



HUST

ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ONE LOVE. ONE FUTURE.



ĐẠI HỌC
BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY
OF SCIENCE AND TECHNOLOGY

COLONOSCOPY POLYP SEGMENTATION AND NEOPLASM CHARACTERIZATION

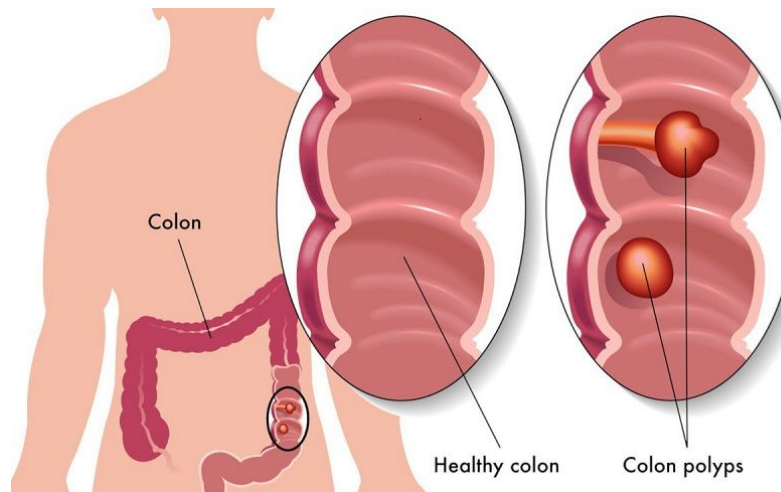
Dinh Viet Cuong-20213566
Chu Quang Tung-20213594

ONE LOVE. ONE FUTURE.

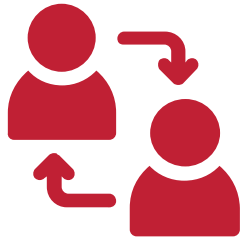
- 1. Introduction**
- 2. Related Work**
- 3. Proposed Method**
- 4. Data Preparation**
- 5. Experiment**
- 6. Conclusion**

Colonoscopy Detection Importance

Colorectal cancer is one of the leading causes of cancer-related deaths worldwide. Early detection and removal of polyps, especially neoplastic polyps, significantly reduce cancer risk. Colonoscopy is a gold standard for detection, but it relies heavily on the skill and experience of clinicians, making automation and accurate analysis vital.

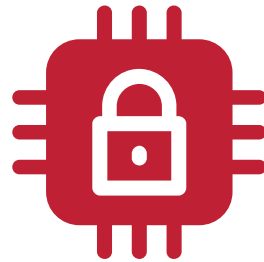


Challenges In Manual Polyp Detection



Subjectivity

Visual interpretation varies between clinicians.



Miss Rates

Small or flat polyps are often missed.



Time-Intensive

Reviewing colonoscopy videos frame-by-frame is laborious.

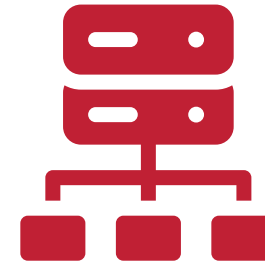
Solution

Artificial Intelligence (AI) can enhance the effectiveness of colonoscopy by:



Segmenting Polyps

Identifying the exact location and boundaries of polyps in real time.



Characterizing

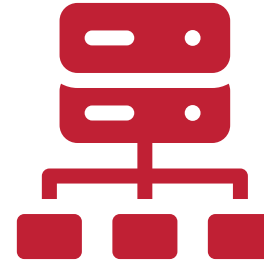
Differentiating between neoplastic (cancerous or precancerous) and non-neoplastic (benign) polyps using advanced machine learning techniques.

Objectives



Segmenting Polyps

Accurately segment polyps from colonoscopy images, leveraging models like U-Net++



Characterizing

Classify segmented polyps into neoplastic or non-neoplastic categories using deep learning.

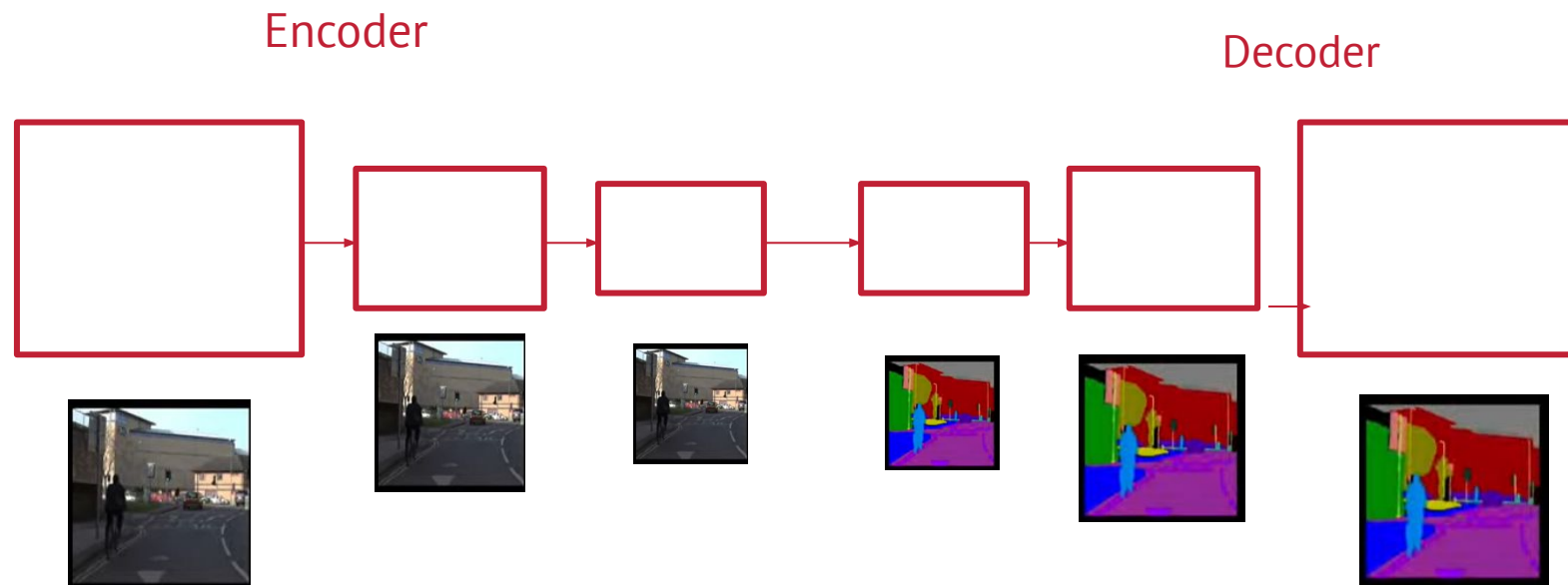
Potential Impact

The successful implementation of this project could:

