balanced Loss

Liangqiao Gui[1], Din Barrameda[2], Zihao Li[3] Xin Wang[1]

[1] College of Integrate Health Sciences, [2] College of Emergency Preparedness, [3] College of Arts and Sciences

Introduction & Motivation

Problem

- 3D medical image segmentation suffers from poor **generalization**
- Models may converge to sharp minima during training
- Severe class imbalance exists in voxel distribution
- Foreground structures (e.g., organs, lesions) occupy very few voxels
- Standard voxel-wise losses are dominated by background

Motivation

- Models tend to under-segment small structures
- Training may converge to sharp minima, leading to poor generalization
- Severe class imbalance further weakens foreground learning

Key Challenge

- Mismatch between: **Highly imbalanced data distribution** and **uniform loss design**
- Leads to: Biased optimization toward background and weak learning signal for foreground

Our Goal

- Improve robustness in segmentation fine-tuning
- Address two complementary challenges:

Optimization → **Sharpness-Aware Minimization**

Imbalance → **Balanced Loss**

- Compare their effectiveness

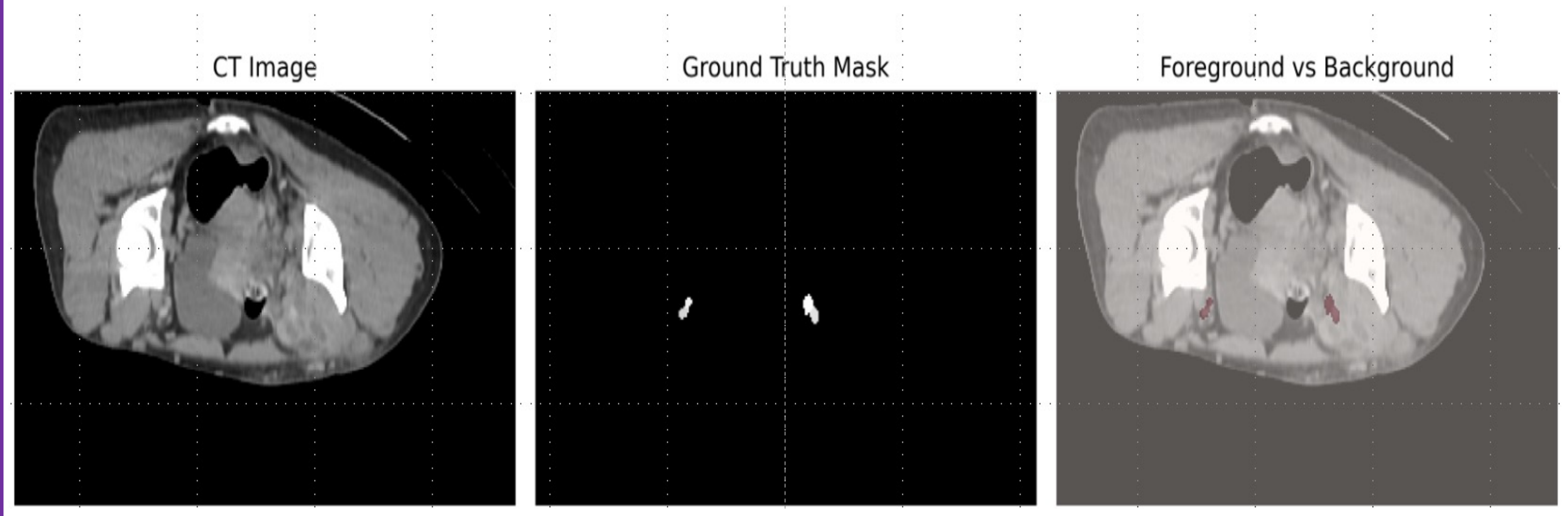

Figure 1. Example of severe foreground–background imbalance in medical image segmentation.
Foreground structures occupy only a very small fraction of voxels (FG ≈ 0.19%), while background dominates (BG ≈ 99.8%).
This imbalance makes standard training biased toward background.

