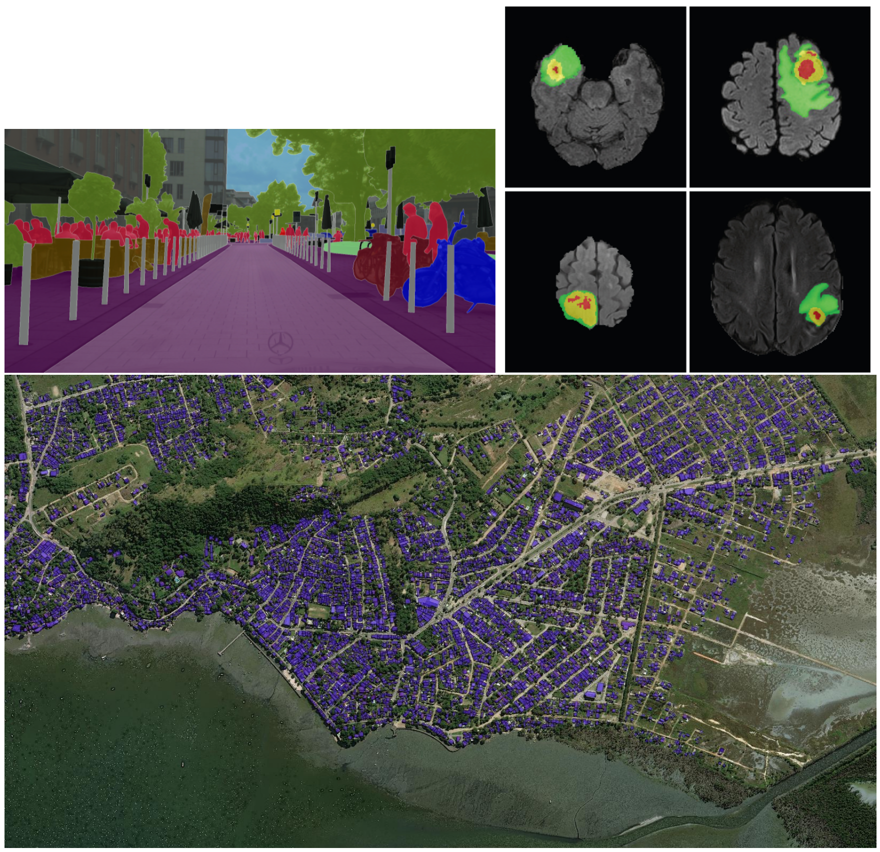
Medical Image Segmentation

Diagnosing gastrointestinal tumors

Ahmadreza Baqerzadeh

1 Problem statement

Segmentation in artificial intelligence is one of the very important areas that aims to determine the precise boundaries of one or multiple targets, requiring each pixel of the image to be labeled. Segmentation has various types, with three commonly used ones being semantic segmentation, instance segmentation, and panoptic segmentation (instance segmentation + semantic segmentation). Segmentation can be used in various fields such as medical image segmentation for disease diagnosis and urban image segmentation for self-driving cars, and so on.



Segmentation is also one of the most commonly used areas in the medical field for disease diagnosis. In medical imaging, segmentation includes: 1.Segmentation of brain MRI images for detecting brain tumors and studying neurological diseases. 2.Segmentation of lung images for lung cancer diagnosis and prognosis. 3.Segmentation of breast MRI images for detecting normal breast tissue and breast cancer, among others.

Segmentation of gastrointestinal images obtained with integrated magnetic resonance imaging (MRI) and MR-Linacs also aids in the detection of gastrointestinal tumors.Diagnosing gastrointestinal tumors is one of the challenges faced by doctors in treating patients. Approximately half of the patients are eligible for radiation therapy, and this procedure should be performed in a way that does not harm the intestines and stomach. By using artificial intelligence (specifically in the field of segmentation), these areas can be identified, helping doctors to expedite patient treatment. Treating patients using this method can reduce daily treatment time from one hour (which can be difficult for patients) to 15 minutes.

Challenges:

1. Complexity of images: Medical images related to the gastrointestinal system have complex structures that can make the analysis and accurate diagnosis of cancerous masses difficult.

2. Size and shape variations of tumors: Gastrointestinal tumors can have different sizes and shapes, which can make their accurate and reliable diagnosis challenging for physicians.

