Short Report: 3D Cardiac MRI Segmentation
Baseline vs Augmented Model Evaluation

1. Overview

This project addresses the task of 3D segmentation from cardiac MRI scans. The dataset consists of 3D MRI volumes with corresponding binary segmentation masks. Because MRI data varies widely in intensity, voxel spacing, and field of view, the pipeline performs mandatory preprocessing steps before model training. The objective is to train and compare:

- A **baseline** 3D U-Net trained without augmentation
- An **augmented** 3D U-Net trained with geometric + MRI-style intensity augmentations

Performance is quantified using Dice, HD95 distance, Sensitivity, and Precision.

2. Methods

2.1 Preprocessing

The following preprocessing steps were applied:

- **Resampling** to 1.5 mm isotropic spacing (SimpleITK)
- **Z-score normalization** with percentile clipping
- **Padding** volumes to sizes divisible by 32
- **Patch extraction:** 96×96×96 patches, 16 per volume, with 90% foreground sampling probability

2.2 Model Architecture

A residual **3D U-Net** was used, consisting of:

- Encoder with LayerNorm + LeakyReLU
- Bottleneck with dropout
- Symmetric decoder with skip connections
- Sigmoid output for binary segmentation

2.3 Augmentation Strategy

Augmentations were carefully designed to be anatomically safe:

- **Geometric (Replay-based 2D-consistent):** RandomRotate90, flips, $\pm 5^{\circ}$ rotation, $\pm 3\%$ translation
- **MRI-Realistic Intensity:** mild gamma shifts, light brightness/contrast changes, rare Rician noise

2.4 Training Setup

Both models were trained identically:

- Optimizer: AdamW (2e-4 learning rate)
- Loss: 0.5 Dice + 0.5 BCE
- Scheduler: warmup + cosine decay (30 epochs)
- Early stopping and patch-based training

3. Experiments

Two experiments were conducted:

- **Baseline model:** no augmentation
- **Augmented model:** geometric + intensity augmentations

Both were evaluated using full-volume sliding-window inference.

4. Results

4.1 Quantitative Metrics (Your Actual Values)

Table 1: Baseline vs Augmented Mean Metrics on Test Set

Model	Dice ↑	HD95 ↓	Sens. ↑	Prec. ↑
Baseline	**0.8965**	**2.6912**	**0.9536**	**0.8995**
Augmented	0.8869	2.6912	0.9503	0.8939

Observation: The baseline model slightly outperforms the augmented model across all metrics.

4.2 Training Curves & Visual Samples

The following plots were generated (available in repository):

- Training–validation loss curves
- Training–validation Dice curves
- Original vs augmented slices
- Predicted vs ground-truth overlays

5. Discussion

Despite multiple experiments, augmentation did not improve performance. Possible reason include:

- **Small sample size:** For small test sets (3 cases), metric shifts may not generalize.

Notably, HD95 remained identical on average, indicating boundary quality was stable even if overlap (Dice) dropped.

6. Conclusion

This project demonstrates a full 3D medical segmentation pipeline including preprocessing, patch-based training, residual 3D U-Net modeling, augmentation, and full-volume evaluation. The key findings are:

- The baseline model achieved the best performance.
- Augmentation provided no measurable benefit and in some metrics slightly decreased performance.