ECE 6545 Deep Learning Image

Project 3: X-ray Image Classification for Infection Diagnosis

Introduction

X-ray imaging is critical for diagnosing infections like COVID-19, enabling rapid treatment. This project uses the covid-chestxray-dataset (914 images) to classify four infection types: COVID-19, Bacterial, Viral, and Other. The main challenges include label imbalance (COVID-19: 64%) and mixed file formats (.nii.gz). We implemented a Simple CNN and ResNet18 to address these issues, applying data augmentation and weighted loss to improve performance.

Methods

We implemented two models:

1. **Simple CNN**: A custom architecture with two convolutional layers (16 and 32 channels), ReLU activations, max-pooling, and a fully connected layer with Dropout (0.5). The output layer has 4 nodes for the infection classes.
2. **ResNet18**: A standard ResNet18 architecture from torchvision, modified for single-channel input (grayscale X-rays) by adjusting the first convolutional layer (1 input channel, 64 output channels). The final fully connected layer was modified to output 4 classes. No pre-trained weights were used.

Training Strategies:

- **Optimizer**: Adam with a learning rate of 0.001.
- **Loss Function**: Weighted cross-entropy loss with weights [1, 3, 5, 2] to address label imbalance.
- **Data Augmentation**: Random horizontal flips and 10-degree rotations to improve generalization.

- **Preprocessing**: Filtered out .nii.gz files, resized images to 224×224, and normalized with mean 0.5 and standard deviation 0.5.

Training was performed for 10 epochs on a TITAN X GPU using PyTorch 2.0.1.

Experiments

Dataset and Metrics

The dataset contains 914 X-ray images from the covid-chestxray-dataset (https://github.com/ieee8023/covid-chestxray-dataset), with the following distribution: COVID-19 (584), Other (231), Bacterial (69), and Viral (30). We split the data into 80% training (731 images), 10% validation (91 images), and 10% testing (92 images). Evaluation metrics include weighted F1 score and confusion matrices.

Baseline

We established a majority baseline that always predicts the most common class (COVID-19, 64%). On the validation set (91 images: 58 COVID-19, 23 Other, 7 Bacterial, 3 Viral), this baseline achieves a weighted F1 score of 0.406, as it correctly predicts all COVID-19 samples but fails on other classes.

