Massachusetts Institute of Technology

# Computer-Aided Segmentation of Gastrointestinal Structures

## 6.8300: Advances in Computer Vision

Srikaran Boya, Joseph Kajon, Zeki Yan

May 14th, 2024

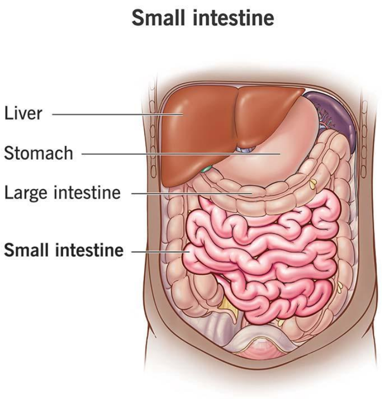
# Introduction

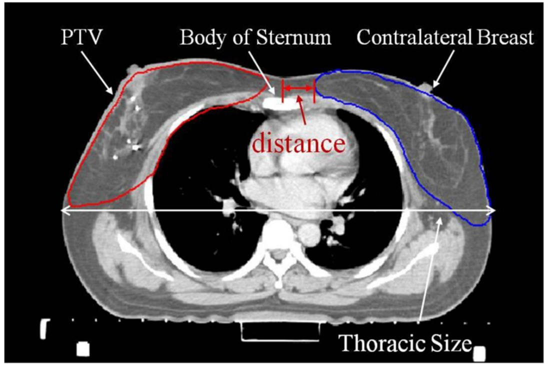
## Problem Statement

- Millions are diagnosed with GI cancer each year
- Treatment requires precise radiation therapy planning

## Current Process

- Manual segmentation of MRI scans
- Labor-intensive & Prone to errors



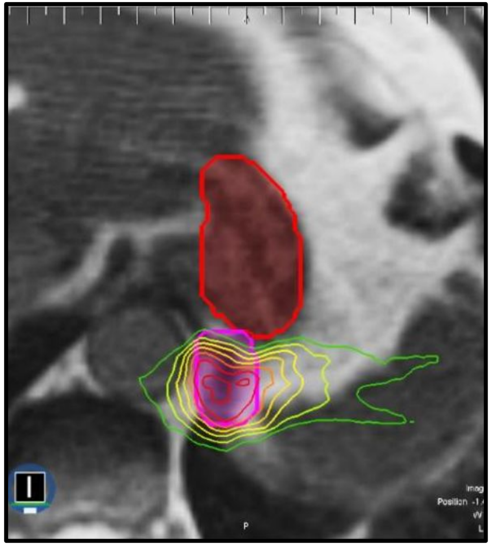

# Proposed Solution

**Objective**

- Develop a deep learning model to automate segmentation of stomach and intestines from MRI scans

**Impact**

- Expedite treatment
- Reduce errors
- Improve patient outcomes



*In figure above, the pink area is tumor, red area is stomach, and other lines are radiation dose levels*

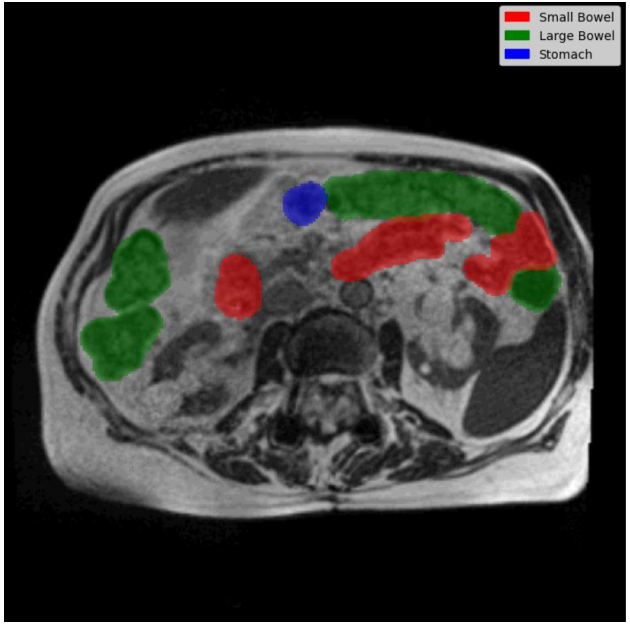
# Related work

- **U-Net Type Neural Networks**
  - Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.
- **DeepLabv3+**
  - Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation, 2017.
  - Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation, 2018.
- **R-CNNs**
  - Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation, 2014.
  - Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask r-cnn, 2018.
- **Multi-Task Learning**
  - Wenjia Bai and Dinggang Shen. Self-supervised learning for cardiac mr image segmentation by anatomical position prediction. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2019, Cham, 2019. Springer
  - Rich Caruana. Multitask learning. Machine Learning, 28:41–75, 1997.

