COVID-19 Lung and Infection Segmentation from Chest X-rays Using U-Net

Lab Report

Abstract—**This study applies a U-Net architecture with residual blocks to two medical image segmentation tasks on the COVID-QU-Ex dataset: lung region segmentation and COVID-19 infection segmentation from chest X-ray images. We compare two loss functions—BCE and BCE+IoU—across both tasks. For lung segmentation ($\sim$33,920 images, 3 classes), both losses achieve strong performance with IoU > 0.95, while the combined BCE+IoU loss yields lower MAE (0.0117 vs. 0.0151). For infection segmentation ($\sim$2,913 COVID-19 images), the task proves substantially harder, reaching IoU of 0.3714 with BCE+IoU loss after 10 epochs.**

I. Introduction

Accurate segmentation of lung regions and infection areas from chest X-rays is critical for COVID-19 diagnosis and severity assessment. Manual delineation is time-consuming and subject to inter-observer variability. This work explores automated segmentation using U-Net with residual blocks on the COVID-QU-Ex dataset, comparing BCE and combined BCE+IoU loss functions on two tasks: (1) lung segmentation across all disease categories, and (2) infection region segmentation for COVID-19 cases.

II. Exploratory Data Analysis

The COVID-QU-Ex dataset contains chest X-ray images organized into three categories: COVID-19, Non-COVID, and Normal, with corresponding segmentation masks.

A. Lung Segmentation Data

The lung segmentation subset includes all three classes with lung boundary masks. All images are 256×256 pixels in grayscale. The dataset is split into training, validation, and test sets totaling approximately 33,920 images across all classes.

B. Infection Segmentation Data

The infection segmentation subset provides infection region masks only for COVID-19 cases (Non-COVID and Normal cases have no infection regions to segment). This results in a smaller dataset of approximately 2,913 images, making the learning task more challenging.

C. Key Differences

Lung masks delineate consistent bilateral lung boundaries, producing relatively uniform mask shapes. Infection masks, in contrast, mark irregular and variable pathological regions whose shape, size, and location depend on disease severity. Additionally, infection regions occupy a much smaller fraction of the image, creating a severe class imbalance at the pixel level.

III. Methods

A. U-Net with Residual Blocks

We implement a 3-level U-Net with residual skip connections ($\sim$8M parameters, base channels = 64). Each residual block consists of two 3×3 convolutions with batch normalization and ReLU, plus a 1×1 shortcut when channel dimensions change. The encoder has three levels (64, 128, 256 channels) with max-pooling, a 512-channel bottleneck, and a symmetric decoder with transposed convolutions and skip connections from the encoder.

B. Loss Functions

We compare two loss functions:

- **BCE Loss**: Binary cross-entropy with logits for pixel-wise classification.
- **BCE + IoU Loss**: A combination with equal weighting ($\alpha = 0.5$):

$$\mathcal{L} = 0.5 \cdot \mathcal{L}_{\text{BCE}} + 0.5 \cdot \mathcal{L}_{\text{IoU}} \quad (1)$$

where $\mathcal{L}_{\text{IoU}} = 1 - \frac{|P \cap G| + \epsilon}{|P \cup G| + \epsilon}$ directly optimizes region overlap.

C. Data Augmentation

Training samples are augmented with random horizontal flips (50%), rotations ($\pm 15°$, 50%), and brightness jitter ($\pm 20\%$, 50%). Masks are transformed identically (nearest-neighbor interpolation for geometric transforms).

D. Training Configuration

All models use Adam optimizer (lr $= 10^{-4}$, weight decay $= 10^{-5}$), ReduceLROnPlateau scheduler (patience = 7, factor = 0.5), gradient clipping at 1.0, batch size 16, and input resolution 256×256. Models are trained for 10 epochs with the best checkpoint selected by validation IoU.

E. Evaluation Metrics

- **MAE**: Mean absolute error between predicted probabilities and ground truth masks.
- **IoU**: Intersection-over-Union at threshold 0.5, measuring overlap between predicted and ground truth regions.

IV. RESULTS

A. Lung Segmentation

Table I summarizes validation performance for lung segmentation. Both loss functions achieve high IoU (> 0.95), indicating accurate lung boundary delineation. The BCE+IoU loss achieves slightly lower MAE (0.0117 vs. 0.0151), suggesting that the IoU component produces sharper predictions with fewer boundary errors.

TABLE I
LUNG SEGMENTATION — VALIDATION RESULTS (10 EPOCHS)

Loss Function	Train Loss	Val Loss	MAE	IoU
BCE	0.0274	0.0243	0.0151	0.9570
BCE + IoU	0.0428	0.0398	0.0117	0.9567

Both models converge steadily over 10 epochs (Fig. 1a–b). The BCE model reaches best IoU at epoch 9 (0.9570), while the BCE+IoU model improves continuously to epoch 10 (0.9567). The near-identical IoU values indicate that both losses are effective for this well-defined segmentation task.
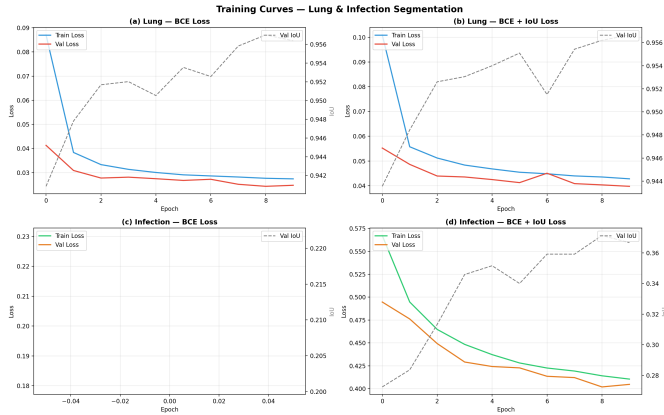

Fig. 1. Training curves for all four experiments. (a) Lung/BCE, (b) Lung/BCE+IoU, (c) Infection/BCE, (d) Infection/BCE+IoU. Solid lines show train/val loss (left axis); dashed lines show val IoU (right axis).

