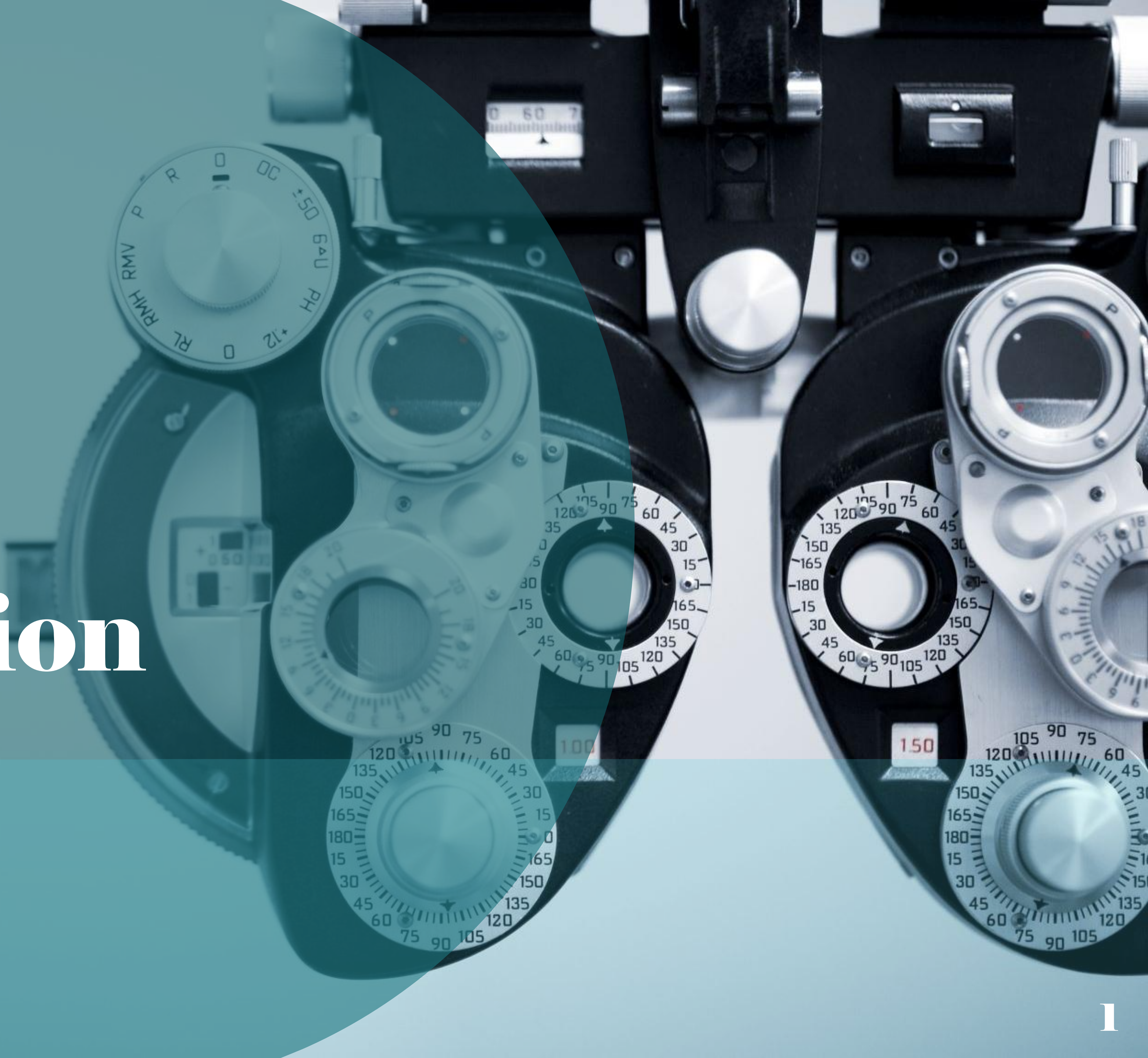
Endoscope Semantic Segmentation



Project Scope and Overview

Scope: advancing semantic segmentation in medical imaging, particularly for computer-assisted surgery.