# Dataset

**Dataset:**

- UW-Madison GI Tract Image Segmentation repository: Kaggle competition
    - 85 cases (patients)
    - MRI scans on different days (normally 3-5 days)
    - Different slices each day
- Segmentation and class information




*Red mask represents small bowel, green mask represents large bowel, and blue mask represents stomach*

# Data Preprocessing

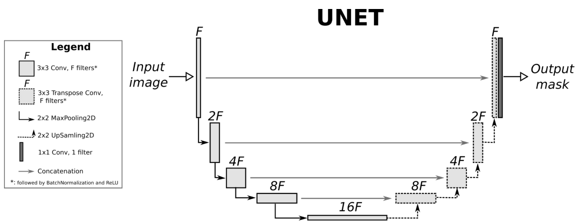
1. Extract metadata from scans (day, case, slice)
2. Converting RLE encoding to masks
3. Train (65%), Val (15%), Test (20%) split
4. Created COCO annotations
5. Developed DataLoader and DataGenerator modules

COCO
Common Objects in Context

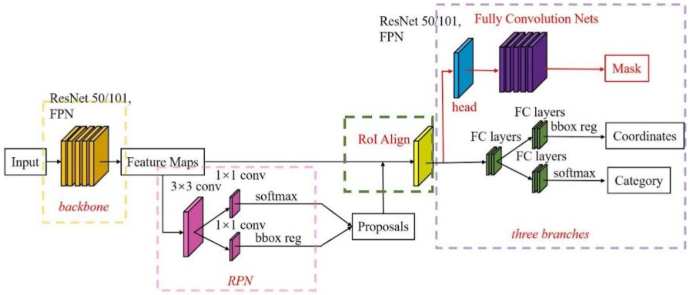

```json
{
    "categories": [
        {
            "id": 0,
            "name": "small_bowel"
        },
        {
            "id": 1,
            "name": "large_bowel"
        },
        {
            "id": 2,
            "name": "stomach"
        }
    ],
    "images": [
        {
            "id": 1,
            "file_name": "case123/case123_day20/scans/slice_0066_266_266_1.50_1.50.png",
            "width": 266,
            "height": 266
        },
        {
            "id": 2,
            "file_name": "case123/case123_day20/scans/slice_0067_266_266_1.50_1.50.png",
            "width": 266,
            "height": 266
        },
```

# Methodology

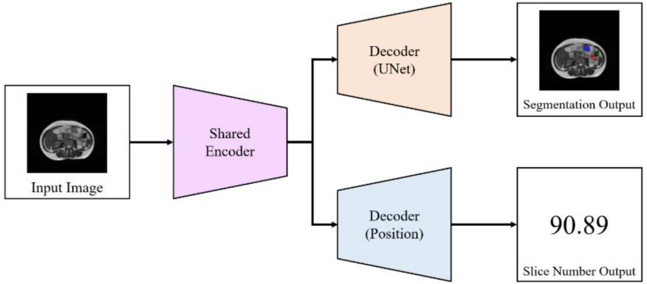
- **Simple U-Net (Baseline)**
  - This model serves as baseline, with a simple and symmetric encoder-decoder architecture
- **Efficient Net:**
  - Efficient Net is used as an encoder in the U-Net architecture. It leverages pretrained weights on imagenet dataset
- **DeepLabv3v+:**
  - Leveraging a powerful backbone ResNeXt101_32x8d for segmentation
- **Mask R-CNN:**
  - Integrate a R-CNN framework with an additional mask prediction branch
- **Multi-Task Learning U-Net:**
  - Simultaneous semantic segmentation and positional regression with shared encoder