Method

Baseline

- SegVol for 3D medical image segmentation
- Fine-tuned on downstream datasets

Sharpness-Aware Minimization

- Seeks flat minima for better generalization
- Replace standard optimizer with SAM during fine-tuning
- Improves generalization and robustness

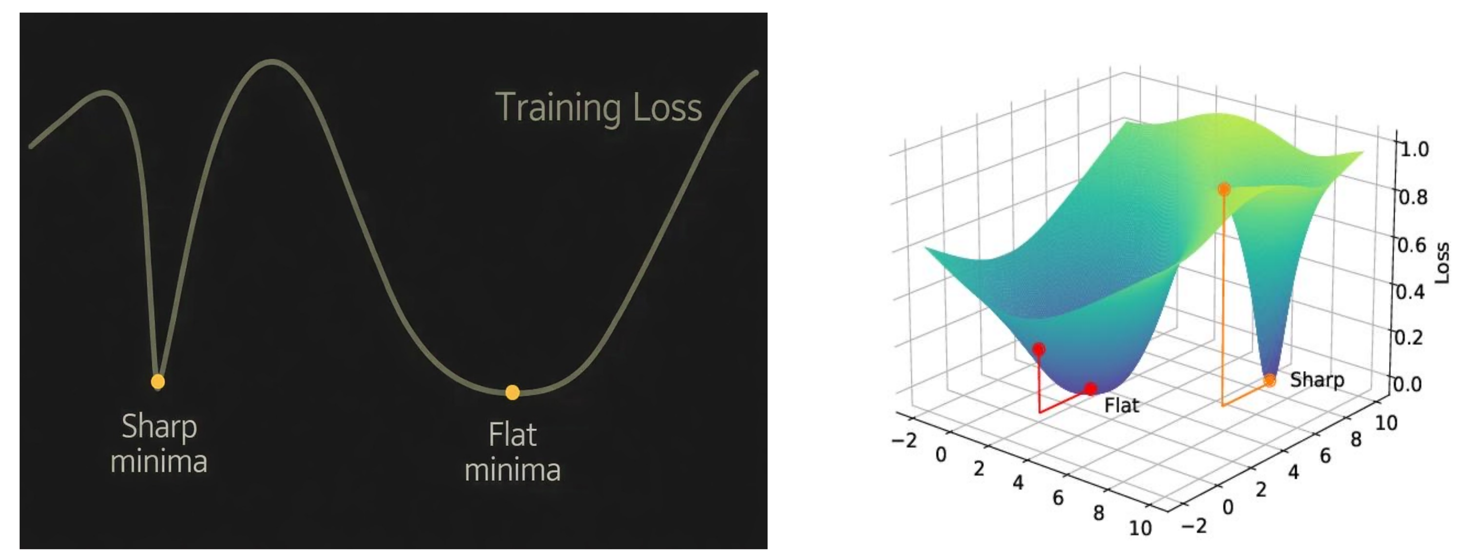

Figure 2. Illustration of sharp and flat minima in the loss landscape.
Left: 2D visualization of sharp vs. flat minima.
Right: 3D loss surface representation.
SAM seeks flatter minima to improve generalization.

Balanced Loss

- Designed for severe foreground–background imbalance
- Standard voxel-wise losses(BCE) are dominated by background
- Logit-adjusted BCE combined with Dice loss
- Improves foreground learning and small structure segmentation