The **main objective** is to develop neural network models that can accurately segment surgical images into distinct classes, such as:

- various tissues
- surgical instruments
- blood vessels
- other critical anatomical structures.

By improving segmentation accuracy, the project aims to enhance real-time surgical navigation and safety, providing essential support for clinical decision-making during operations

Dataset Overview

The CholecSeg8K dataset is organized in a hierarchical structure that simplifies access and usage. Here's how the dataset is structured:

I. **Top-Level Directories**:
   Each folder is named *video01*, *video02*, etc., and represents an entire surgical video clip.

II. **Segment Directories**:
   Within each video folder, the video is split into multiple segments.
   Each segment is named with the video ID and the starting frame number (e.g., *video01_00080* starts at frame 80).

III. **Frame and Image Files**:
   Each segment contains **80 consecutive frames**, and for each frame, there are **4 image files**:

- The raw image frame
- The annotation tool mask (hand-drawn by experts)
- The color mask (for visualization, with distinct class colors)
- The watershed mask (used for training, with class IDs encoded as grayscale values)
  → This totals **320 images per segment.**

Annotations: Every frame is annotated at the pixel level for **13 distinct classes**, including tissues, instruments, and blood vessels. Both the color and watershed masks include these annotations for visual and computational purposes.

Mask Overview

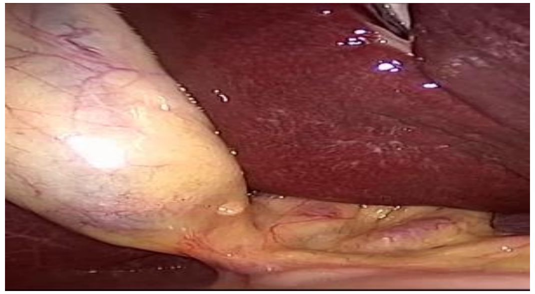
Each image frame in the dataset is accompanied by three types of masks, each serving a distinct purpose in the segmentation pipeline:

1. Original Image Frame
   The raw endoscopic image captured during surgery.Serves as the input for the segmentation model.(Image: frame_100_endo.png)

