Knee Feature Extraction and Similarity Comparison

Introduction

This technical report details the development of a pipeline for extracting and comparing deep learning features from a segmented 3D knee CT volume. The methodology combines semi-automated mask preparation followed by a novel 2D-to-3D model conversion approach, enabling feature analysis and similarity assessment across different regions.

Methodology

Mask Generation and Segmentation

The 3D knee CT data was carefully divided into three parts, clearly separating the tibia (green), femur (red), and background (grey). Semi-automated mask preparation techniques that integrated algorithmic approaches with manual verification were used because variable mask shapes and sizes were observed across different CT slices. The watershed algorithm is used to effectively detach connected small blob artefacts from the main bone structures. Batch processing with tailored thresholds and minimum size parameters was implemented in place of a unified strategy to handle inconsistent object sizes, preserve bone structures, and remove artefacts. A systematic labelling strategy was developed where the watershed algorithm identifies the tibia and femur, combined with background volume to obtain the colourful output.

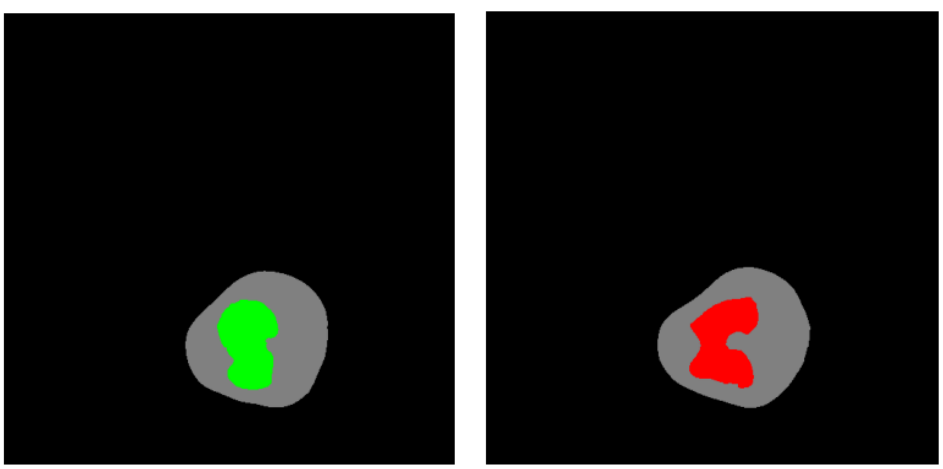

Figure 1: Segmentation results

Model Conversion

A pretrained DenseNet121 model from the PyTorch Vision library was successfully converted for 3D volumetric image processing by systematically transferring input/output channels from 2D convolution layers directly to 3D equivalents while converting kernel size, stride, and padding parameters to 3D tuples (H, W, D). The 2D weights were replicated along the depth dimension and normalized by depth value (D) to maintain consistent scaling across volumetric data. The batch normalization weights were transferred directly without modification to ensure proper feature normalization in 3D space. The pooling layers were converted by creating 3D kernel sizes, padding, and stride configurations since pooling layers contain no learnable weights. The converted model processed 3D volumetric data correctly.

Feature Extraction

Feature extraction was implemented using hook mechanisms to capture representations from different layers. The last, third-last, and fifth-last convolutional layers were targeted to capture feature maps. The forward propagation hooks were registered to the model, which automatically captured layer outputs during model inference and stored results in a structured dictionary format for systematic access. Finally, global average pooling was applied to each extracted feature map, producing fixed N-dimensional vectors. The extracted features from the ultimate layers enable downstream similarity analysis between the three segmented regions of the tibia, femur, and background.

Feature Comparison

A similarity analysis was conducted across regions using cosine similarity metrics. This process systematically computed similarity scores for three critical comparisons: tibia and femur (bone-to-bone comparison), tibia and background (bone-to-tissue contrast), and femur and background (bone-to-tissue contrast). The similarity computations used characteristics from three different layers on three different region pairs. The cosine similarity metric was employed to assess correlations between feature vectors, yielding robust similarity scores for each region pair. All similarity scores were systematically recorded in CSV format for documentation and analysis purposes.

	Image Pair	last	third-last	fifth-last
0	tibia <-> femur	0.994712	0.990259	0.881864
1	tibia <-> background	0.777343	0.679824	0.613073
2	femur <-> background	0.778106	0.658943	0.542630

Figure 2: Similarity Matrix

Future Directions

Enhancing the current segmentation approach with advanced image processing algorithms or deep learning-based methods could significantly improve consistency across other datasets. The converted 3D DenseNet121 model can be applied for downstream tasks such as classification, segmentation, and object detection, providing a strong foundation for multi-task learning in volumetric medical data. Additionally, the segmentation masks produced through this pipeline are highly exact and can serve as high-quality training data for supervised segmentation models. A deeper analysis of cosine similarity results and careful validation of low-level features may guide dataset selection and preparation strategies, ultimately improving model performance and generalizability.

Conclusion

The developed pipeline presents a straightforward framework for 3D feature analysis, effectively combining classical image processing techniques with deep learning methodologies. The 3D-converted DenseNet121 model demonstrates strong potential in capturing meaningful features from a CT scan. The use of feature extraction and similarity analysis offers valuable insights at different levels of the model. Although not fully automated, the system effectively bridges traditional preprocessing with deep learning feature extraction, making it a solid foundation for further research and clinical applications.