Multiclass Image Segmentation with Fully Convolutional Neural Network Architecture

Key components include the model architecture, training strategies, evaluation metrics, and insights inspired by the Fully Convolutional Transformer (FCT) [3]

Data Preprocessing

Data [1] was normalized and converted from `.nii.gz` to `.tif` [2] format using bicubic interpolation for images and nearest-neighbor interpolation for labels to preserve quality and discrete class boundaries. All images were resized to 256x256 dimensions to address dataset size mismatch, ensuring uniformity across the dataset.

Model Architecture

The model's core design is based on a Fully Convolutional Network (FCN) with five convolutional blocks, each containing three convolutional layers followed by max-pooling. Transposed convolution layers were used for upsampling to restore original dimensions. Though inspired from FCT, the architecture did not incorporate convolutional attention and wide-focus modules to extract hierarchical context due to complexity. Regularization techniques, including Batch Normalization and dropout, were employed to enhance generalization and prevent overfitting.

Loss Function

Dice Loss, tailored for segmentation tasks, addressed class imbalances by optimizing pixel-wise overlap between predictions and ground truth. This loss function emphasizes maximizing segmentation accuracy and is effective for datasets with sparse foreground classes..

Training Strategy

The training process utilized the Adam optimizer with a batch size of 16 over 100 epochs. GradientTape was employed for efficient operation tracking and weight updates. This strategy ensured iterative improvements in segmentation performance while maintaining computational efficiency.

Performance Metrics

- **Dice Coefficient (Weighted)**: **0.7485**
- Recall (Weighted): 0.9821
- Precision (Weighted): 0.9830
- **Average IoU Score (Jaccard Index)**: **0.6292**

Interpretation

The Recall, and Precision scores are notably high due to the dominance of background pixels in the dataset, which skew these metrics upward. However, **the Dice Coefficient and Average IoU Score is more meaningful** for assessing the model's segmentation quality as they directly evaluate the overlap between predicted and ground truth segmentation maps [6]. The moderate IoU and Dice coefficient scores suggests areas for refinement, particularly in capturing fine-grained details in complex cardiac regions.

Prediction overlay of 5 random test images were also visualised

References

1. Cardiac Atlas Project. Data source: [https://www.cardiacatlas.org/]
2. Data transformation techniques: J. Creinhold's Gist. [https://gist.github.com/jcreinhold/01daf54a6002de7bd8d58bad78b4022b
3. Fully Convolutional Transformer: Tragakis et al., "The Fully Convolutional Transformer for Medical Image Segmentation". [https://arxiv.org/pdf/2206.00566] https://github.com/Thanos-DB/FullyConvolutionalTransformer/tree/main
4. Dice Loss in Medical Image Segmentation. https://cvinvolution.medium.com/dice-loss-in-medical-image-segmentation-d0e476eb486
5. GPT - 4o
    a. For tweaking codes in Model Architecture and Training strategy
    b. For generating code of prediction overlay
6. Understanding Evaluation Metrics in Medical Image Segmentation https://medium.com/@nghihuynh_37300/understanding-evaluation-metrics-in-medical-image-segmentation-d289a373a3f