3. Natural variations in the gastrointestinal system: The presence of natural variables such as bowel movements, digestion, and gastric motility can make the diagnosis of gastrointestinal tumors more difficult and lead to errors and incorrect results in segmentation.

AI targets:

1. Accurate identification of cancerous masses: As mentioned, for high doses of X-ray radiation, they should be directed towards the tumor and should not cause any harm to the stomach and intestines. Therefore, tumor segmentation needs to be done with high accuracy.

2. Reduction of errors caused by different dimensions of masses: Cancerous masses can have various dimensions, and their type should be detected in all dimensions.Misdiagnosing any type of these masses with different dimensions has a detrimental effect on the treatment of patients.

Medical targets:

1. Faster detection of masses and quicker treatment process: According to studies conducted by Carbone universities, in normal conditions and manually, it takes approximately 1 hour to identify the precise area of these masses for daily treatment and deliver radiation to the masses. This time can be reduced to 15 minutes using artificial intelligence.

2. Early detection of masses: Early detection improves the treatment process and helps physicians diagnose these masses earlier and initiate treatment before the patient's condition worsens. Artificial intelligence can increase the chances of patients' survival.

3. Accurate and more reliable diagnosis: Although manually determining the area of masses by physicians is somewhat accurate, in some cases, it may introduce human errors and jeopardize the patient's health. Using artificial intelligence helps make this process more reliable.

2 Related Works

1.U-Net: Convolutional Networks for Biomedical Image Segmentation

Article: link

Code: link

U-Net is a type of convolutional neural network (CNN) that has been widely used for biomedical image segmentation tasks. It was specifically designed for handling the challenges of segmenting biomedical images, such as medical scans and histological slides.

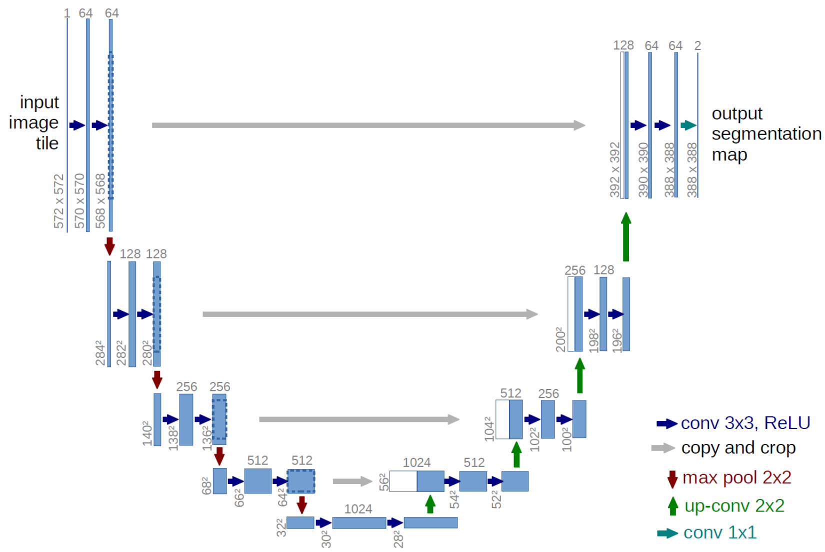
The name "U-Net" comes from the shape of its architecture, which resembles an upside-down "U". The network consists of an encoder path and a decoder path. The encoder path is responsible for capturing context and extracting hierarchical features from the input image, while the decoder path reconstructs the segmented image based on the extracted features.

The encoder path typically consists of a series of convolutional and pooling layers that gradually reduce the spatial dimensions of the input image while increasing the number of channels (features). This allows the network to capture high-level contextual information at different scales.

The decoder path uses upsampling and concatenation operations to recover the spatial resolution lost during the encoding process. Upsampling layers expand the feature maps back to the original size, while concatenated skip connections combine the feature maps from the corresponding encoding layers. This enables the network to preserve detailed information from earlier layers and produce accurate segmentation maps.

U-Net also uses skip connections between the encoder and decoder paths to facilitate information flow and improve segmentation accuracy. These skip connections allow the network to access both low-level and high-level features, enabling precise localization of objects in the segmented image.

Overall, U-Net has been proven effective for a wide range of biomedical image segmentation tasks, including organ segmentation, tumor detection, cell segmentation, and more. Its architecture and design principles have inspired numerous variations and adaptations within the field of medical imaging.