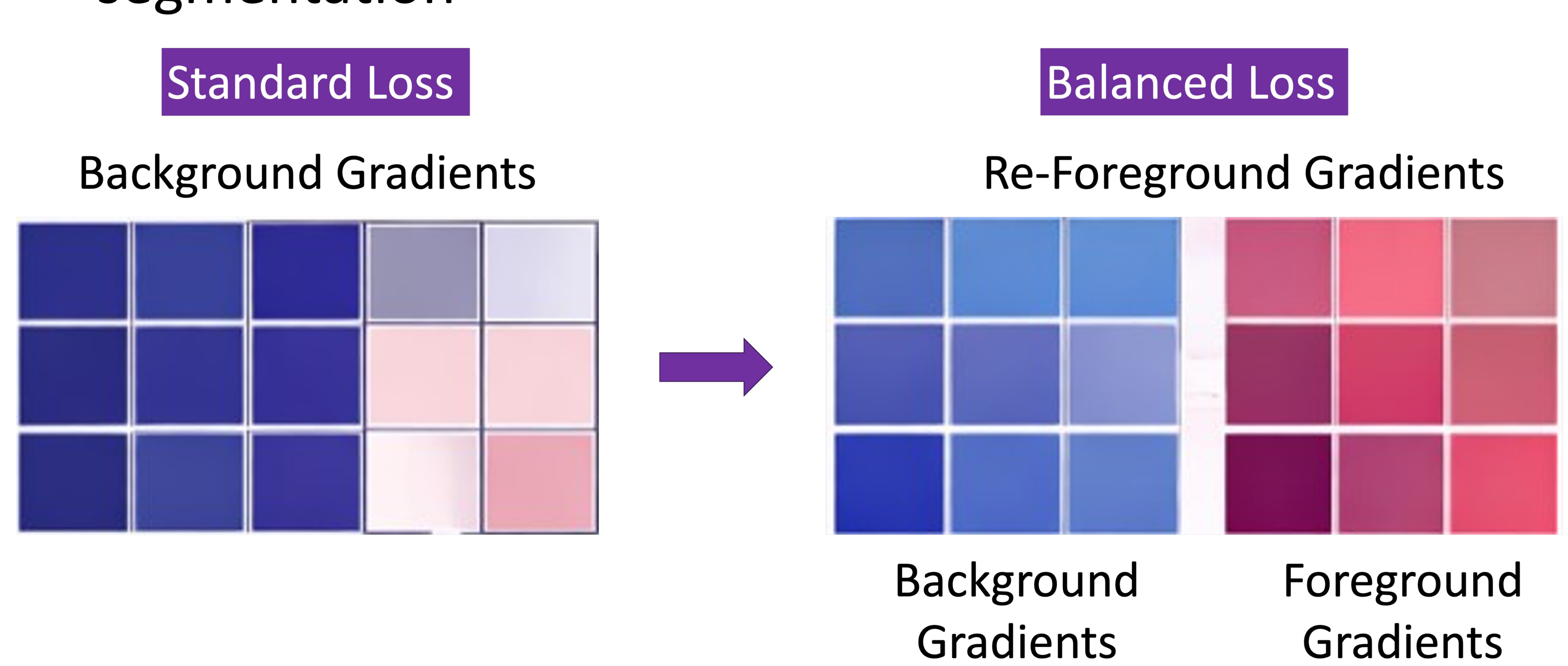

Figure 3. Illustration of foreground–background imbalance and the effect of Balanced Loss.
Left: Standard voxel-wise loss (BCE) is dominated by background gradients, leading to weak foreground learning.
Right: Balanced Loss reweights gradients to strengthen foreground learning and improve small structure segmentation.

Results

Average Dice is computed across validation cases, representing overall segmentation performance. Balanced Loss achieves the highest average score.