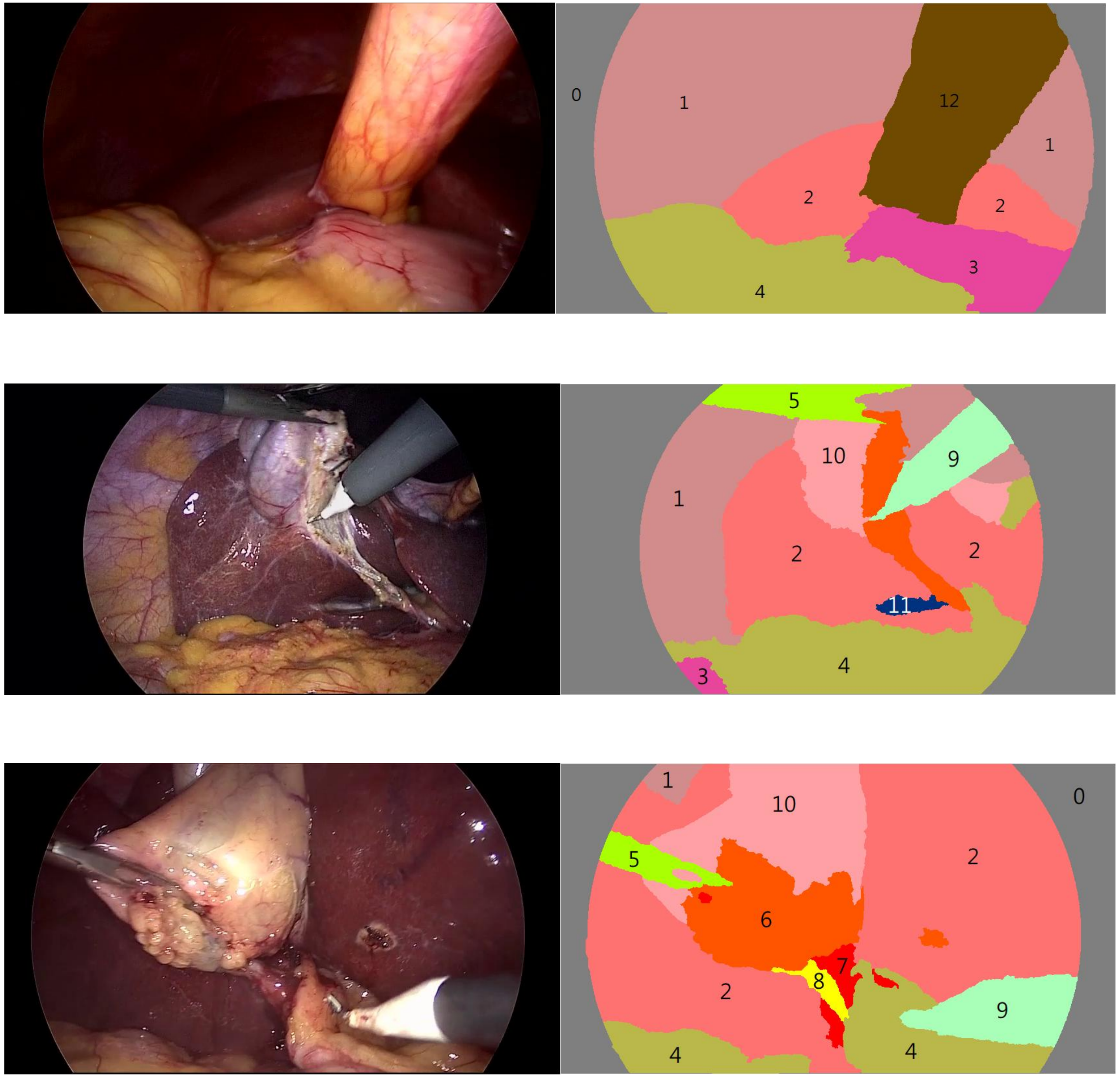
2. Annotation Tool Mask
   Hand-drawn mask created by medical experts.Provides detailed pixel-level annotations.Serves as the foundation for generating both the color and watershed masks.(Image: frame_100_endo_mask.png)

3. Color Mask
   Derived from the annotation tool mask.Assigns a unique RGB color to each class (e.g., tissue, instrument, blood).Designed for easy visual inspection and interpretation.(Image: frame_100_endo_color_mask.png)

4. Watershed Mask
   Also generated from the annotation tool mask.Encodes each class using a unique grayscale value (R=G=B).Suitable for training and automated processing as it maps directly to class IDs.(Image: frame_100_endo_watershed_mask.png)