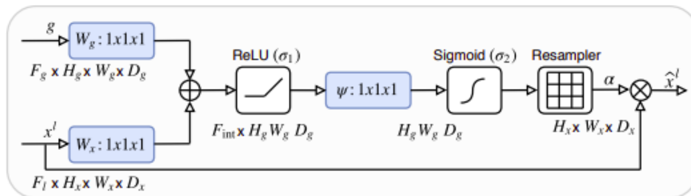


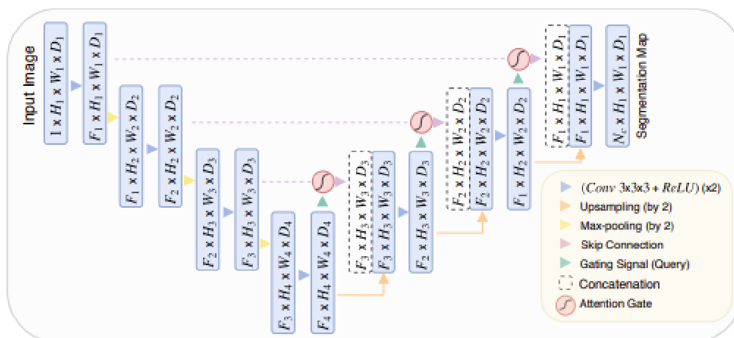
2.Attention U-Net: Learning Where to Look for the Pancreas
Article: link
Official code: link

The article introduces two methodologies: Fully Convolutional Network (FCN) and Attention Gates for Image Analysis.

FCN is a type of convolutional neural network that outperforms traditional approaches in medical image analysis. It learns domain-specific image features using stochastic gradient descent optimization and shares learned kernels across all pixels. FCN's convolution operations effectively exploit the structural information in medical images. It has been successfully applied to tasks such as cardiac MR, brain tumor, and abdominal CT image segmentation.

Attention Gates (AGs) are integrated into the U-Net architecture, which is commonly used for image segmentation tasks due to its good performance and efficient GPU memory usage. AGs aim to capture a large receptive field and semantic contextual information. They progressively downsample the feature-map grid and identify salient image regions using attention coefficients. These coefficients are computed using additive attention, which allows for focus on subsets of target structures. AGs suppress feature responses in irrelevant background regions without the need to crop a region of interest between networks.





3.UNET 3+: A FULL-SCALE CONNECTED UNET FOR MEDICAL IMAGE SEGMENTATION

Article: link

Official code: link

The article describes the UNET 3+ architecture, which is a neural network designed for medical image segmentation. The methods used in the article include the following:

1. UNET Architecture: The UNET 3+ architecture is an extension of the original UNET architecture. UNET is a popular architecture known for its effectiveness in image segmentation tasks. It consists of an encoder-decoder structure with skip connections that enable the fusion of information at different resolutions.

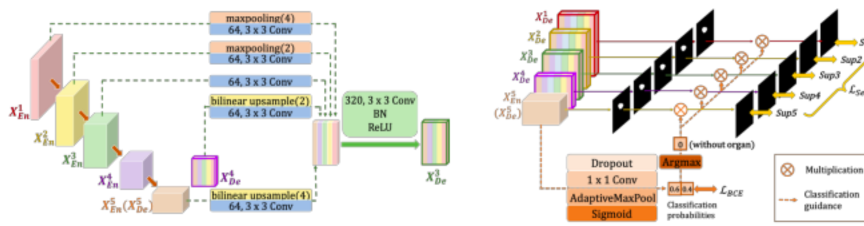
2. Full-Scale Skip Connection: UNET 3+ introduces a full-scale skip connection that allows information from all resolutions to be fused at each level of the network. This helps capture both local and global information effectively, improving the segmentation accuracy.

3. Densely Connected Convolutional Blocks: The UNET 3+ architecture incorporates densely connected convolutional blocks. These blocks enhance feature learning and promote information flow across different layers of the network. Densely connected connections allow each layer to directly access the feature maps of all preceding layers, facilitating the propagation of information.

4. Training and Evaluation: The authors train and evaluate the UNET 3+ architecture using various medical imaging datasets. The datasets include tasks

such as tumor segmentation and organ segmentation. Training involves optimizing the network parameters using suitable loss functions and optimization algorithms. Evaluation is performed by measuring the segmentation accuracy and comparing the results with state-of-the-art segmentation methods.

