3D Lung Segmentation Challenge

Oktay Kurt
Università degli Studi di Milano-Bicocca
o.kurt@campus.unimib.it
910544

1 Introduction

Accurate 3D lung segmentation from CT images is critical for the diagnosis and treatment of pulmonary diseases. This project builds upon a 3D U-Net to tackle the AAPM Thoracic Auto-segmentation Challenge dataset. The primary objective is to reduce overfitting and enhance generalization through:

- A 5-level U-Net architecture (deeper than the baseline).
- Strong data augmentations (flips, rotations, elastic transforms).
- Dropout, L2 weight decay, and learning rate schedules.

Various learning rates, dropout rates, and weight decays were tested; early stopping was also employed. The final Jupyter Notebook that implements these improvements has been uploaded along with the trained weights for inference.

2 Methods

2.1 Model Architecture

The network is a 5-level 3D U-Net that processes volumes of size $64 \times 64 \times 64$. Each encoder level consists of:

- Two 3D convolutions (kernel size = $3 \times 3 \times 3$), each followed by Batch Normalization and ReLU.
- Dropout (0.5) and L2 weight decay (10^{-5}) to reduce overfitting.
- A MaxPooling3D step to halve spatial dimensions.

The decoder mirrors this structure using Conv3DTranspose for upsampling, concatenating skip connections from the corresponding encoder level.

2.2 Data Preprocessing & Augmentation

- **HU Windowing:** $[-1000, 300]$ clipped, then normalized to $[0, 1]$.
- **Lung Binarization:** Masks thresholded to 0 or 1.
- **Augmentation:**
  - Random axis flips.
  - Random 3D rotations up to $\pm 10°$.
  - Elastic deformations with B-Spline transforms.

This combination was iteratively tuned to increase variance in training data and reduce overfitting.