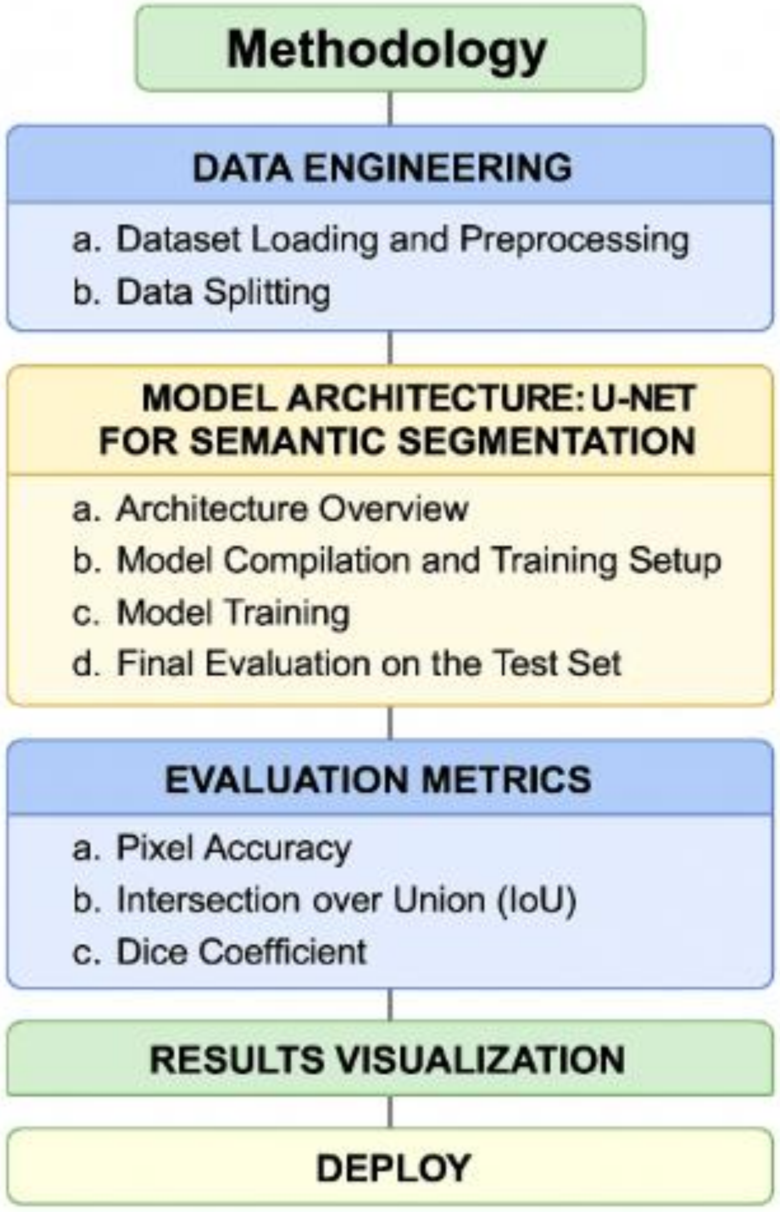
Dataset Examples of Labeling



Class Information Table

Class Number	Class Name	RGB Hexcode
Class 0	Black Background	#505050
Class 1	Abdominal Wall	#111111
Class 2	Liver	#212121
Class 3	Gastrointestinal Tract	#131313
Class 4	Fat	#121212
Class 5	Grasper	#313131
Class 6	Connective Tissue	#232323
Class 7	Blood	#242424
Class 8	Cystic Duct	#252525
Class 9	L-hook Electrocautery	#323232
Class 10	Gallbladder	#222222
Class 11	Hepatic Vein	#333333
Class 12	Liver Ligament	#050505

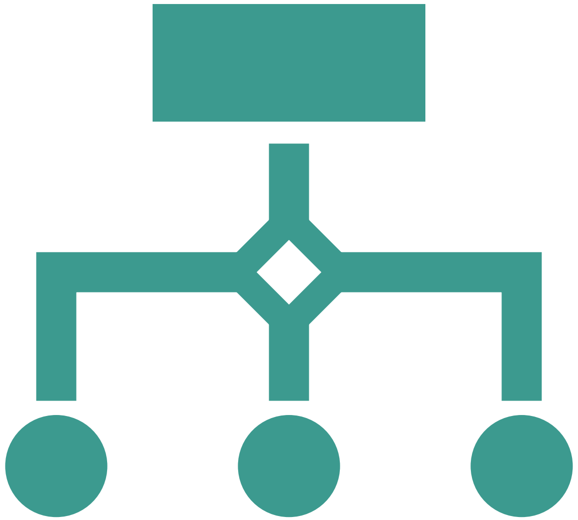
Solution Methodology



Data splitting

To evaluate our model's performance effectively, we divide the dataset into three parts:

- Training set (60%)
  Used to train the neural network and update weights during learning.
- Validation set (20%)
  Used during training to monitor model performance, tune hyperparameters, and apply early stopping.
- Test set (20%)
  Set aside until the very end. Used to evaluate the model's true generalization performance on completely unseen data.