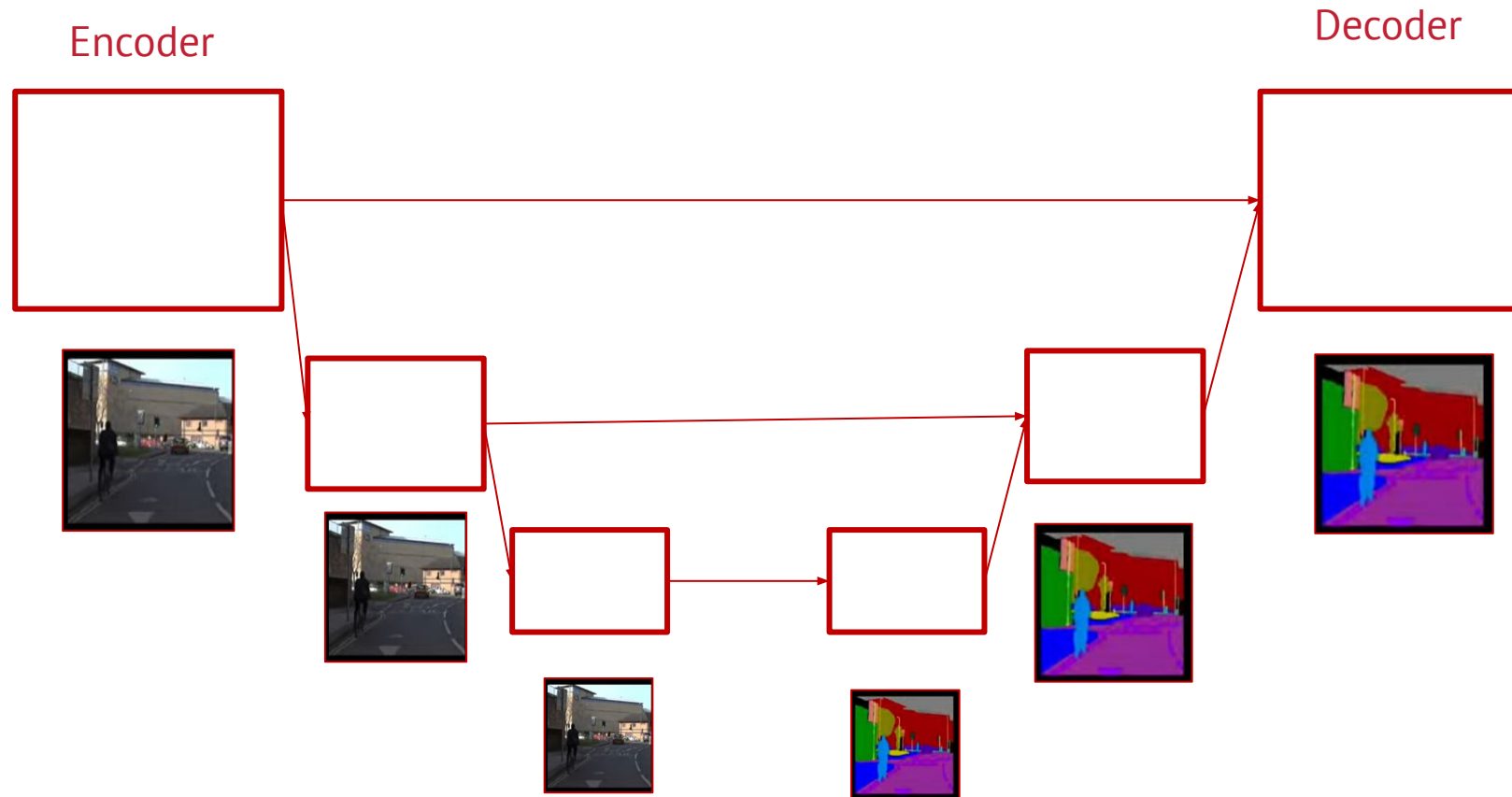
- Reduce the workload on clinicians.
- Improve early detection rates, especially for high-risk polyps.
- Enhance diagnostic accuracy and support clinical decision-making.

UNET

Convolutional Neural Network (CNN) designed to perform semantic segmentation — where every pixel in the image is classified into a category.