U-Net Architecture

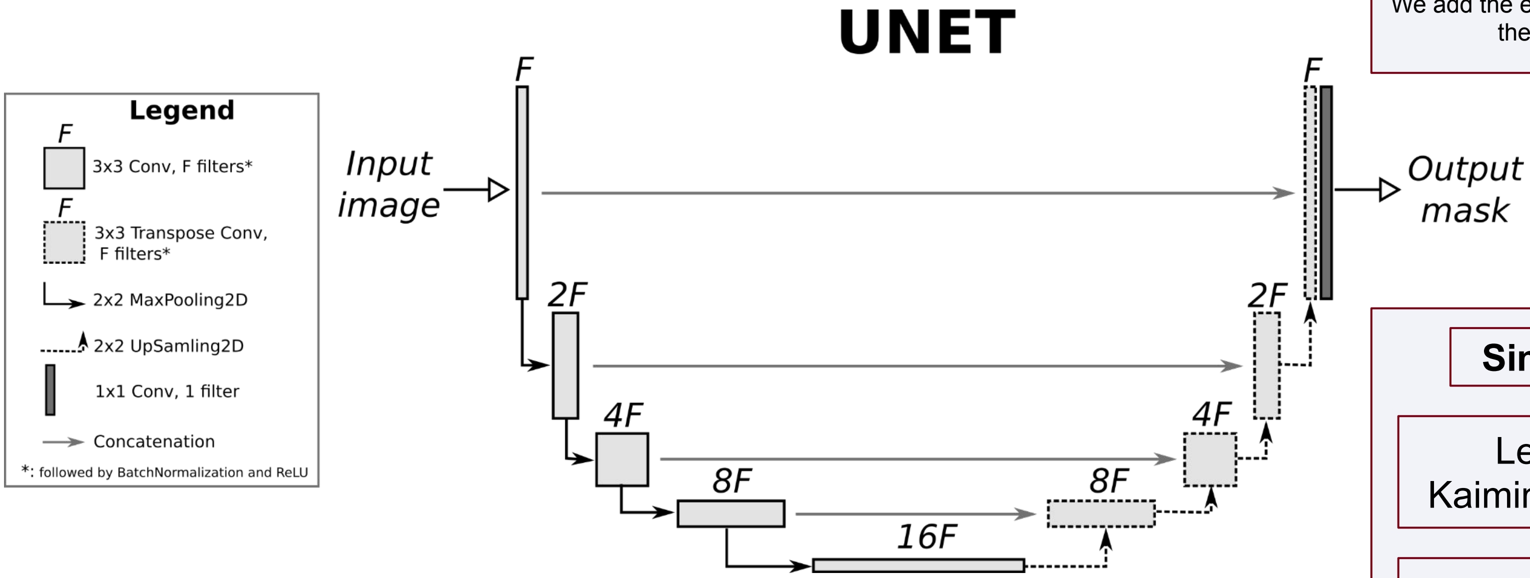



Mask R-CNN Architecture




Multi-Task Learning Architecture

