Gastro-Intestinal Tract Segmentation Using Multi-Task Learning and FiLM

Bryan Chia, Helen Gu, Nicholas Lui

CS 231N Spring 2022, Statistics Department, Stanford University

Stanford Computer Science

Background / Introduction

In 2019, approximately **2.5 million people were being treated for gastro-intestinal tract** cancer worldwide using **radiation therapy**. Radiation therapy requires manually tracing out the stomach and intestines in MRI scans in order to ensure that x-ray beams are directed to avoid those crucial organs. Deep learning and computer vision methods can be employed to **help segment the stomach and intestines**, improving speed and efficiency of treatment.

Problem Statement

Our project investigates methods to improve upon baseline methods of semantic segmentation in medical imaging. Traditionally, improving upon the performance of semantic segmentation models requires training on larger quantities of labelled data; however, **labelling GI scans is time-consuming and laborious**. We aim to explore whether we can improve the performance of conventional semantic segmentation models by **incorporating contextual information from scans that requires no additional annotation cost**.

Dataset

Our dataset is a Kaggle dataset consisting of 85 different cases (patients) with roughly 3-5 days of scans per patient. Each day consists of ~150 scan slices. In a scan, every pixel takes on **at least 1 out of 4 classes:** (i) Stomach, (ii) Small Intestine, (iii) Large Intestine, (iv) Others/Background.

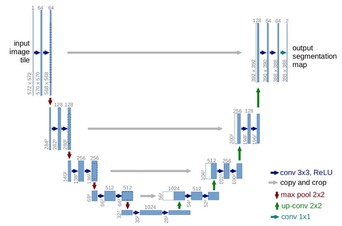
Processing steps:

1. Run length decoding for mask decoding
2. Create one mask for each organ class; masks concatenated together
3. Min-max image normalization, resizing scans to 224x224 with bilinear interpolation to retain [0,1] values for masks
4. We split our dataset in a 70/20/10 ratio. Two ways of splitting data:
   Patient-day: Entire days for a given patient are withheld from the train set; mimics Kaggle unseen test set
   Patient: Entire cases are withheld from the train set → model is not able to overfit to patient-specific features

Lee et al (2022) - UW-Madison GI Tract Image Segmentation Competition

Methods

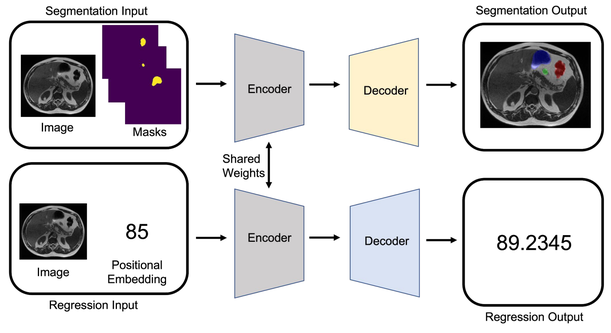
Our baseline is a **UNet**, which comprises an encoder and decoder portion. The encoder downsamples the image and extracts feature representations. The decoder upsamples the image and obtains the mask labels.



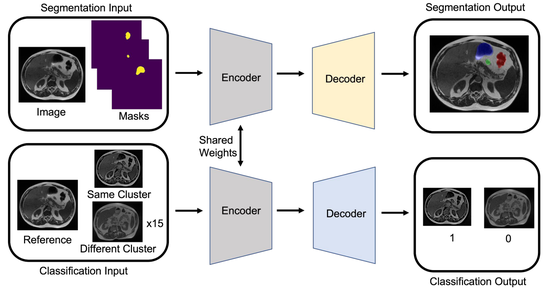
We experiment with three techniques:

- **Multi-task Learning:** Train segmentation task alongside an auxiliary task which shares similarities with the main task + doesn't require manual labelling -> additional signal!
  Auxiliary tasks:

Task 1: Predicting the slice position of a scan (in image name) from anatomical structure



Task 2: Contrastive Learning - Positives are adjacent scans from the same person and day



- **Feature-wise Linear Modulation:** Condition intermediate model output on image metadata (slice position, case number, etc.)
- **Different encoder architecture:** We experiment with a ResNet50 encoder backbone and a more parsimonious backbone with fewer Conv layers and no skip connections.

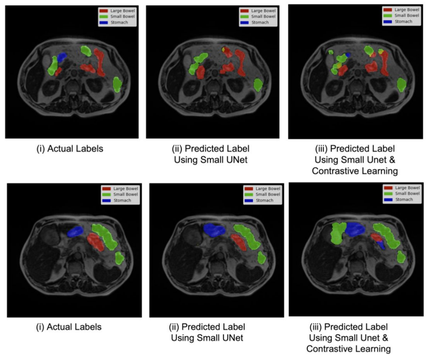
Experiments & Analysis

Model	Auxiliary Task	Overall	Large Bowel	Small Bowel	Stomach
Small UNet	-	0.8242	0.7903	0.8058	0.8763
Small UNet (Multitask)	Contrastive Learning	0.8095	0.7748	0.7944	0.8594
Small UNet (FiLM)	-	**0.8314**	**0.7957**	**0.8114**	**0.8872**

Table 1: Test Dice Coefficient Results (Patient-Day Split)

Model	Auxiliary Task	Overall	Large Bowel	Small Bowel	Stomach
Small UNet	-	0.7992	0.7507	0.7876	0.8593
Small UNet (Multitask)	Contrastive Learning	**0.8093**	**0.7763**	**0.7948**	0.8568
Small UNet (FiLM)	-	0.8026	0.7628	0.7846	**0.8605**

Table 2: Test Dice Coefficient Results (Patient Split)



Contrastive Learning *helps* by over-predicting classes with less data

Contrastive Learning *hurts* by over-predicting classes with less data

Conclusions & Future Work

1. **Multi-task learning can help or hurt**, depending on allocation of weights in loss function, and distributions of train and test data
2. Since **contrastive learning** learns global relationships between images, it **tends to over-predict the presence of small classes**
3. Adding image **metadata is universally helpful**
4. In the case of medical imaging, using a smaller architecture leads to **faster convergence** despite less expressivity
5. Potential avenues of future work:
   - Adaptive weight loss functions, improving contrastive learning setup (MoCoV2 and data augmentation), different architectures (DeepLab), ensembling