By incorporating these methods, the UNET 3+ architecture demonstrates improved segmentation accuracy and performance compared to existing methods. The combination of the UNET architecture, full-scale skip connections, and densely connected convolutional blocks contributes to the effectiveness of the proposed approach.



4.Stepwise Feature Fusion: Local Guides Global
Article: link
Official code: link

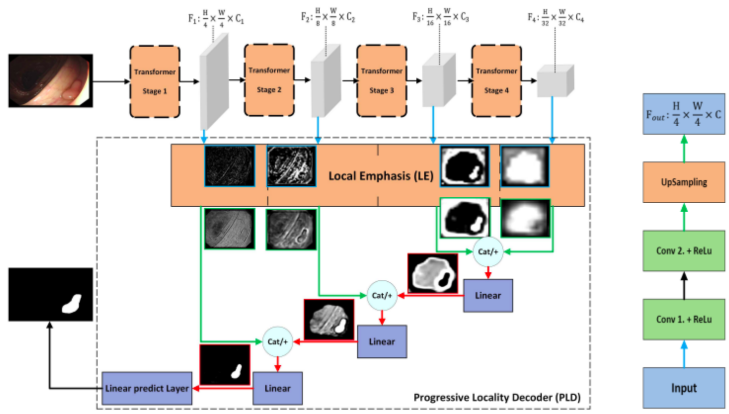
The methodology described in this section focuses on enhancing the generalization ability and multi-scale feature processing capability of the model for polyp segmentation.

1. Transformer Encoder: Instead of using a CNN as the encoder, a Transformer based on a pyramid structure is employed to provide better generalization ability and multi-scale feature processing. The encoder design from PVTv2 and Segformer is adopted, utilizing convolution operations to replace the positional encoding operation of traditional Transformers. This ensures consistency of spatial information and maintains excellent performance and stability.

2. Aggregate Local and Global Features Stepwise (PLD): The authors introduce a novel multi-stage feature aggregation decoder called PLD (Pyramid Locality Decoder) to address the lack of local and detailed information processing in existing Transformer models. PLD consists of the Local Emphasis (LE) module and the Stepwise Feature Aggregation (SFA) module. LE module uses the local receptive field of convolution kernels to emphasize neighboring features and reduce attention dispersion. SFA progressively fuses features from different levels in the feature pyramid, guiding the attention of the model to critical regions.

3. Stepwise Segmentation Transformer: Two models, SSFormer-S (Standard) and SSFormer-L (Large), are proposed based on different encoder scales. These models achieve state-of-the-art performance and competitive results in various polyp segmentation benchmarks, as well as in the ISIC-2018 and 2018 DATA Science Bowl competitions.

Overall, the methodology involves utilizing Transformer encoders, introducing PLD for multi-stage feature aggregation, and developing SSFormer models for polyp segmentation tasks.



(a) Overview of SSFormer (b) LE

5.DDANet: Dual Decoder Attention Network for Automatic Polyp Segmentation

Article: link

Official code: link

The DDANet architecture described in the section consists of three key components: residual blocks, squeeze and excitation blocks, and the overall DDANet architecture.

1. Residual Block: The authors introduce a residual block to address the challenges of vanishing or exploding gradients as the network depth increases. The residual block comprises two 3x3 convolutions, batch normalization, and a ReLU activation function. It also includes a skip-connection that connects the input with the output of the residual block, facilitating better gradient flow during backpropagation.

2. Squeeze and Excitation Block: To address the equal importance treatment of every feature channel in CNNs, the authors introduce a squeeze and excitation layer. This layer acts as a channel-wise attention mechanism, re-weighting each feature channel to create a more accurate feature map. It consists of two steps: compressing feature maps using global average pooling and passing them through a 2-layer neural network to scale the feature channels.

3. DDANet Architecture: The proposed DDANet architecture follows an encoder-decoder design, similar to ResUNet++. It combines the features of residual learning and the squeeze and excitation network. DDANet consists of a single encoder shared by dual decoders. The encoder network includes four encoder blocks, while each decoder network includes four decoder blocks. Skip connections are used to fetch features from earlier layers at their original resolution, increasing feature representation strength and aiding gradient flow. The decoder blocks output a segmentation mask and a reconstructed grayscale image.

Overall, the DDANet architecture incorporates residual blocks, squeeze and excitation blocks, and an encoder-decoder design to improve performance in image segmentation tasks.