# U-Net Type Models




**Efficient Net**

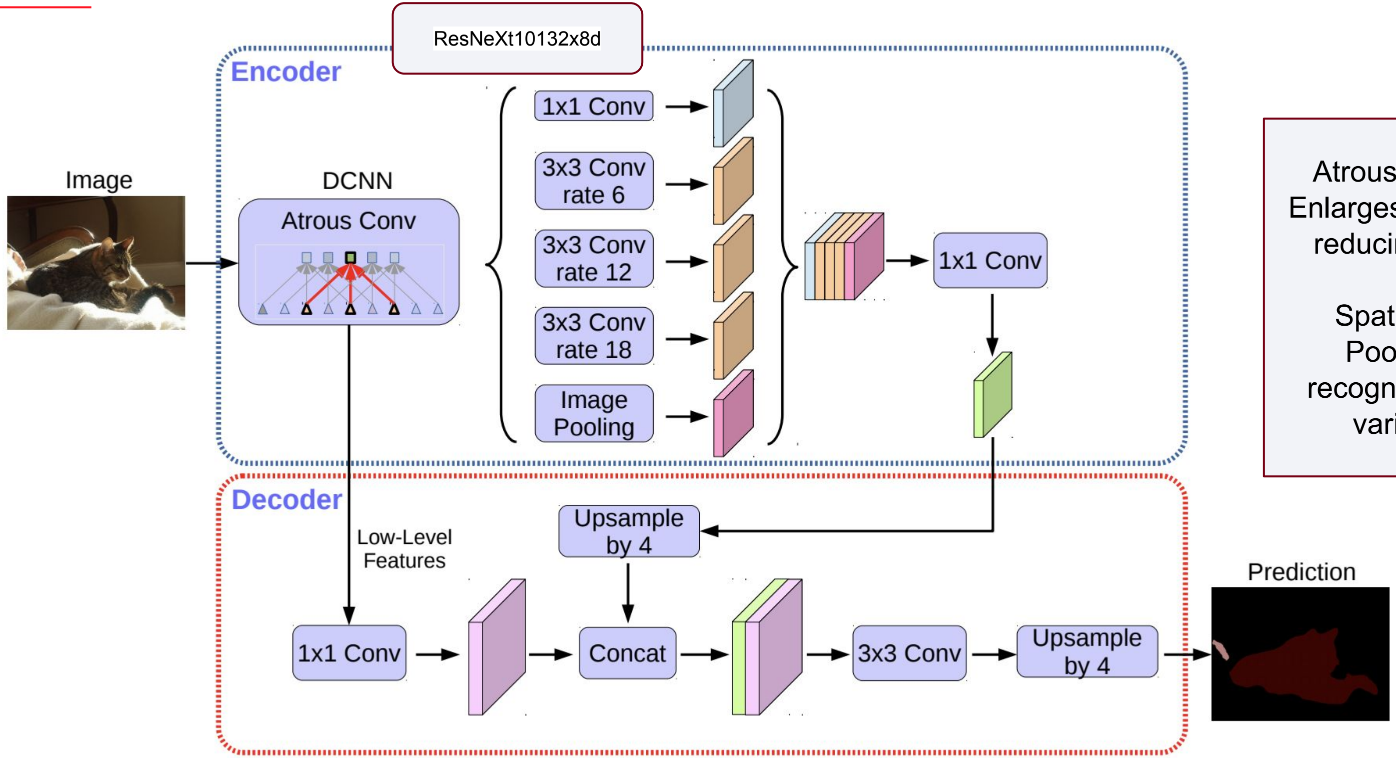
We add the efficient net architecture to the encoding layer