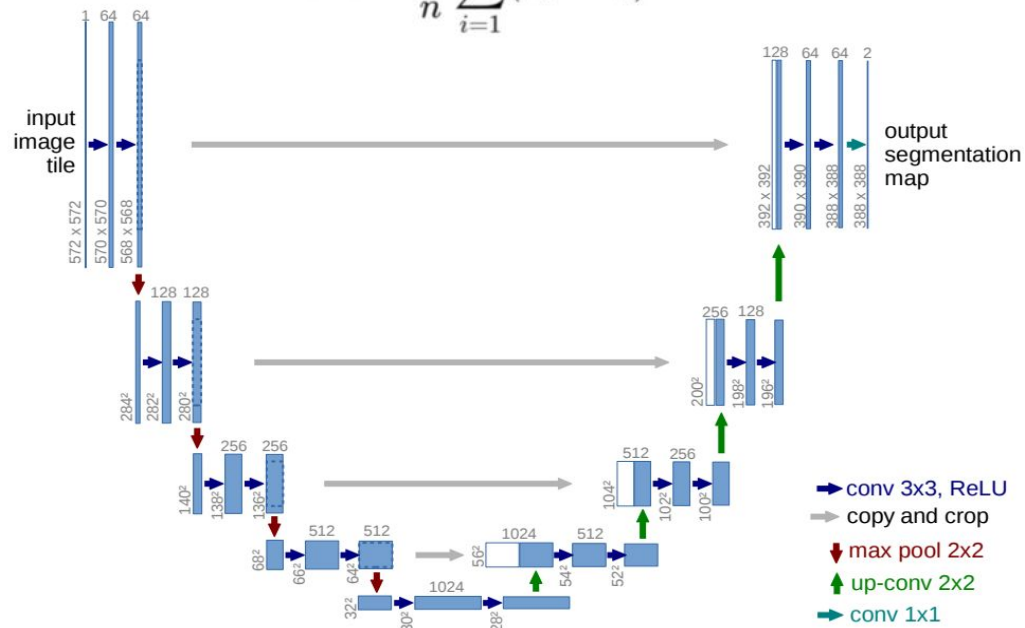
II,RELATED WORK



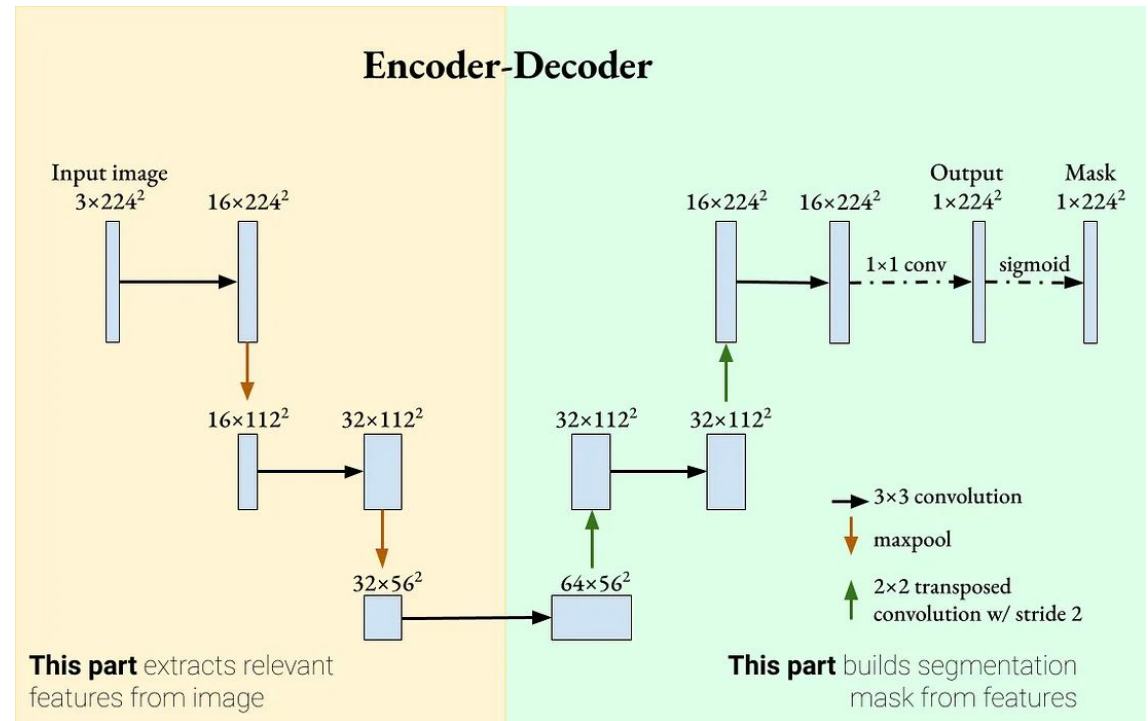
II,RELATED WORK

2

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

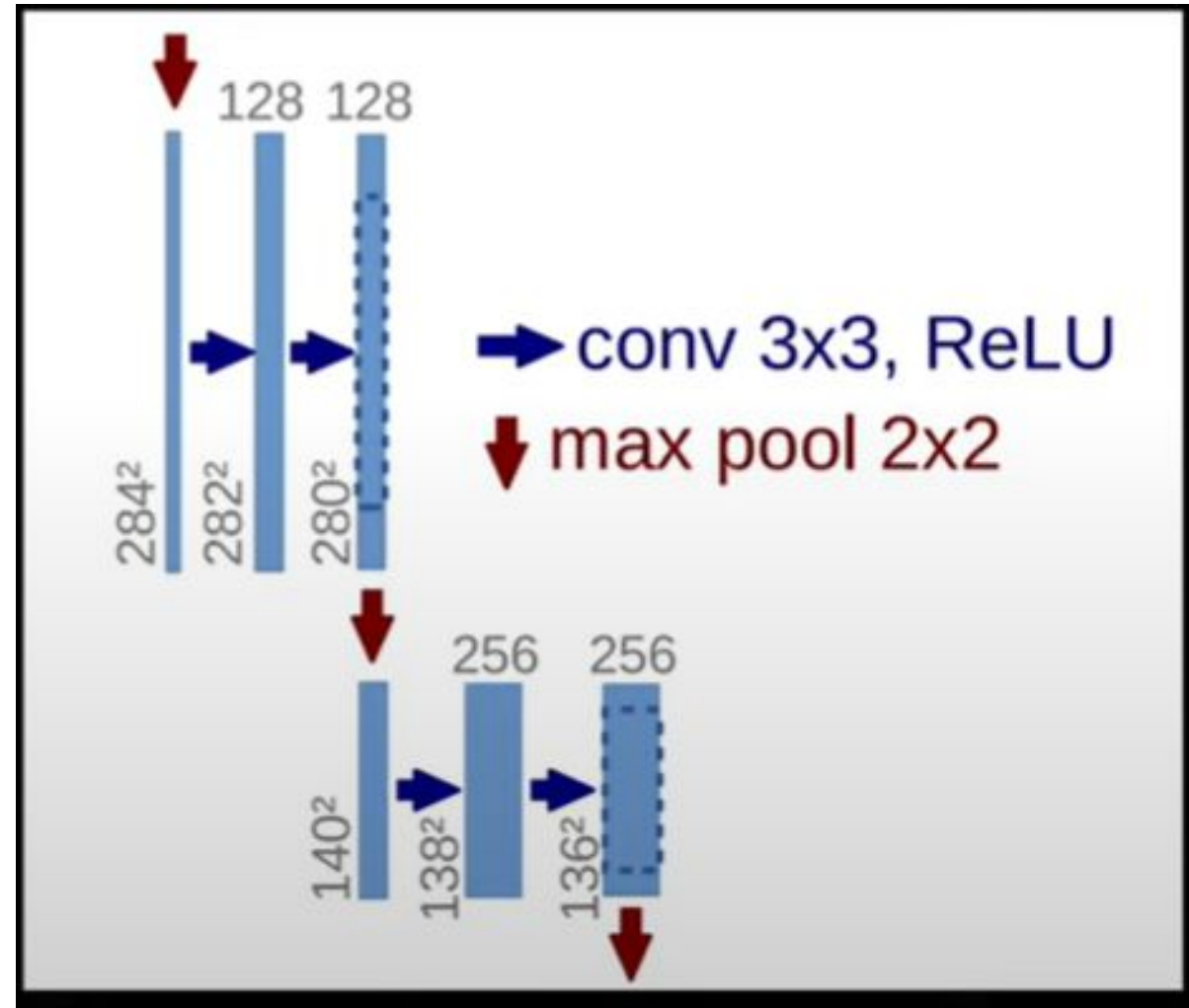


Encoder-Decoder



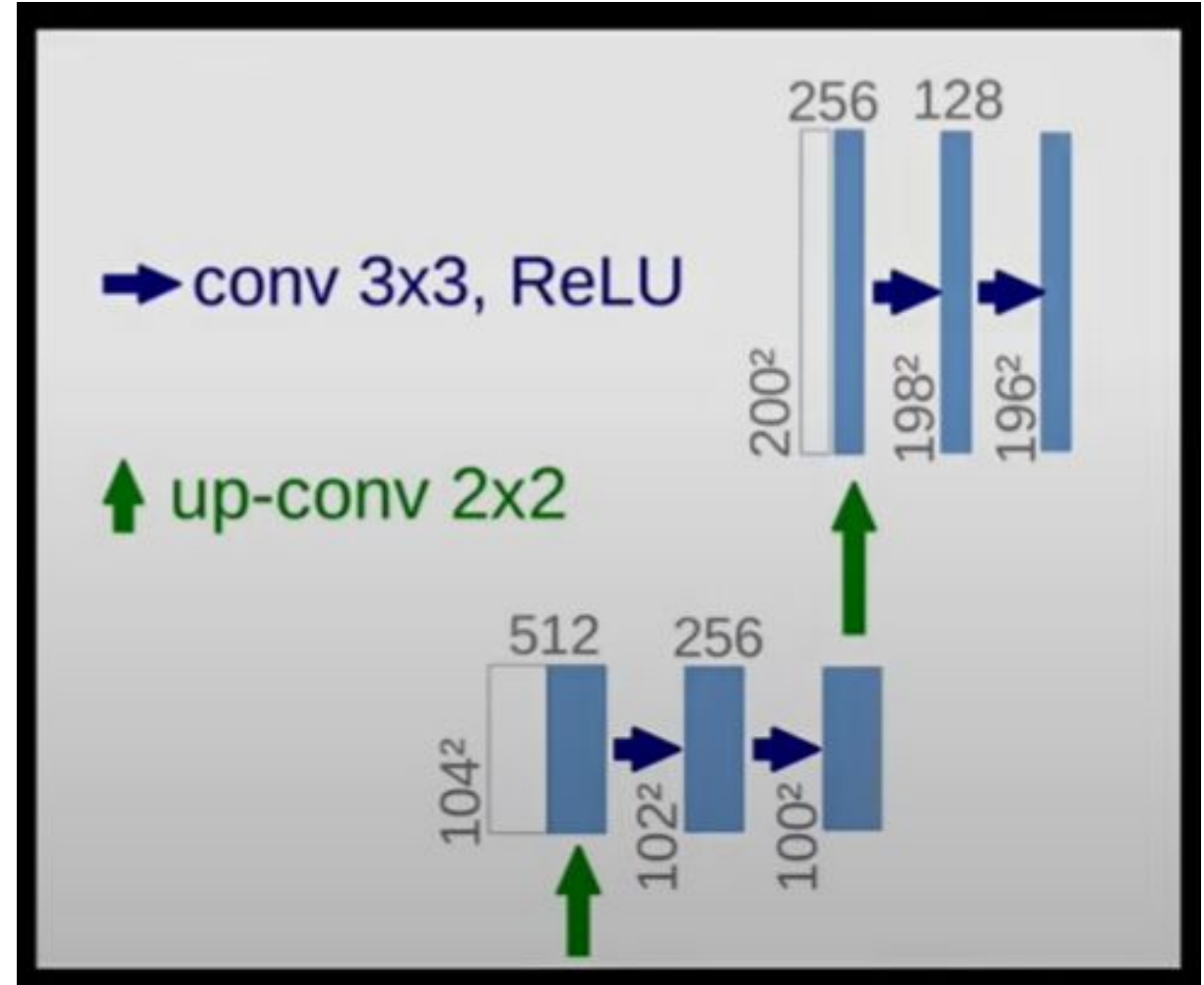
Encoder

- Series of repeated 3x3 convolutional layers extract features from the input image
 - Each convolutional layers is followed by a ReLU layers
 - 2x2 max pooling layers to downsample the feature maps → reduce spatial dimensions while increasing the depth (number of channels) of the feature maps
