

Practical Work 4: Advanced Vision, Segmentation, and 3D Data

Louis Fippo Fitime

Claude Tinku

Kerolle Sonfack

Department of Computer Engineering, ENSPY
University of Yaoundé I

October 8, 2025

Abstract

This fourth Practical Work focuses on advanced computer vision tasks that require high precision and architectural complexity. Students implement a U-Net architecture for semantic segmentation of medical images and explore the theoretical and practical challenges of processing 3D volumetric data using Conv3D layers, while applying Deep Learning Engineering best practices.

1 Part 1: Segmentation and MLOps Best Practices

1.1 Semantic Segmentation and the U-Net Architecture

Semantic segmentation consists of assigning a class label to every pixel in an image. This task is particularly important in medical imaging, where accurate localization of structures such as organs or tumors is critical.

Output Type

Unlike image classification, which outputs a single class label, a semantic segmentation model produces a tensor of spatial dimensions similar to the input image. For a 2D image of size $H \times W$, the output is typically a tensor of size $H \times W \times C$, where C represents the number of classes (or a single channel with sigmoid activation for binary segmentation).

U-Net Structure and Skip Connections

The U-Net architecture is composed of an encoder (contracting path) and a decoder (expansive path). The decoder path progressively restores spatial resolution using upsampling operations while refining predictions.

Skip connections in U-Net concatenate feature maps from the encoder to the decoder at corresponding resolutions. Unlike ResNet skip connections, which perform element-wise addition, U-Net skip connections preserve fine-grained spatial details that are crucial for precise localization.

Loss Functions for Medical Segmentation

In medical datasets, the foreground region (e.g., a tumor) often occupies a very small portion of the image compared to the background. In such cases, standard categorical cross-entropy becomes inadequate as it is dominated by background pixels.

Alternative loss functions such as the **Dice loss** or **IoU loss** are preferred because they directly optimize region overlap and are more robust to class imbalance.

1.2 Engineering Practices: Experiment Tracking

Reliable experiment comparison is essential in advanced deep learning workflows.

MLflow Naming Convention

A strict naming convention for MLflow runs should include the model architecture, optimizer, and loss function, for example:

UNet_Adam_DiceLoss

This enables efficient tracking and comparison across multiple experiments.

Custom Metrics Logging

Metrics such as the Dice Coefficient or IoU are not always available as default Keras metrics. They must be implemented as custom functions and explicitly logged to MLflow to enable consistent evaluation and reproducibility.

2 Part 2: Semantic Segmentation on Medical Data

2.1 Exercise 1: Implementing the U-Net Architecture

The U-Net model consists of convolutional blocks combined with downsampling and upsampling operations. Skip connections allow the decoder to recover spatial information lost during downsampling, leading to more accurate segmentation masks.

Training Strategy

The model is compiled using a segmentation-specific loss function and trained while monitoring metrics such as Dice or IoU. These metrics provide a more meaningful evaluation than accuracy for pixel-wise prediction tasks.

2.2 Exercise 2: Segmentation-Specific Metrics

Dice Coefficient

The Dice Similarity Coefficient measures the overlap between the predicted mask and the ground truth mask:

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|}$$

It is particularly effective for small target regions.

IoU Metric

The Intersection over Union (IoU) metric evaluates the ratio between the intersection and the union of the predicted and ground truth masks. IoU is generally more sensitive to small segmentation errors than Dice, especially when the target region is small.

3 Part 3: 3D Convolutions and Volumetric Data

3.1 Conv3D for Volumetric Data

Medical imaging modalities such as CT and MRI produce volumetric data represented as $D \times H \times W \times C$ tensors. Conv3D kernels operate across depth, height, and width, enabling the network to capture spatial correlations between adjacent slices, which is impossible with Conv2D layers.

Engineering Trade-offs

Conv3D layers are computationally expensive and memory-intensive. Engineers must carefully balance kernel size, number of filters, and input depth to avoid excessive memory usage while maintaining sufficient representational power.

3.2 Exercise 3: Conv3D Block and MLOps Discipline

A simple Conv3D block demonstrates how volumetric data is processed. MLflow is used to log model configurations, hyperparameters, and performance metrics, ensuring traceability and reproducibility of experiments.

4 Conclusion

This practical work highlighted advanced computer vision techniques for semantic segmentation and volumetric data analysis. By combining U-Net architectures, custom metrics, Conv3D layers, and MLOps best practices, students gained a comprehensive understanding of real-world challenges in medical image analysis.