**Simple U-Net**

Leaky ReLU
Kaiming Initialization

Downsampling - 4 blocks
Upsampling - 4 blocks

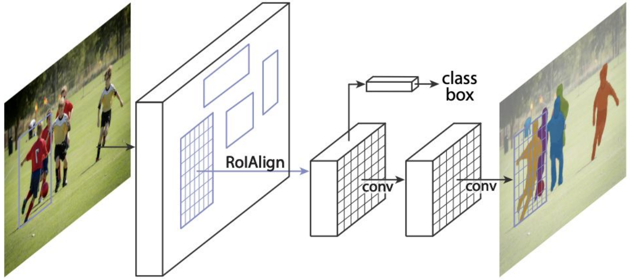
# DeepLabv3+



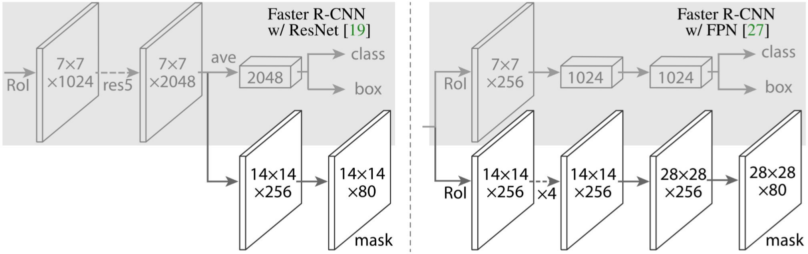

Atrous convolution: Enlarges field view w/o reducing resolution

Spatial Pyramid Pooling: helps recognize objects of various sizes

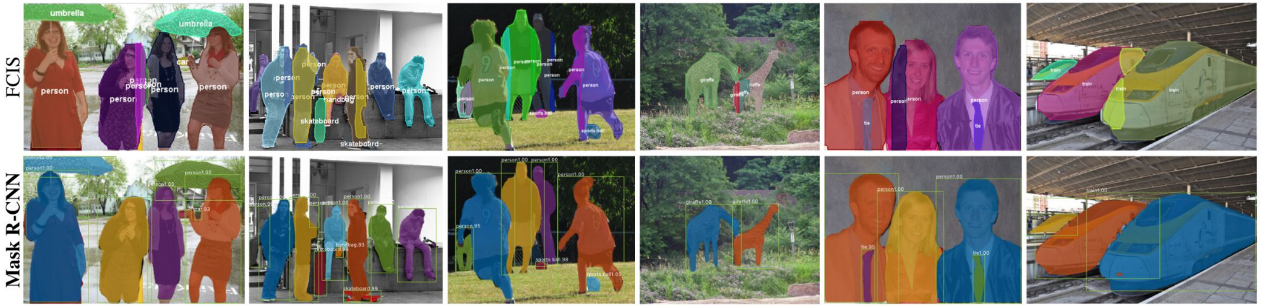
# Mask R-CNN




**Mask R-CNN framework for instance segmentation**

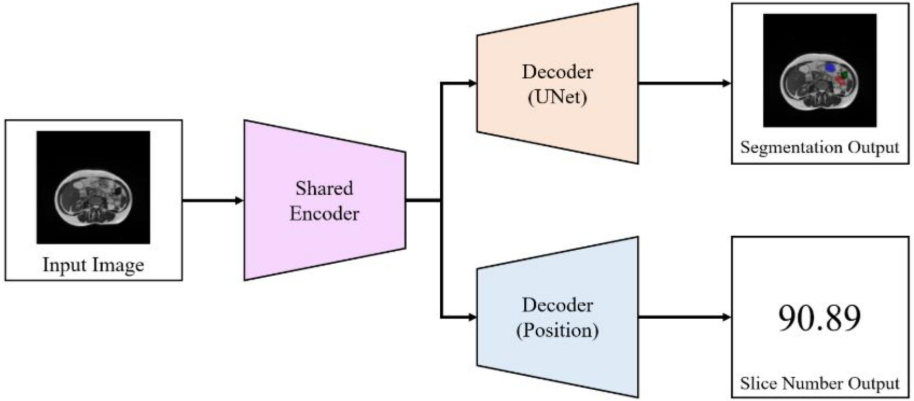


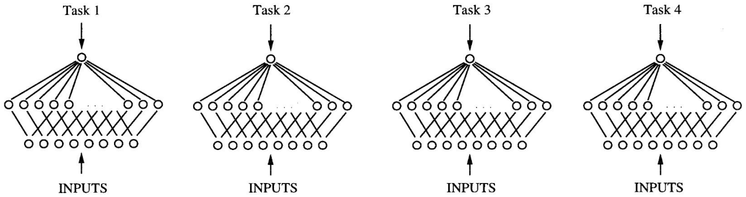

**Mask R-CNN Head Architecture**



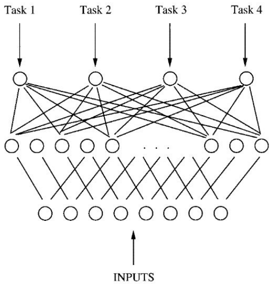