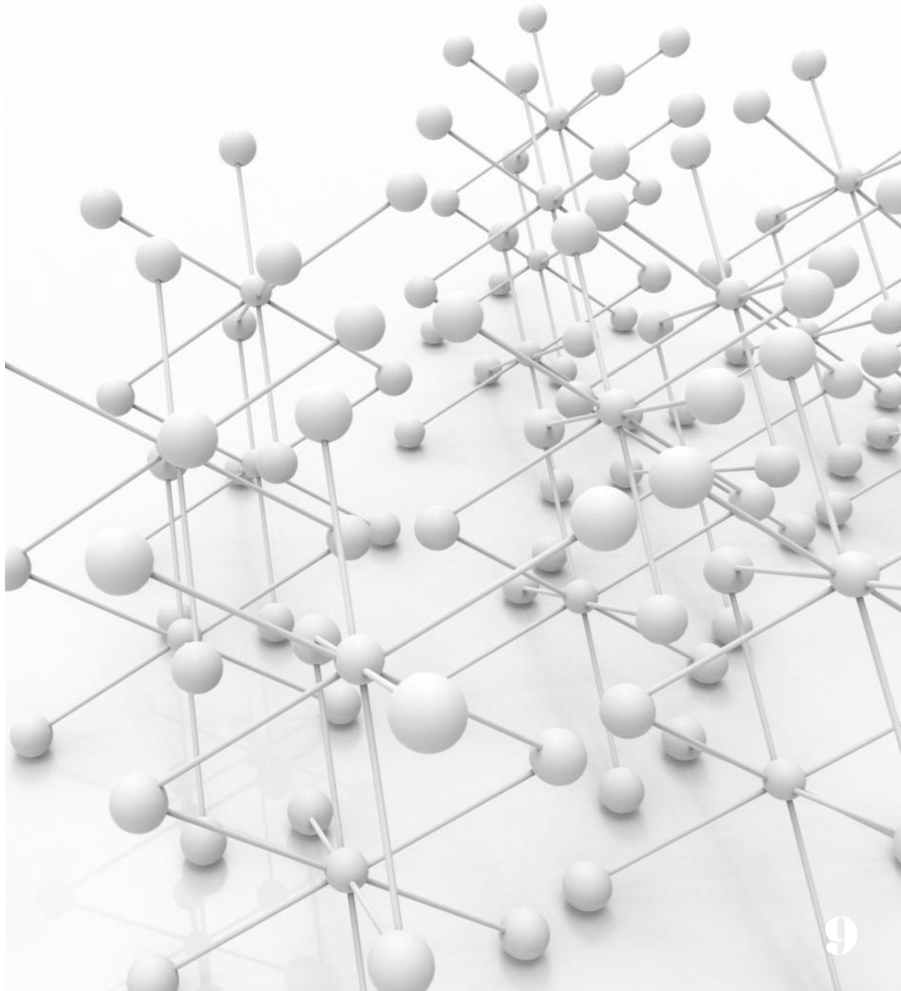
Model Architecture: U-Net for Semantic Segmentation

