

# A Fixed-Point Model for Pancreas Segmentation in Abdominal CT Scan

A research work funded by the FELIX project

Yuyin Zhou<sup>1</sup>, Lingxi Xie<sup>1</sup>, Wei Shen<sup>1,2</sup>, Yan Wang<sup>1</sup>, Elliot K. Fishman<sup>3</sup>, Alan L. Yuille<sup>1</sup>

<sup>1,3</sup>The Johns Hopkins University / School of Medicine

<sup>2</sup>Shanghai University



Scan & Bookmark!

## ABSTRACT

Deep neural networks have been widely used for automatic organ segmentation from abdominal CT scans. However, the segmentation accuracy of some small organs (e.g., the pancreas) is sometimes below satisfaction, arguably because deep networks are easily disrupted by the complex and variable background regions which occupies a large fraction of the input volume. In this paper, we formulate this problem into a fixed-point model which uses a predicted segmentation mask to shrink the input region. This is motivated by the fact that a smaller input region often leads to more accurate segmentation. In the training process, we use the ground-truth annotation to generate accurate input regions and optimize network weights. On the testing stage, we fix the network parameters and update the segmentation results in an iterative manner. We evaluate our approach on the NIH pancreas segmentation dataset, and outperform the state-of-the-art by more than 4%, measured by the average Dice-Sørensen Coefficient (DSC). In addition, we get 62.43% DSC in the *most challenging* case, which guarantees the reliability of our approach in clinical applications.

## CONTRIBUTION

We propose a novel framework for organ segmentation in CT and other data formats. Note that medical imaging data are often 3D volumes, which are different from those 2D images. Our technical contributions can be summarized as:

1. We suggest to use 2D segmentation models to process 3D data. This alleviates the issues of data and memory. 3D information is integrated via multi-slice segmentation and fusing prediction from different viewpoints.
2. We design a coarse-to-fine framework for single-viewpoint segmentation. This model works well for small targets such as the pancreas or the pancreatic cyst.

Our framework obtains the state-of-the-art segmentation accuracy (82.37%) on the NIH pancreas segmentation dataset.

[Our codes are available online!](#)

## REFERENCES

Key references are numbered as in the paper.  
[6] J. Long *et al.*, Fully Convolutional Networks for Semantic Segmentation, *CVPR*, 2015.  
[9] H. Roth *et al.*, DeepOrgan: Multi-level Deep Convolutional Networks for Automated Pancreas Segmentation, *MICCAI*, 2015.

## ACKN.

This work was supported by the Lustgarten Foundation for Pancreatic Cancer Research and NSFC No. 61672336. We thank Dr. Seyoun Park and Zhuotun Zhu for their enormous help, and Weichao Qiu, Cihang Xie, Chenxi Liu, Siyuan Qiao and Zhishuai Zhang for discussions.

## THE PROPOSED APPROACH

### The FELIX Project

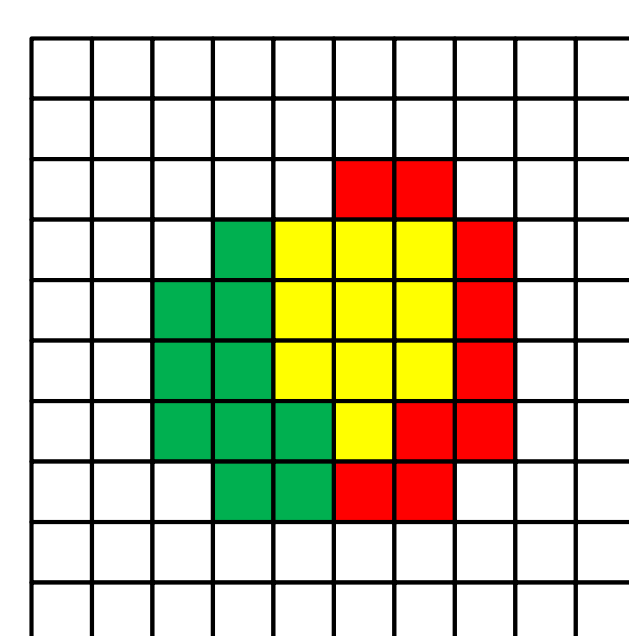
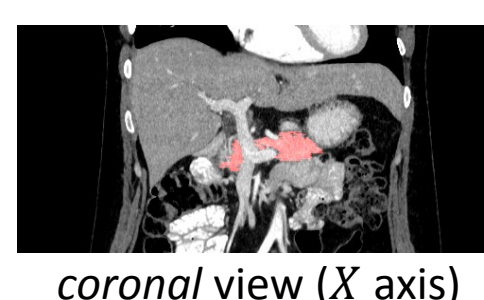
The pancreatic cancer is a major killer to humans. As the symptom is very difficult to detect at an early stage, the cancer has often spread to other organs at the time of diagnosis, causing a very high death rate (5-year survival rate is merely 7.7%).