- allowing the model to capture more abstract features at different scales.
- Double channels after each downsampling operation



Decoder

- Series of repeated 3x3 convolutional layers extract features from the input image
- Each convolutional layers is followed by a ReLU layers
- Upsampling, followed by 2x2 conv layers
- → increase the spatial dimensions of the feature maps
- Halve channels after upsampling conv



Connecting Paths

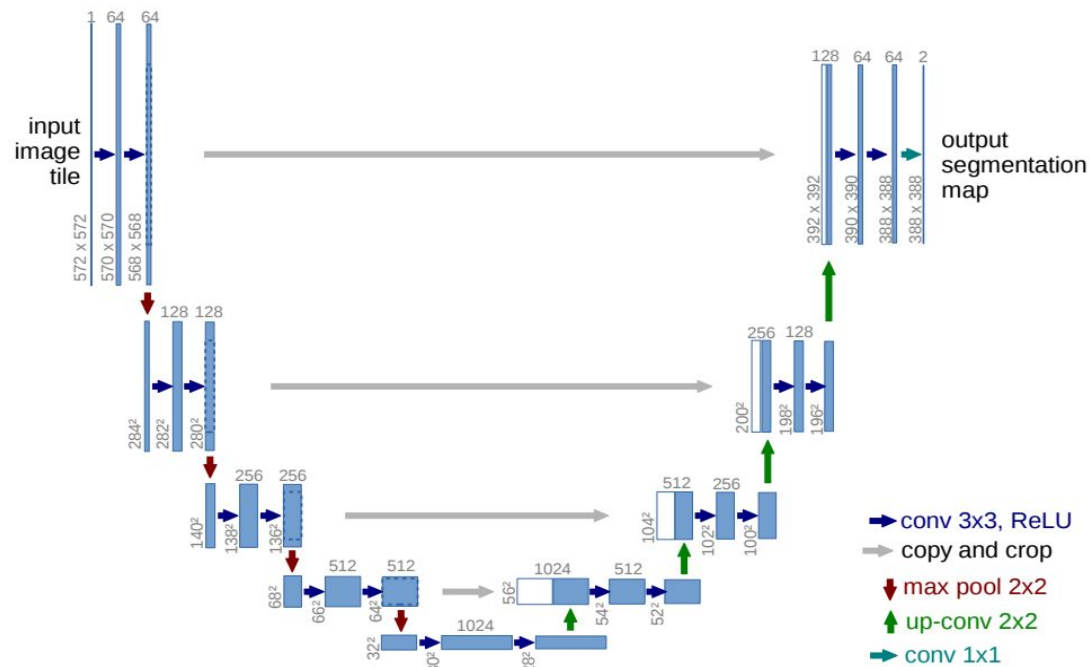
Feature Concatenation:

- Incorporates skip connections from the encoder
- Features from corresponding layers in the encoder are concatenated with the upsampled features in the decoder.
→ *restore spatial information that may have been lost during downsampling.*

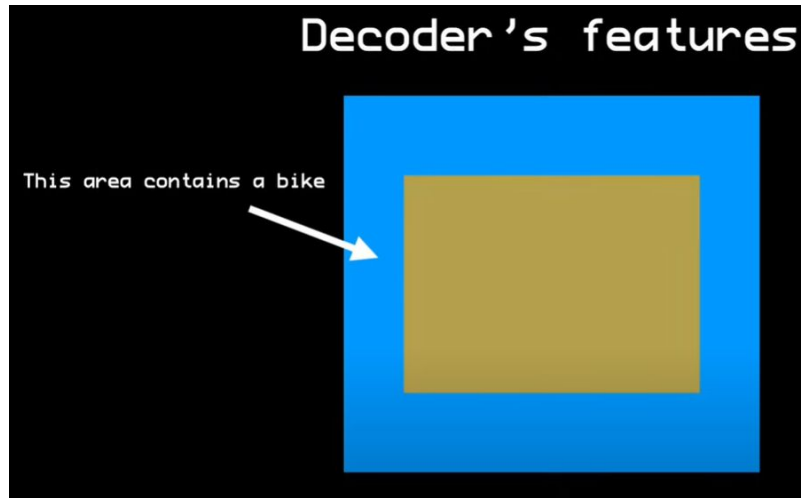
Combining Information:

- By combining features from both the encoder and decoder, the model can leverage both semantic information (from the decoder) and spatial information (from the encoder)

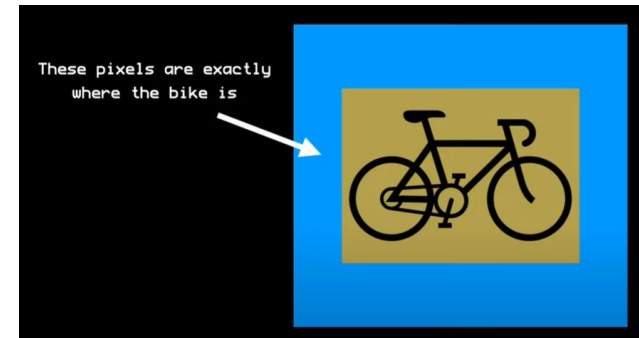
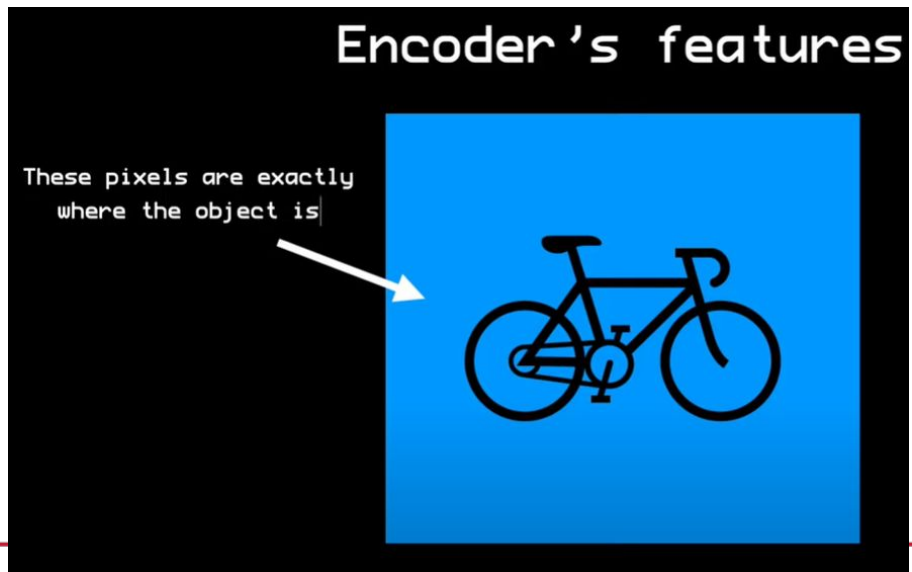
2