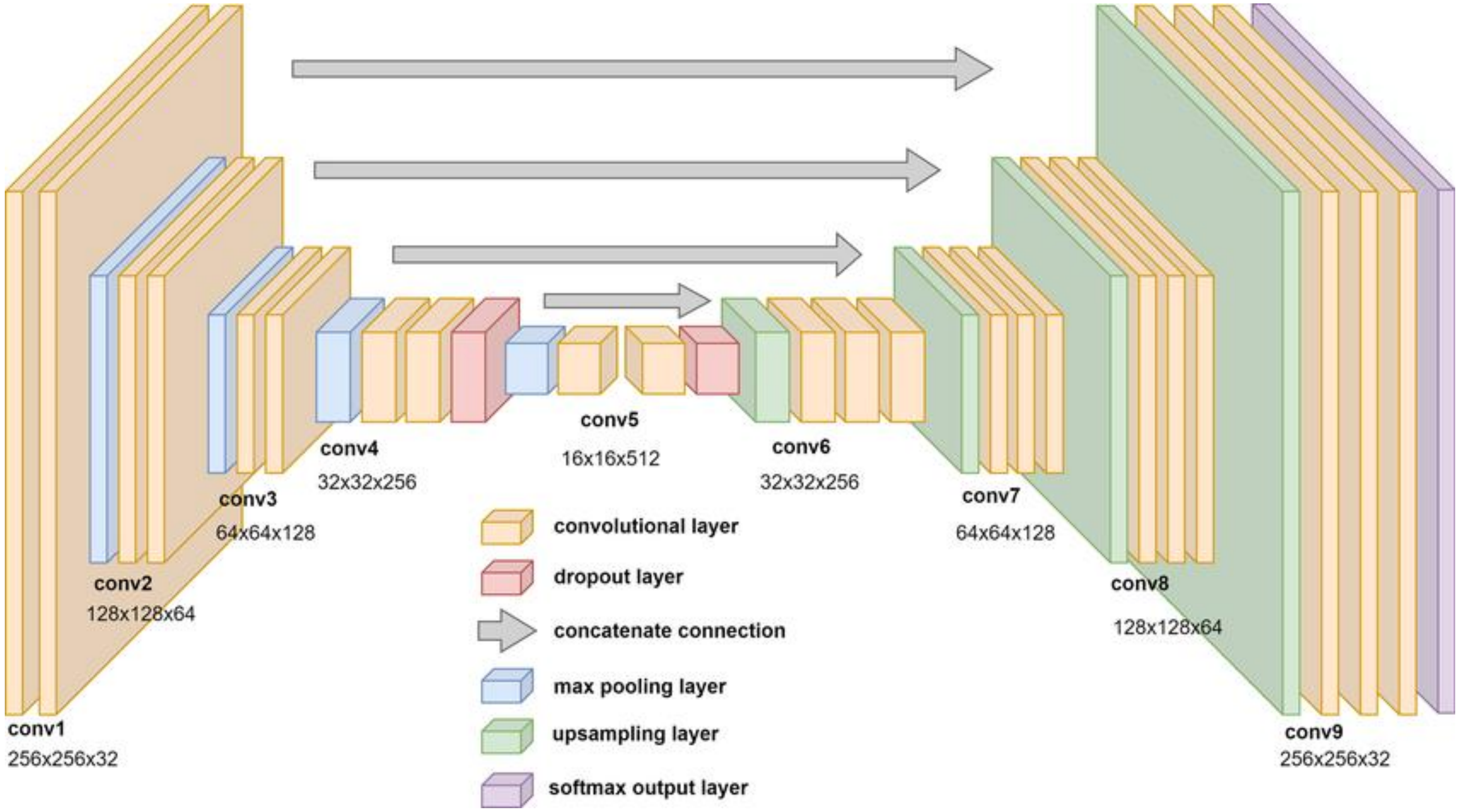
- For this project, we use a **U-Net architecture** — a popular encoder–decoder convolutional neural network designed for semantic segmentation tasks in biomedical imaging.

- U-Net is built to capture both global context and fine-grained local details, thanks to its **skip connections** that link the encoder and decoder paths.

- **1. Encoder (Contracting Path):**
- **2. Decoder (Expanding Path):**
- **3. Output Layer:**

- This architecture enables the model to segment objects at different scales and accurately preserve spatial information.





U-Net Architecture

Model Compilation and Training Setup

Loss Function:

-> SparseCategoricalCrossentropy is used since the target masks contain integer class labels (not one-hot encoded).

Optimizer:

-> Adam — a robust and widely used optimizer for deep learning, with a default learning rate.

Metrics:

-> We track pixel-wise accuracy during training.

-> More detailed metrics like IoU and Dice coefficient will be computed separately after training.

Early Stopping:

-> To prevent overfitting, we use early stopping with patience = 5.

-> This means training will stop if the validation loss does not improve for 5 consecutive epochs. The best-performing model weights are automatically restored.

Evaluation Metrics

Evaluation Metric	Result
Accuracy on the Test Set	0.9899
Loss on the Test Set	0.0414
Pixel Accuracy	98.94%
Intersection over Union (IoU)	50.19%
Dice Coefficient	51.67%

Thank you!

Author: Mariana-Ionela Muntian
Technical University of Cluj-Napoca
Bachelor of Computer Science, 3rd year
Module: Image Processing