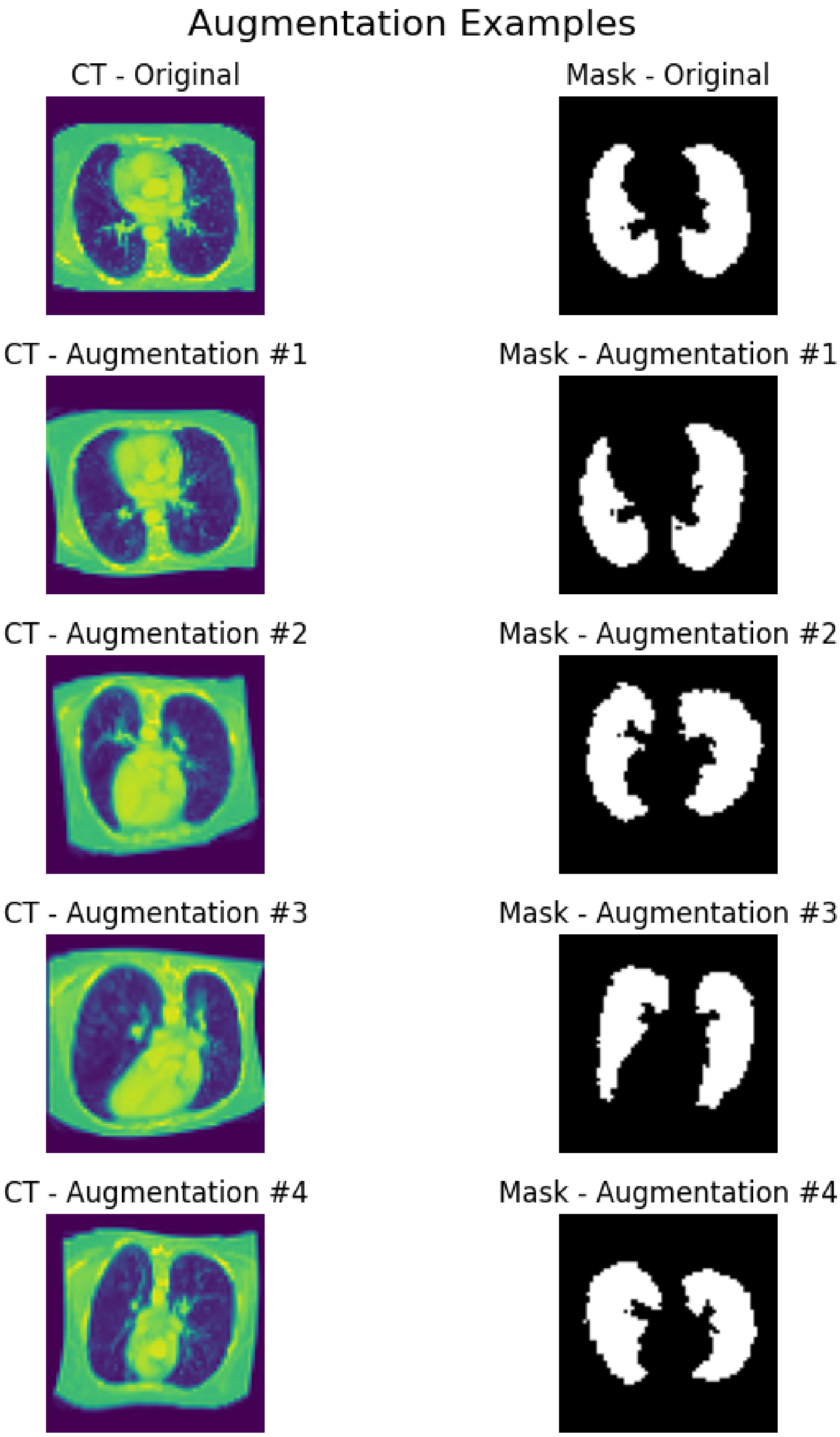

Figure 1: Demonstration of augmentations on a single CT slice. Elastic warping is particularly visible around the lung boundaries.

2.3 Training Setup

- **Loss Function:** A blend of binary cross-entropy and Dice loss.
- **Optimizer:** Adam, with initial learning rates $\{1 \times 10^{-4}, 5 \times 10^{-4}, 1 \times 10^{-3}\}$. The best results came from 5×10^{-4} with a ReduceLROnPlateau schedule (patience = 6).

- **EarlyStopping:** Monitored validation Dice, patience = 15.
- **Batch Size:** 1 volume.
- **Train/Val/Test Split:** ∼80% train, 10% validation (6 patients), 10% test (6 patients).

3 Results

3.1 Training Curves

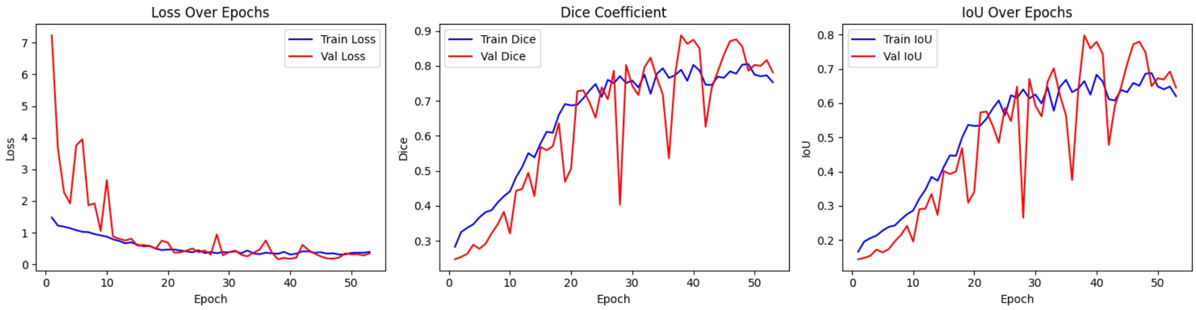

Figure 2: Training metrics: (a) Loss over epochs, (b) Dice over epochs, (c) IoU over epochs. Initial Dice values were low (0.3) but gradually improved to ∼ 0.85–0.90. Spikes in validation metrics arise due to the small validation set.