II, RELATED WORK

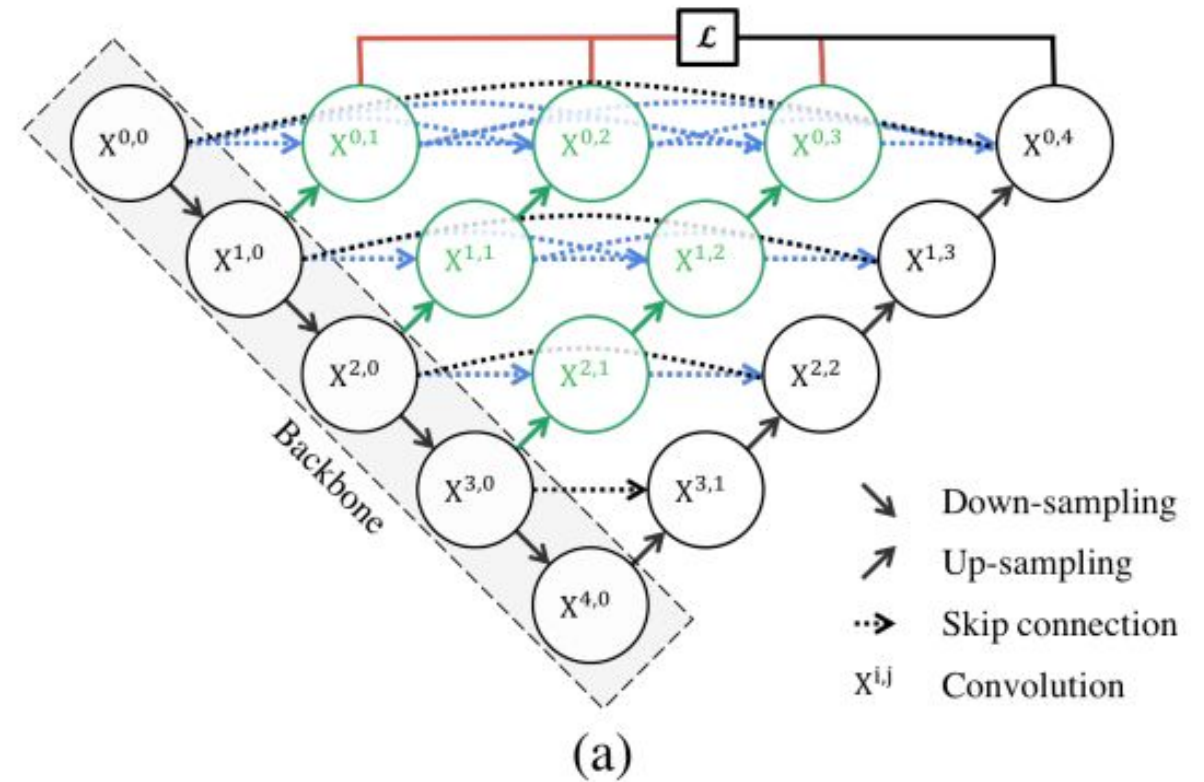


+



III, PROPOSED METHOD

Aspect	Unet	Unet++
Skip Connections	Simple skip connections (concatenation of encoder and decoder features).	Nested dense skip connections between encoder-decoder at various levels.
Main Goal	Capture features at different resolutions to localize and classify regions.	Improve feature reuse, gradient flow, and fine-grained segmentation by reducing the semantic gap between encoder and decoder.
Feature Refinement	Minimal refinement in skip connections.	Each feature map is progressively refined at different levels.
Accuracy	May struggle with small objects and fine edges due to limited connections.	Better at segmenting objects with irregular shapes and fine edges.



UNET++ (NESTED UNET)

Model selection rationale:

1. Segmentation of Polyps with Complex and Fine Boundaries
 - Polyp segmentation requires precise identification of irregular and fine shapes (especially for small or partially visible polyps).
 - UNet's simple skip connections may not be enough to recover such fine details.
 - UNet++ has nested dense skip connections, meaning the decoder gets progressively refined feature maps from different levels, leading to better feature reconstruction.
2. Bridging the Semantic Gap
 - In UNet, the encoder extracts high-level features, but the decoder tries to reconstruct low-level spatial details, leading to a semantic gap.
 - UNet++ introduces intermediate dense skip pathways that progressively refine and propagate features from the encoder to the decoder.
 - This improves the segmentation mask's accuracy near boundaries (e.g., transitions from background to red or green regions).

Resnet-50 Encoder

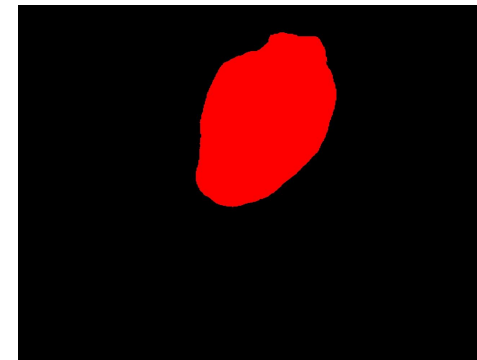
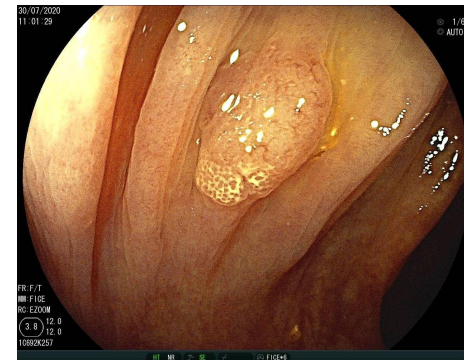
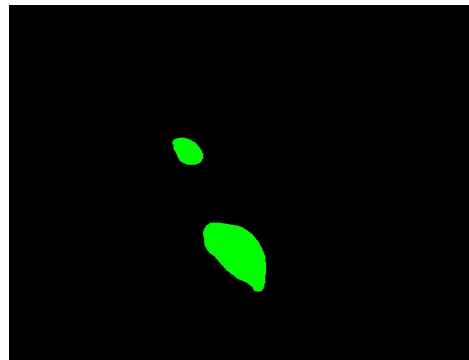
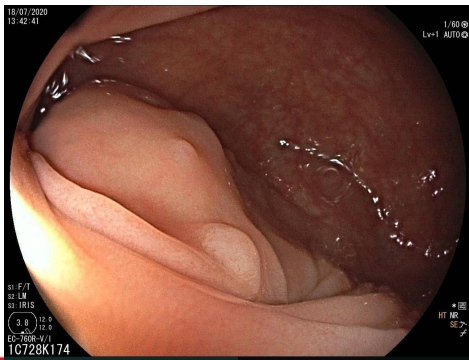
Resnet-50 as the backbone for feature extraction.

- Pre-trained on ImageNet (a large dataset of millions of labeled images).
- The encoder already "knows" how to detect low-level features (e.g., edges, corners) and high-level features (e.g., shapes, patterns).
- Speeds up training and improves performance, especially with limited training data (like in medical imaging).
- Instead of learning from scratch, the model starts from a strong baseline of general image features.
- ResNet-50 has 50 layers, making it deep enough to capture complex patterns in the data (like irregular shapes of polyps).

Dataset Description

BKAI-IGH NeoPolyp-Small is a public dataset released by BKAI Research Center, Hanoi University of Science and Technology in collaboration with Institute of Gastroenterology and Hepatology (IGH), Vietnam.