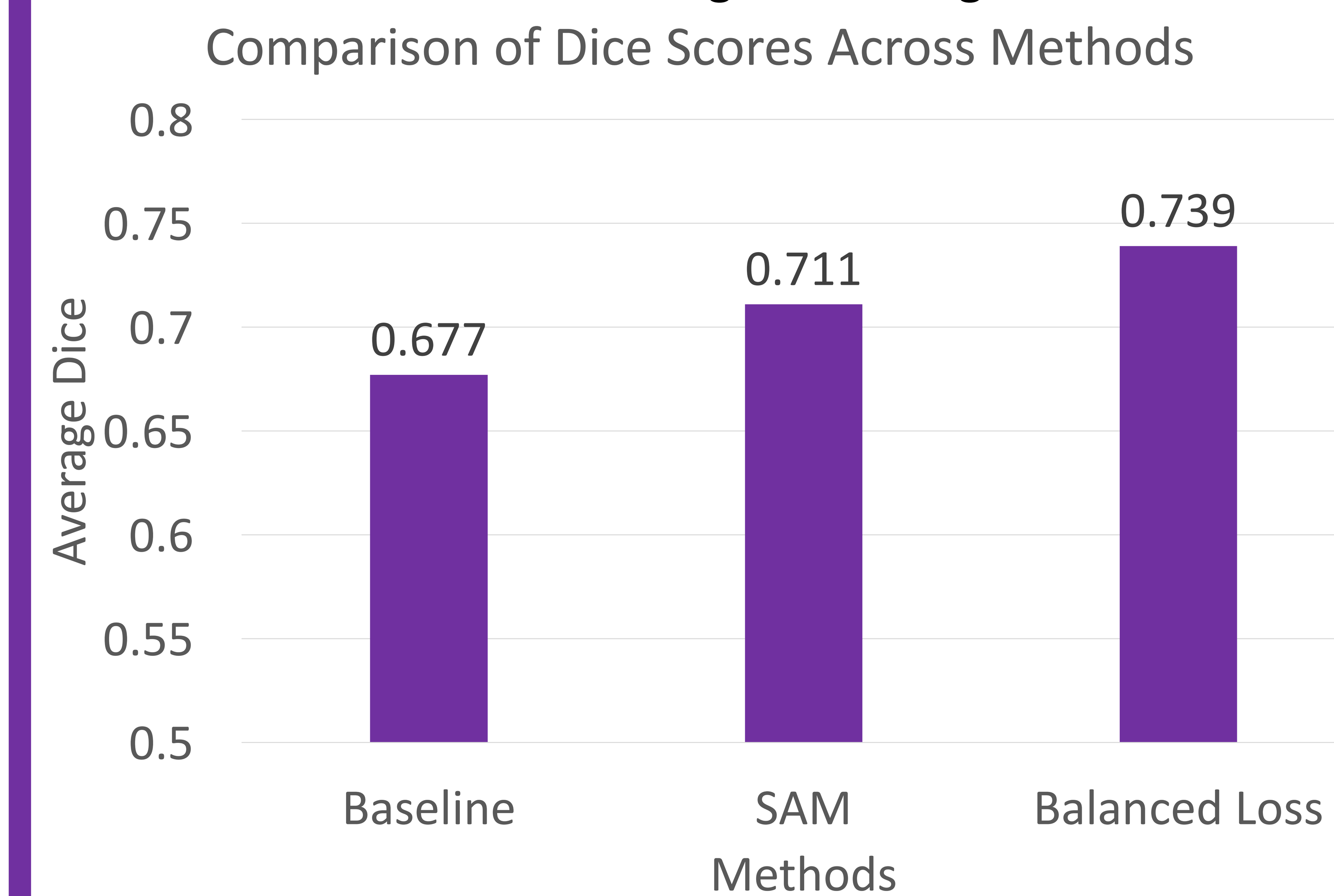

Figure 4. Comparison of Dice scores across different methods. Balanced Loss achieves the best performance, outperforming both baseline and SAM-based optimization.

Key Findings

- Both SAM and Balanced Loss improve segmentation performance
- Balanced Loss provides greater improvement than SAM
- Higher Dice scores indicate improved overlap with ground truth, balanced Loss enhances performance under severe class imbalance

Future work

- Combine SAM and Balanced Loss for joint optimization
- Explore applying both strategies simultaneously
- Evaluate performance on more datasets and tasks

Reference

- Du et al. (2024). SegVol: Universal and Interactive Volumetric Medical Image Segmentation
- Foret et al. (2021). Sharpness-Aware Minimization for Efficiently Improving Generalization
- Menon et al. (2020). Long-tail Learning via Logit Adjustment