3.2 Visual Evaluation on Validation Data

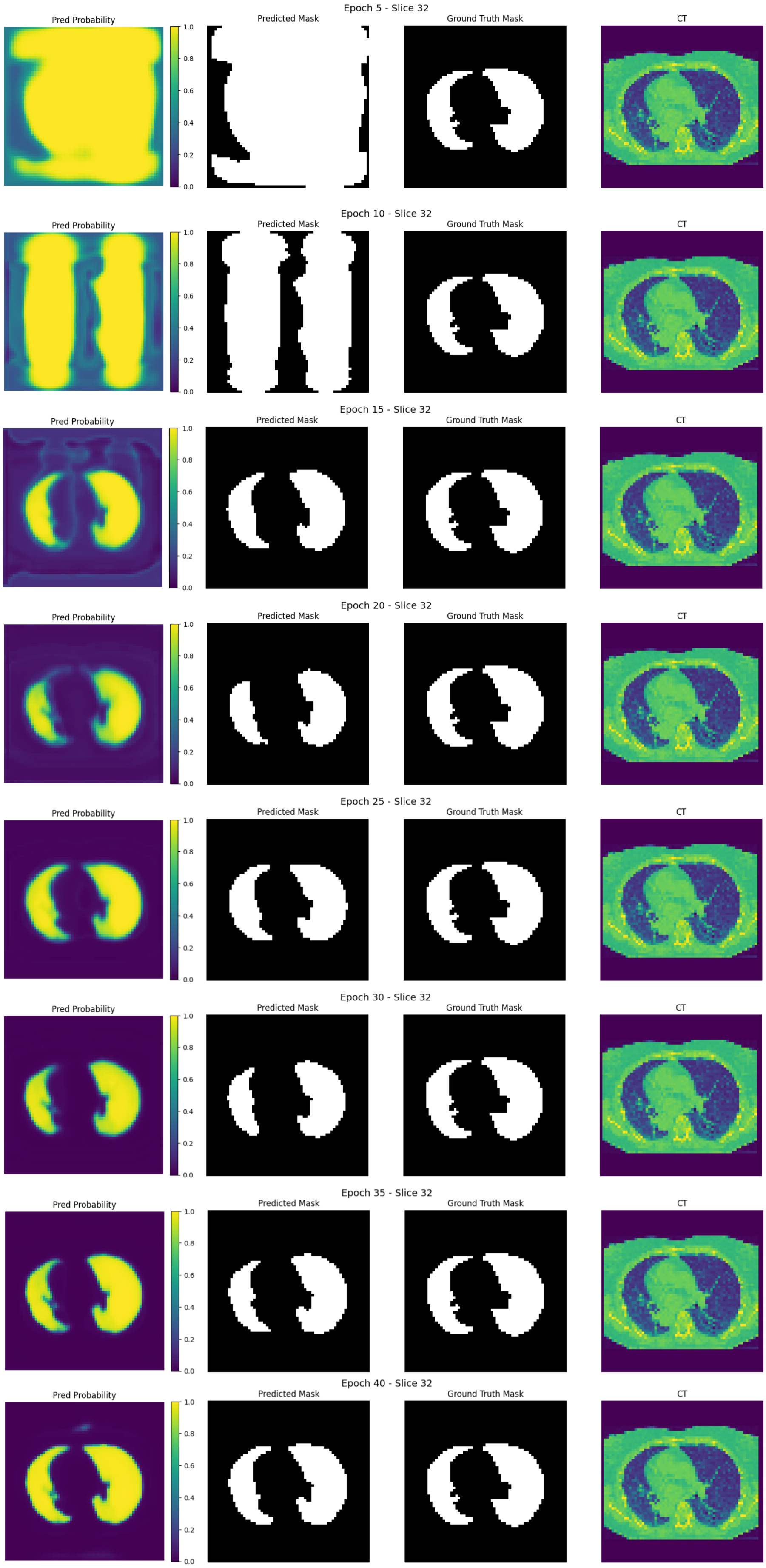

Figure 3: Validation predictions at different epochs. Early epochs show fragmented masks, while later epochs approach near-complete lung coverage.

3.3 Test Results

Final Test Metrics:

- Test Loss = 0.1586
- Test Dice = 0.8864
- Test IoU = 0.7962

These scores confirm robust lung segmentation with minimal overfitting.

