TASK III REPORT

Overview

This project focuses on the conversion of a 2D pretrained DenseNet121 model into a 3D CNN and its application in analyzing medical CT scan volumes. The core objective is to extract meaningful features from different bone regions (Tibia, Femur, and Background) using deep convolutional layers and assess the similarity between these regions using cosine similarity.

Objectives

- Convert a pretrained 2D DenseNet121 into a 3D model.
- Extract features from selected 3D convolution layers (last, 3rd-last, 5th-last).
- Apply global average pooling to generate fixed-size feature vectors.
- Compare feature similarity across Tibia, Femur, and Background regions.
- Save results and analysis for downstream tasks and visualization.

Methodology

1. 2D to 3D Model Conversion

- A custom PyTorch class DenseNet3D was created to inflate a pretrained DenseNet121 2d model into 3D.
- 2D layers (Conv2d, BatchNorm2d, Pooling) were converted to 3D counterparts.
- The model was adapted to handle grayscale CT volumes (1 channel).

2. Feature Hook Registration

- The last, third-last, and fifth-last convolutional layers were captured using forward hooks.
- Each hook stored the intermediate output (feature map) from the respective layer.

3. Feature Vector Extraction

- Global Average Pooling (GAP) was applied to each 3D feature map to produce an N-dimensional vector.
- Features were extracted independently for Tibia, Femur, and Background masks.

4. Similarity Analysis

- Cosine similarity was computed between feature vectors for the following region pairs:
  - Tibia vs. Femur
  - Tibia vs. Background
  - Femur vs. Background
- This was repeated for each of the three selected layers.

5. Result Organization

- The final DataFrame included each region pair and cosine similarity across three depths.
- Results were exported to a CSV file.

Results Summary

Region Pair	Last Layer	3rd-Last	5th-Last
Tibia-Femur	0.9807	0.9252	0.9464
Tibia-Background	0.8067	0.7433	0.6759
Femur-Background	0.8304	0.7846	0.6318

- Tibia and Femur show consistently high similarity, indicating their structural proximity.
- Background shows decreasing similarity from last to fifth-last layer, indicating low-level feature separation.

Observations

- Deeper layers (last) abstract over finer details, resulting in high similarity even across distinct regions.
- Shallower layers (5th-last) retain more structural uniqueness, helpful in distinguishing bone from background.
- The model was not fine-tuned on CT data, so deeper features are likely biased toward pretrained ImageNet representations.

Conclusion

This pipeline effectively demonstrates how 3D CNNs can extract features from medical volumes and how different layers impact similarity across anatomical regions. These features can be further used for classification, segmentation, or clinical analysis tasks.

Refine can be done

- Experiment with other models.
- Replace GAP with attention-based pooling for better regional specificity.
- Expand to multi-sample analysis and statistical evaluation.