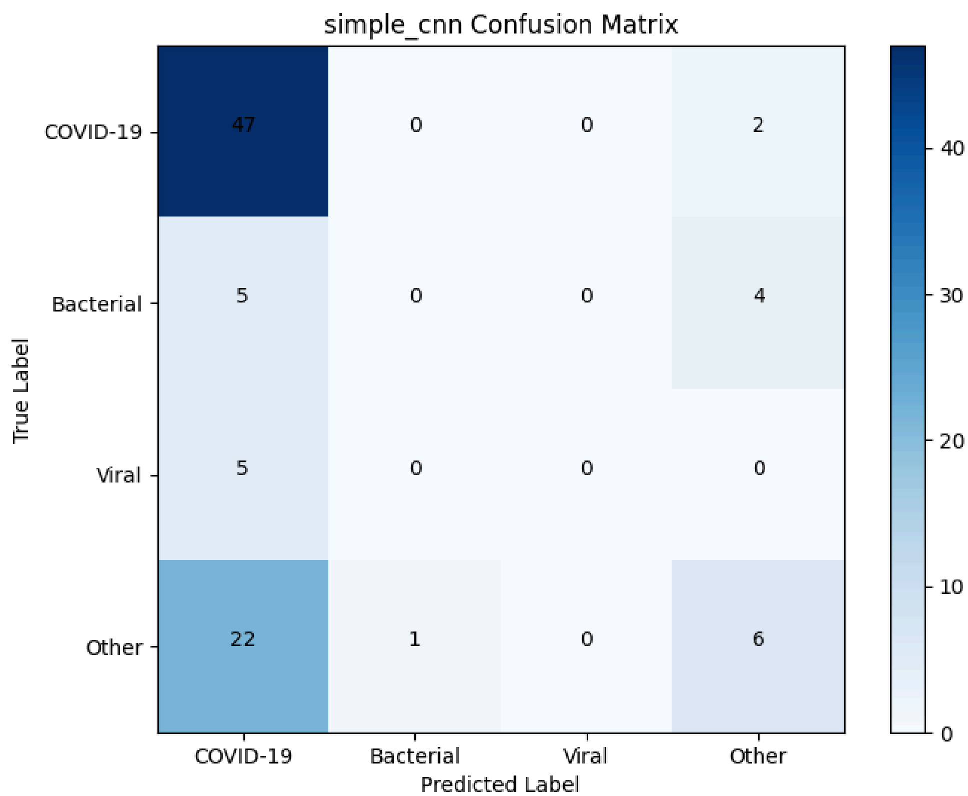
Results

- **Simple CNN**: Achieved a validation F1 score of 0.5917 at epoch 10, significantly outperforming the baseline (0.406). The confusion matrix (Figure 1) shows a high accuracy for COVID-19 (95.9%), but poor performance for Bacterial (0%) and Viral (0%), with most samples misclassified as COVID-19 or Other. The loss curve (Figure 2) indicates stable training loss (1.5 to 1.1), but validation loss fluctuates (1.2 to 1.3), suggesting overfitting.
- **ResNet18**: Achieved a validation F1 score of 0.5764 at epoch 7, also outperforming the baseline (0.406). The confusion matrix (Figure 3) shows a lower COVID-19 accuracy (73.5%), but better performance for Viral (60% accuracy). Bacterial accuracy remains low (11.1%). The loss curve (Figure 4) shows stable training loss (1.4 to 1.1), but validation loss fluctuates more (1.1 to 1.7), indicating stronger overfitting than Simple CNN.

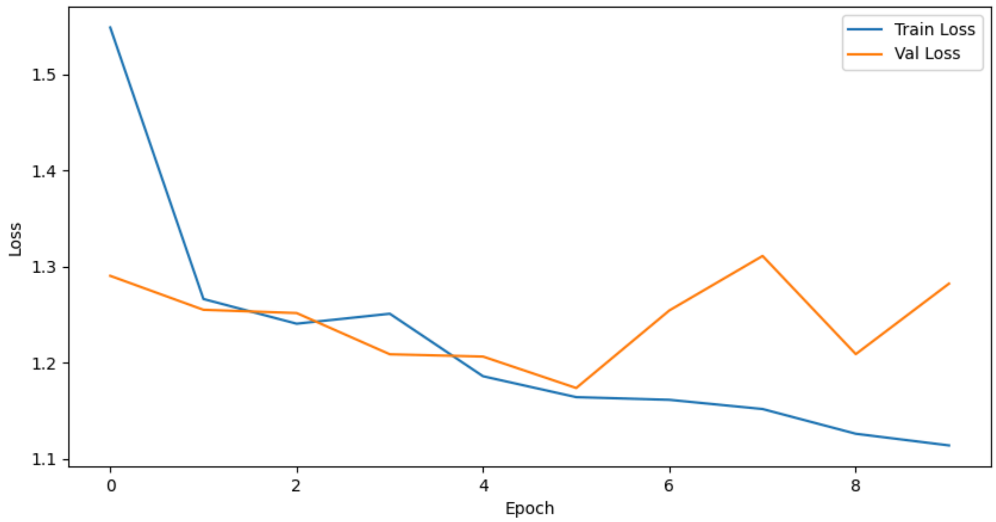
Analysis

Simple CNN outperforms ResNet18 in F1 score (0.5917 vs. 0.5764) and COVID-19 classification, likely due to its simpler architecture being less prone to overfitting on this small dataset. However, ResNet18 performs better on smaller classes like Viral (60% vs. 0%), suggesting it captures more diverse features. Both models significantly outperform the majority baseline (F1: 0.406), indicating effective learning. Label imbalance severely impacts both models, with small classes (Bacterial, Viral) often misclassified as COVID-19. The weighted loss helps but is insufficient. Overfitting is evident in both models, especially ResNet18, due to the lack of pre-trained weights and limited data.