Sample predictions (Fig. 2) show that both loss functions produce visually accurate lung masks closely matching the ground truth.

B. Infection Segmentation

Table II shows results for infection segmentation using BCE+IoU loss. The substantially lower IoU (0.3714) reflects the difficulty of this task: infection regions are small, irregularly shaped, and highly variable across patients.

TABLE II
INFECTION SEGMENTATION — VALIDATION RESULTS (10 EPOCHS)

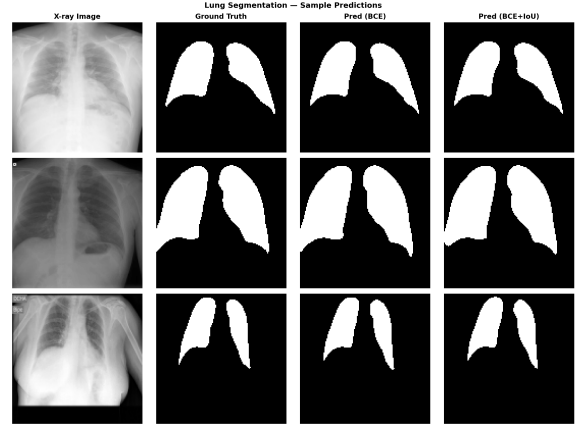| Loss Function	Train Loss	Val Loss	MAE	IoU
BCE + IoU	0.4105	0.4020	0.0438	0.3714

Fig. 2. Lung segmentation sample predictions on the test set. Columns: input X-ray, ground truth mask, BCE prediction, BCE+IoU prediction.

The training curve (Fig. 1d) shows continuous improvement without plateau, suggesting that longer training could further improve performance. Sample predictions (Fig. 3) reveal that both models capture the general infection location but struggle with precise boundary delineation.
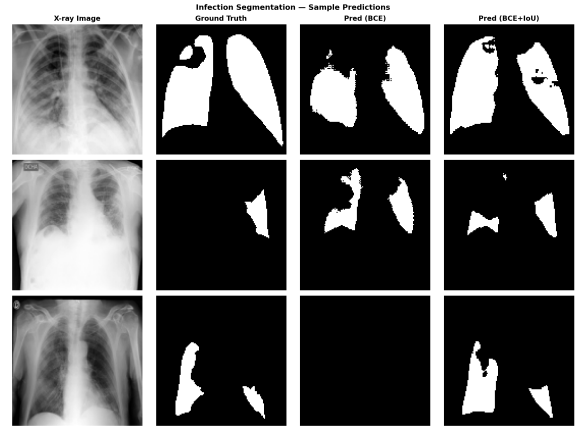

Fig. 3. Infection segmentation sample predictions on the test set. Columns: input X-ray, ground truth mask, BCE prediction, BCE+IoU prediction.

The gap between lung and infection segmentation highlights how pixel-level class imbalance and mask variability affect model learning.

C. Comparison Across Tasks

TABLE III
OVERALL COMPARISON — BEST VALIDATION IoU

Task	Loss	MAE	IoU
Lung	BCE	0.0151	**0.9570**
Lung	BCE + IoU	0.0117	0.9567
Infection	BCE + IoU	0.0438	0.3714

Lung segmentation is a well-posed problem with large, consistent target regions, enabling high accuracy even with

limited training. Infection segmentation requires modeling subtle, sparse patterns with far less training data, explaining the performance gap.

V. DISCUSSION

Loss function impact. For lung segmentation, BCE and BCE+IoU yield comparable IoU, but the combined loss produces lower MAE, indicating that the IoU term encourages more precise boundaries. For infection segmentation, the IoU component is essential to avoid trivial all-zero predictions caused by severe class imbalance.

Task difficulty. Lung boundaries are anatomically consistent and occupy a large image fraction, making them easy targets. Infection regions are diffuse, variable, and occupy a small fraction of the lung area, presenting a harder optimization landscape.

Limitations. Training was limited to 10 epochs due to computational constraints, and the infection model had not yet plateaued. Longer training, additional augmentation (e.g., elastic deformation), and architectural modifications (e.g., attention gates, deeper encoder) could improve infection segmentation.

VI. CONCLUSION

This study demonstrates that a residual U-Net achieves strong lung segmentation (IoU > 0.95) from chest X-rays with only 10 epochs of training, with both BCE and BCE+IoU losses performing comparably. Infection segmentation is substantially harder (IoU $= 0.37$) due to smaller and more variable target regions and limited training data. The combined BCE+IoU loss is recommended for both tasks as it directly optimizes for region overlap and produces sharper boundaries. Future work should explore longer training schedules, attention mechanisms, and multi-task learning to improve infection segmentation accuracy.