**Performance on Microsoft COCO dataset compared to FCIS model**

# Multi-Task Learning U-Net




**Single Task Backpropagation of four tasks with the same inputs**




**Multi-Task Backpropagation of four tasks with the same inputs**

# Evaluation Metrics

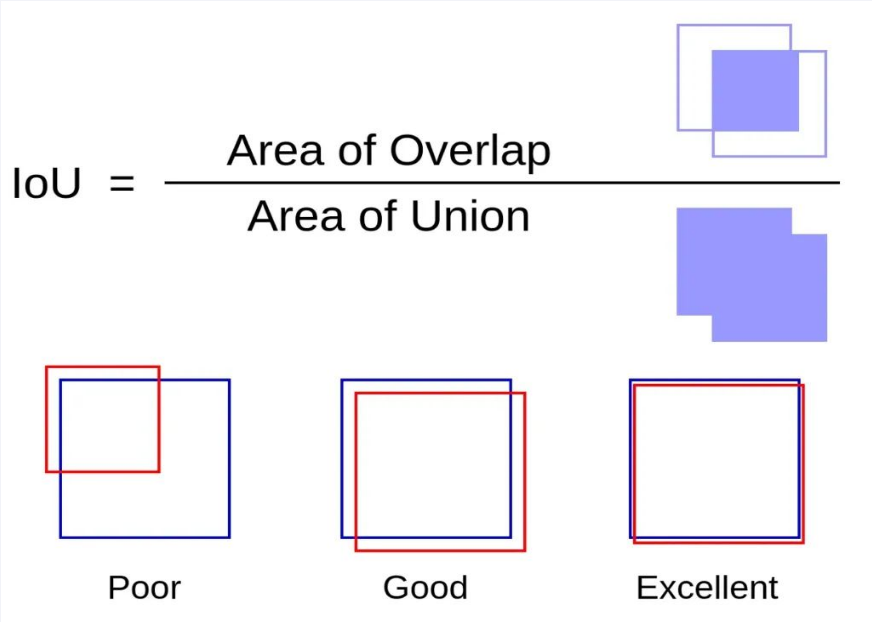
### Dice Loss

- Measures overlap between predicted and true segmentation masks
- Lower is better

$$C = \frac{2 \times |M_1 \cap M_2|}{|M_1| + |M_2|}$$

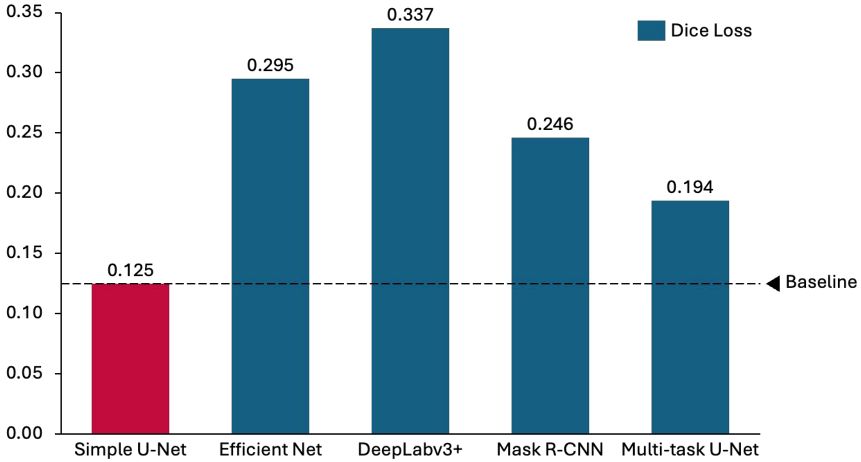
$$L_{\text{Dice}} = 1 - C$$

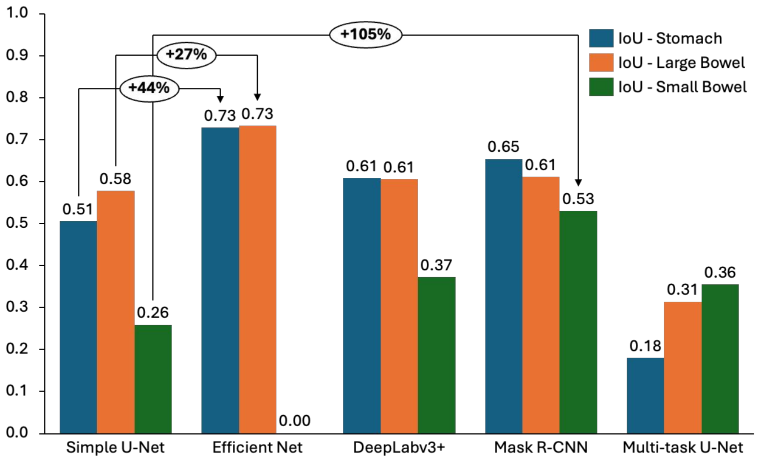
### Intersection over Union



# Model Performance

- Metrics are calculated based on test (20%) dataset
- Model training on Google Colab platform with A100 GPU