- Size: 375.42MB (2200 files)
- Number: 1200 images (1000 - train and 200 - test)
- Ground-truth explanation
- **Red color:** neoplastic polyps - Dangerous, likely cancerous.
- **Green color:** non-neoplastic polyps - Less dangerous, benign.
- **Black color:** background

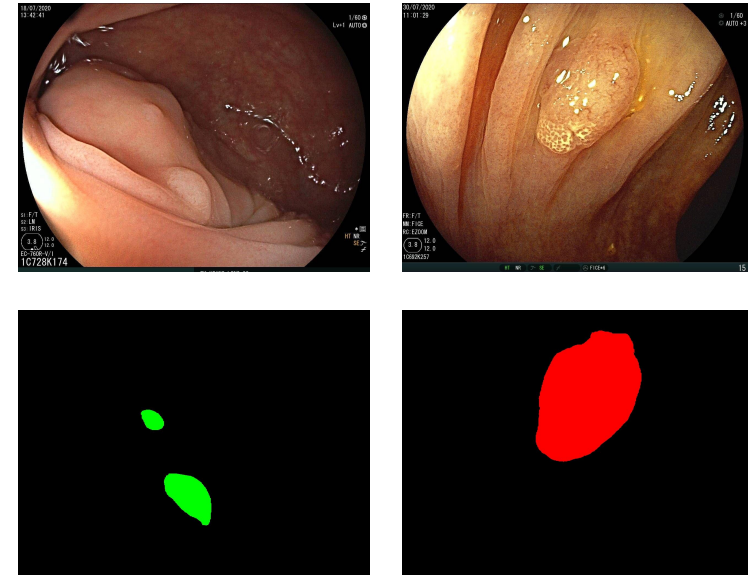


Read Mask

In order to prepare the ground-truth segmentation masks for training a deep learning model.

- Convert img to HSV (Hue, Saturation, Value)
- Detecting Red Regions
- Detecting Green Regions
- Combining Red and Green Masks

Pixel Color	Class	Meaning
Black	0	Background
Red	1	Neoplastic polyps
Green	2	Non-neoplastic polyps



Data Augmentation (Train Set)

- Data augmentation is the process of randomly altering the input images to create a more diverse dataset. → prevent overfitting and improve generalization ability of the model
- Horizontal Flip
- Vertical Flip
- Random Gamma: Brighten or Darken the image
- RGB Shift: Change the color balance

Data Preprocessing (Train + Val Set)

Resize: image to 256x256

Scaling vs Formatting

- Normalizes the image pixel values to match the ImageNet pre-trained model's expectations. (Pixel values are normalized to have a mean of 0 and a standard deviation of 1.)
- Converts the NumPy array image to a PyTorch Tensor (with channels-first format (C, H, W)).

Model Inference (Forward Pass)

- The preprocessed image tensor is passed through the UNet++ model.
- The model processes the image through multiple encoder and decoder layers, generating a feature map at each stage.
- The final output from the model is a segmentation map containing class scores (logits) for each pixel, indicating the likelihood that a pixel belongs to each class (background, neoplastic polyp, or non-neoplastic polyp).

Model Inference (Forward Pass)

- The preprocessed image tensor is passed through the UNet++ model.
- The model processes the image through multiple encoder and decoder layers, generating a feature map at each stage.
- The final output from the model is a segmentation map containing class scores (logits) for each pixel, indicating the likelihood that a pixel belongs to each class (background, neoplastic polyp, or non-neoplastic polyp).

Training & Validation

- Epoch: 200
- Batch Size: 8
- Learning Rate: 0.0001 with Adam optimizer
- Use cross-entropy loss (between the predicted mask and the ground-truth mask.) → helps the model learn to classify each pixel as background, red (neoplastic polyps), or green (non-neoplastic polyps).

$$\mathcal{L}(y, \hat{y}) = - \sum_{c=1}^C y_c \log(\hat{y}_c)$$

Results

Original Image: 9fc7330398846f67b5df7cdf3f33c3ca.jpeg



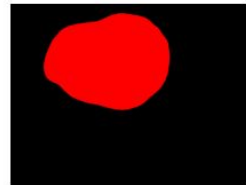
Predicted Mask



Original Image: 3c84417fda8019410b1fcf0625f608b4.jpeg



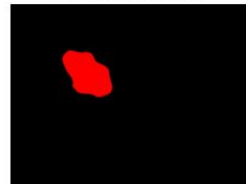
Predicted Mask



Original Image: a9d45c3dbc695325ded465efde988dfb.jpeg



Predicted Mask



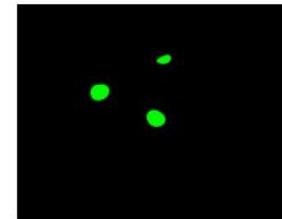
Original Image: 7936140a2d5fc1443c4e445927738677.jpeg

Predicted Mask

Original Image: 7936140a2d5fc1443c4e445927738677.jpeg



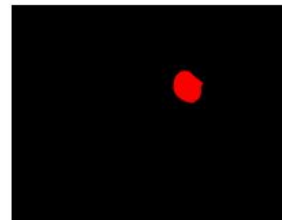
Predicted Mask



Original Image: 2cd066b9fdb3bbc04a3afe1f119f21b.jpeg



Predicted Mask



Evaluations

Dice coefficient: evaluate the similarity between the predicted segmentation (output of the model) and the ground truth segmentation

- 1 means perfect overlap (predicted and ground truth are identical).
- 0 means no overlap.

$$\frac{2*|X \cap Y|}{|X|+|Y|}$$

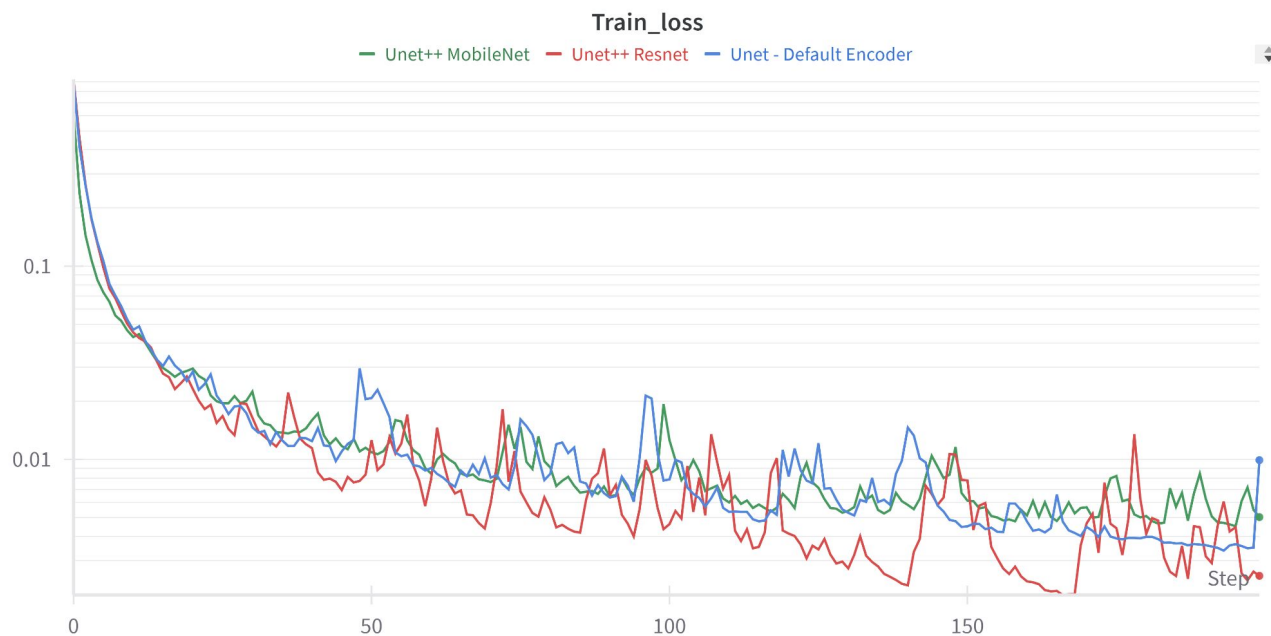
Where:

- X : Set of pixels predicted as belonging to the target class (e.g., neoplastic polyp).
- Y : Set of pixels in the ground truth for the target class
- $|X \cap Y|$: The number of pixels that are correctly predicted (intersection between X and Y).
- $|X|$: The total number of pixels predicted as belonging to the target class.
- $|Y|$: The total number of pixels in the ground truth for the target class.

Model	Dice Coefficient
Unet++ with Resnet50 Encoder	0.74620
Unet with Resnet34 Encoder	0.73228
Unet with mobilenetV2 Encoder	0.67029

Our proposed model slightly outperforms other models in terms of the Dice score.

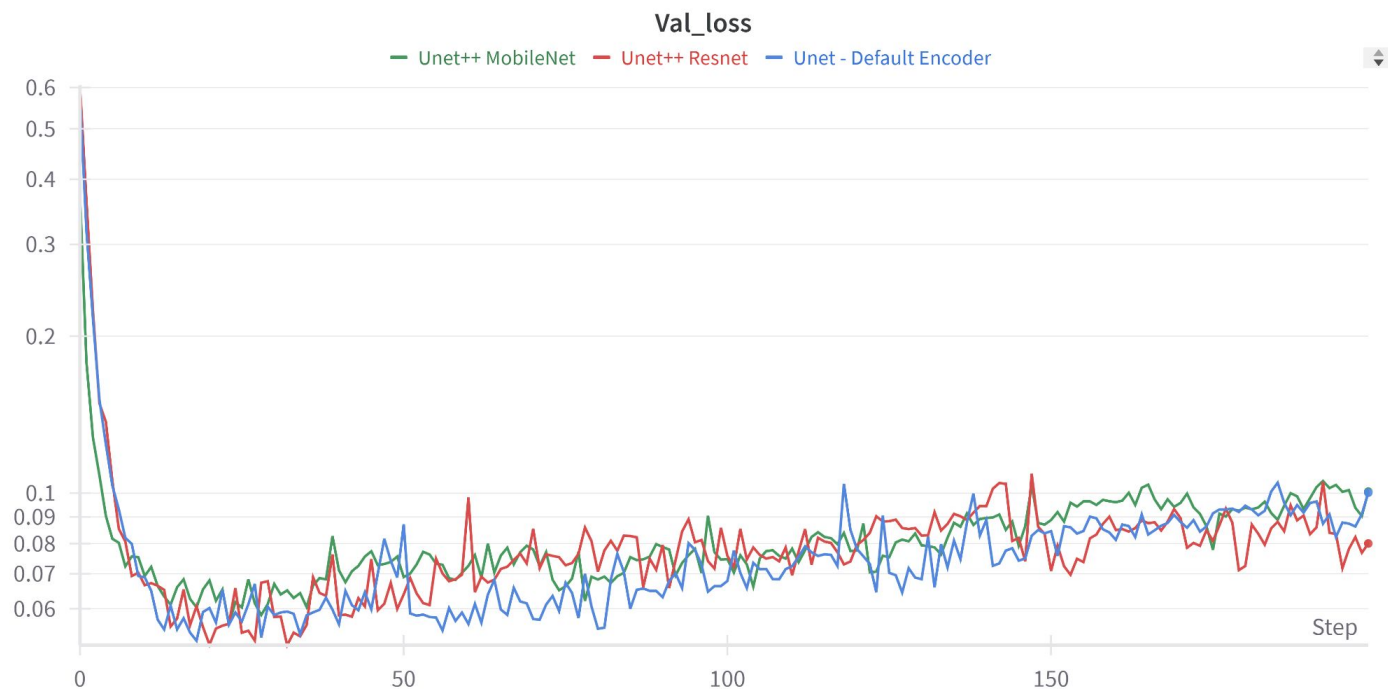
VI, Result & Conclusion



UNet++ vs UNet:

- UNet++ (both ResNet and MobileNet encoders) outperforms the default UNet in both training and validation loss.
- This is due to the nested skip connections in UNet++, which refine the feature maps and help reduce the semantic gap between the encoder and decoder.

VI, Result & Conclusion



Encoders Comparison:

- MobileNet encoder results in the most stable and low loss values, likely due to its lightweight architecture and strong generalization capabilities.
- ResNet encoder provides rich features but requires careful regularization to avoid overfitting.
- Default encoder performs poorly, as it may not extract meaningful features for complex segmentation tasks.

In conclusions,

- For Faster Convergence and Lightweight Models: Continue with UNet++ with MobileNet.
- For Higher Accuracy and More Complex Features: UNet++ with ResNet may still be beneficial if you can address its fluctuations (e.g., through learning rate tuning or regularization).
- For Simpler Use Cases: Avoid using the default UNet encoder if high performance is needed, as shown by the consistently higher validation loss.

Future work,

- Integrating attention mechanisms into UNet++ could further improve the model's ability to focus on relevant regions of interest, reducing false positives and false negatives.
- The use of multi-scale feature extraction techniques can be investigated to handle polyps of varying sizes more effectively.
- Employing transformer-based encoders could improve the model's capability to capture long-range dependencies within colonoscopy images.
- Deploying these models in real-time clinical settings and validating their performance in diverse datasets from different institutions would ensure their generalizability and practical applicability in real-world scenarios.

A large, stylized graphic of the letters 'HUST' in white, set against a red background. The letters are composed of a dense pattern of small dots, creating a halftone effect. The graphic is centered on the left side of the slide.

HUST

THANK YOU !