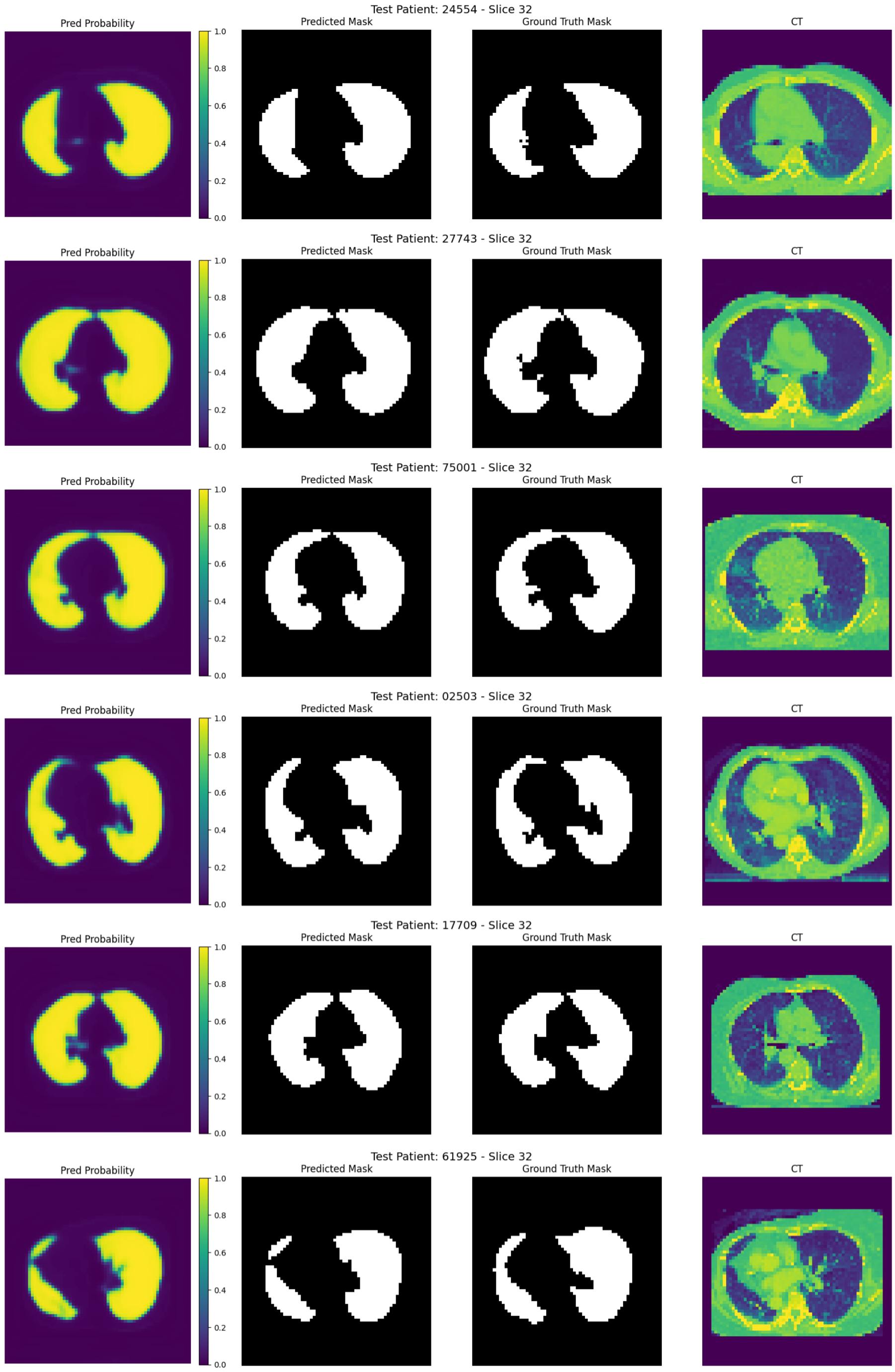

Figure 4: Predicted masks align well with ground truth lung structures across diverse patient anatomies.

4 Discussion and Conclusions

Overfitting was effectively reduced by:

- High dropout (0.5) in the U-Net convolution blocks.
- Weight decay (10^{-5}).
- Extensive augmentations (random flips, rotations, elastic).
- Adaptive learning rate via ReduceLROnPlateau and EarlyStopping.

Despite minor oscillations of the validation metric due to the limited validation set (6 patients), the final test Dice of ~ 0.886 demonstrates strong lung coverage.

Future enhancements might include:

- Attention-based 3D U-Nets or residual blocks for more nuanced features.
- Larger input resolutions ($128 \times 128 \times 128$) if memory allows, capturing finer lung boundaries.
- Cross-validation or a bigger validation split for more stable metric tracking.

Overall, the 5-level 3D U-Net with strategic regularization and augmentation achieved reliable segmentation while mitigating overfitting. The final Jupyter Notebook and weights have been uploaded, enabling reproducible inference and further experimentation.

References

1. AAPM Thoracic Auto-segmentation Challenge, 2017.
2. Original data on The Cancer Imaging Archive (TCIA): "LCTSC" dataset.