**Mask R-CNN** is overall the best model (high score and balanced in three organs)

# Result Visualization

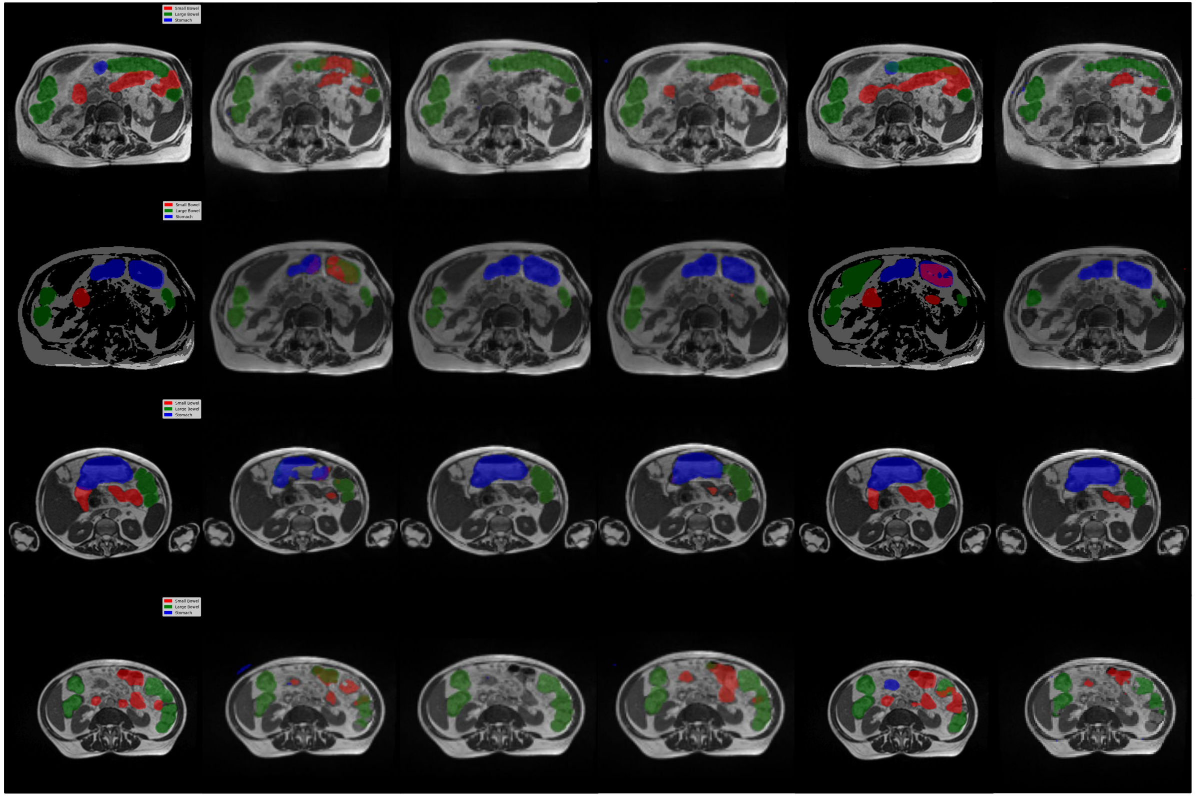
Ground Truth | Simple UNet | Efficient UNet | DeepLabv3+ | Mask R-CNN | Multi-task UNet

Train 1

Train 2

Test 1

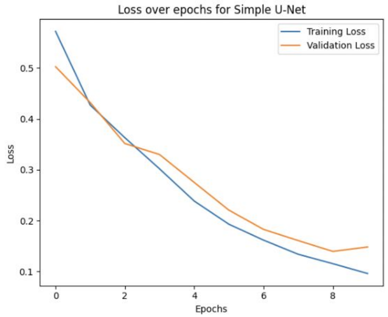
Test 2



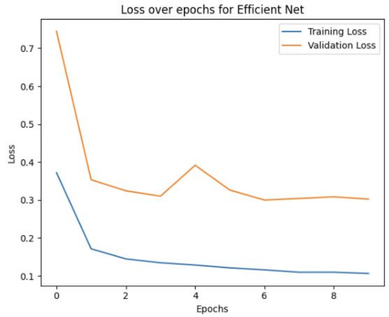
**Red: Small Bowel; Green: Large Bowel; Blue: Stomach**