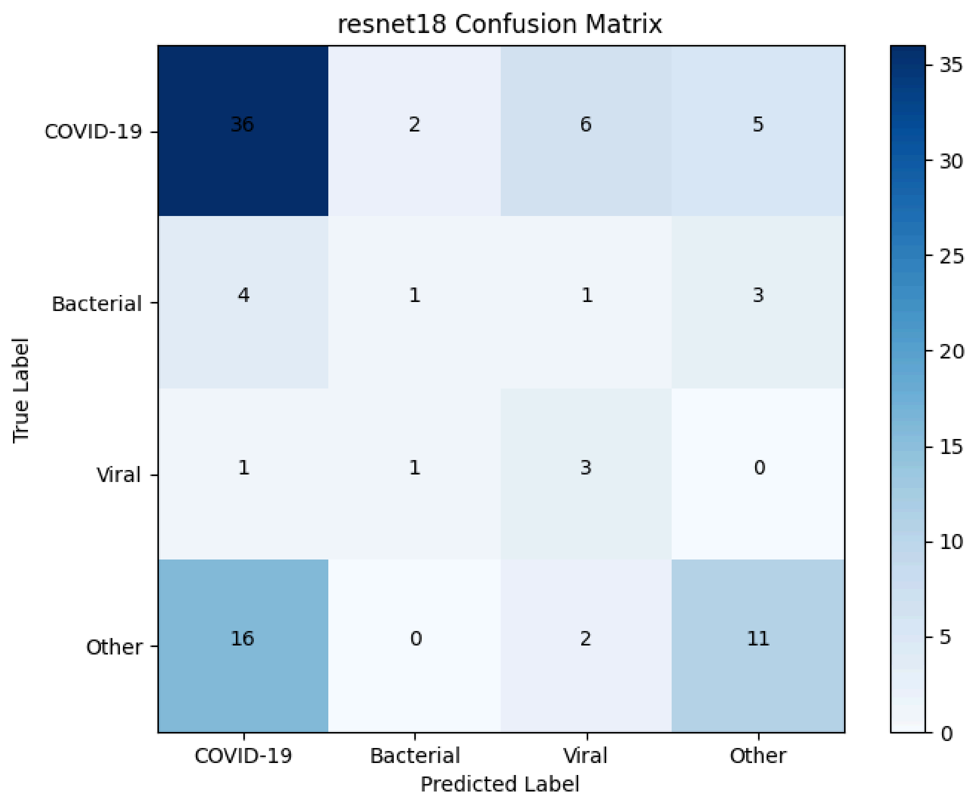
- Figure 1: Simple CNN Confusion Matrix



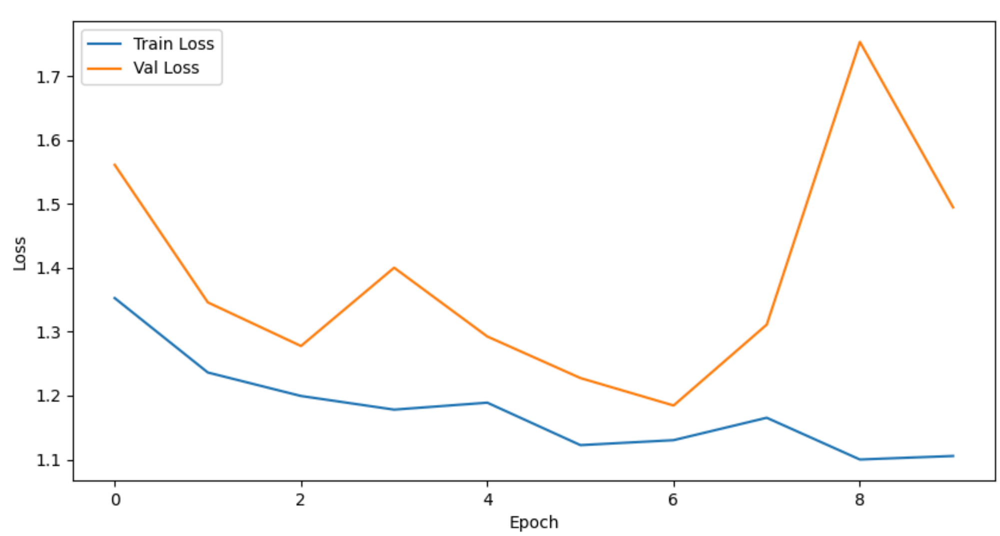
- Figure 2: Simple CNN Loss Curve



- Figure 3: ResNet18 Confusion Matrix



- Figure 4: ResNet18 Loss Curve



Conclusion

Simple CNN slightly outperforms ResNet18 (F1: 0.5917 vs. 0.5764) and the majority baseline (F1: 0.406), but ResNet18 better handles smaller classes like Viral (60% accuracy). Label imbalance significantly affects performance on Bacterial and Viral classes. Data augmentation and weighted loss were partially effective. Future improvements could include using pre-trained ResNet18 weights, increasing epochs, adjusting the weighted loss, or exploring Transformer models.

Code Citation

We used PyTorch 2.0.1 and torchvision's ResNet18. The dataset is from https://github.com/ieee8023/covid-chestxray-dataset.