# Error Analysis

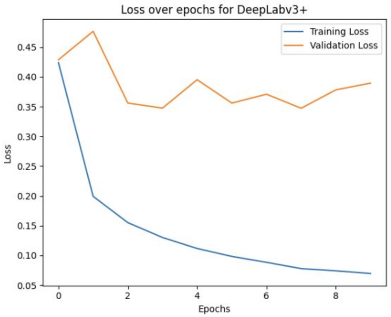
## Simple UNet



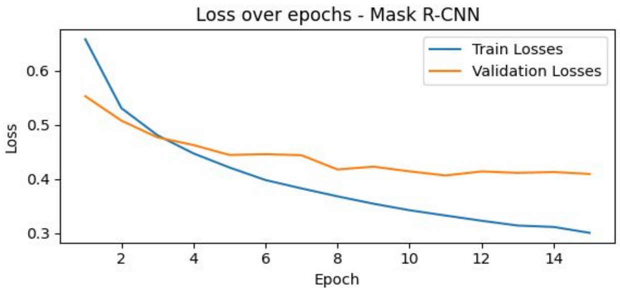

## Efficient Net



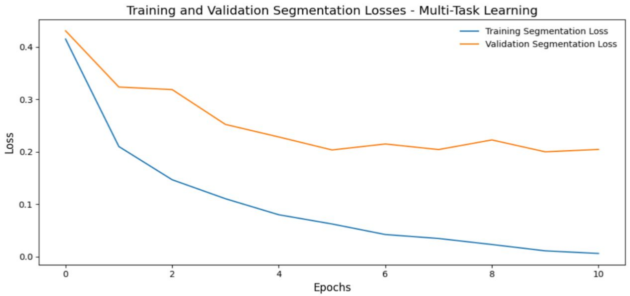

## DeepLabv3+




## Mask R-CNN



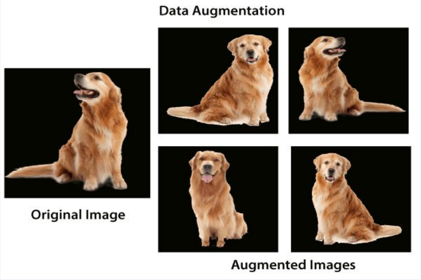

## Multi-task Learning U-Net

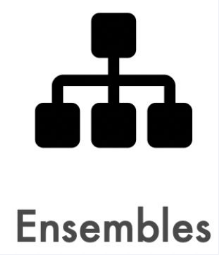


# Conclusion & Future Work

- Out of the 5 models, the simple baseline model performed well with respect to the dice loss metric
- While advanced models and multi-task learning did not consistently outperform the baseline, some showed notable improvements in specific metrics, leading to slight enhancements in test set performance for certain classes

## Future Directions




Improve the positional encoding

# References


1. Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation, 2017.
2. Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation, 2018.
3. Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation, 2014.
4. Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask r-cnn, 2018.
5. Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollar, and C. Lawrence Zitnick. Microsoft coco: Common objects in context.In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors,Computer Vision – ECCV 2014, pages 740–755,Cham, 2014. Springer International Publishing.
6. Jelena Novosel, Prashanth Viswanath, and Bruno Arsenali. Boosting semantic segmentation with multi-task self-supervised learning for autonomous driving applications. In Proc. of NeurIPS-Workshops, volume 3, 2019.
7. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.
8. Wenjia Bai and Dinggang Shen. Self-supervised learning for cardiac mr image segmentation by anatomical position prediction. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2019, Cham, 2019. Springer
9. Rich Caruana. Multitask learning. Machine Learning, 28:41–75, 1997.
10. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee, 2009.
11. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. Proceedings of the IEEE international conference on computer vision, pages 1026–1034, 2015.

Massachusetts Institute of Technology

# Any Questions? Email us!


**Srikaran Boya**
srikaran@mit.edu

**Joseph Kajon**
kajon145@mit.edu

**Zeki Yan**
zikaiyan@mit.edu

Massachusetts Institute of Technology

# Appendix

# Model Results on Metrics

Table 1. Model results

| Model | DiceLoss | IoU-S | IoU-LB | IoU-SB |
|---|---|---|---|---|
| Simple U-Net | **0.125** | 0.506 | 0.578 | 0.259 |
| Efficient Net | 0.295 | **0.728** | **0.733** | 0.000 |
| DeepLabv3+ | 0.337 | 0.609 | 0.606 | 0.372 |
| Mask R-CNN | 0.246 | 0.654 | 0.611 | **0.530** |
| Multi-task U-Net | 0.194 | 0.180 | 0.314 | 0.355 |