The FELIX project is aimed at applying modern approaches in computer vision to assist doctors in diagnosing the pancreatic cancer at an early stage. We start from the CT scanned images. This is a challenging task, as the pancreas is a small organ with irregular shape and a blurring boundary. In this work, we present an approach for accurate pancreas segmentation.

### Data and Evaluation

Data: the NIH pancreas segmentation dataset [9]

- 82 CT samples collected from **healthy** people
- Moderate resolution along the axial view
- Evaluation: the DSC score



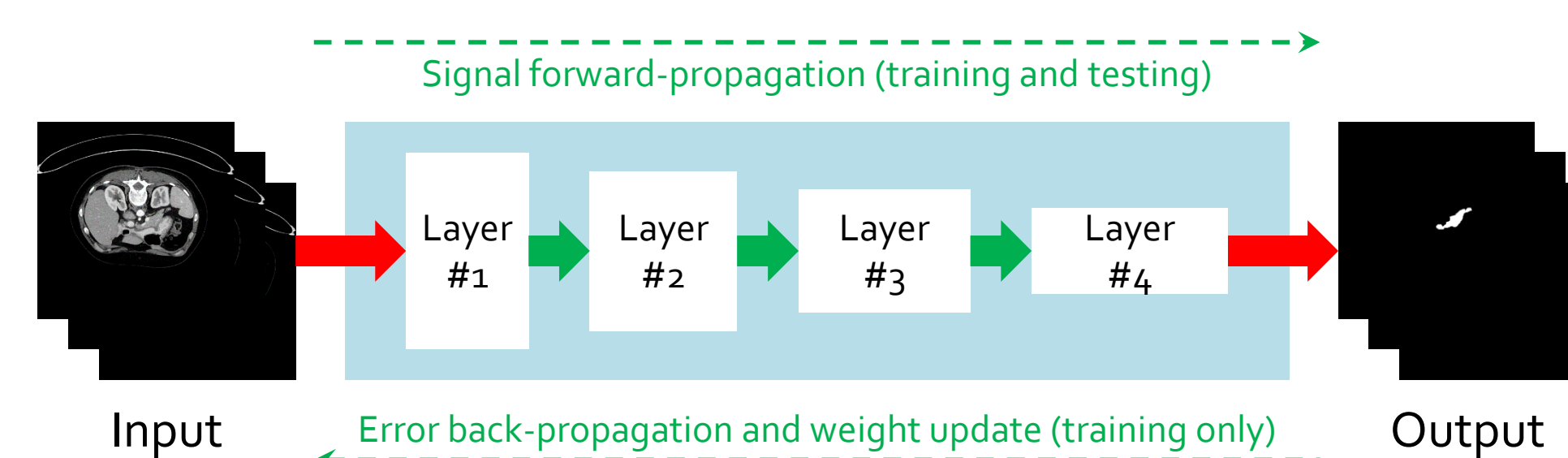
Ground-truth (19 voxels)  
Prediction (20 voxels)  
Overlap (10 voxels)

$$DSC = \frac{2 \times 10}{19 + 20} = 51.28\%$$

Note: evaluation in 3D volumes!

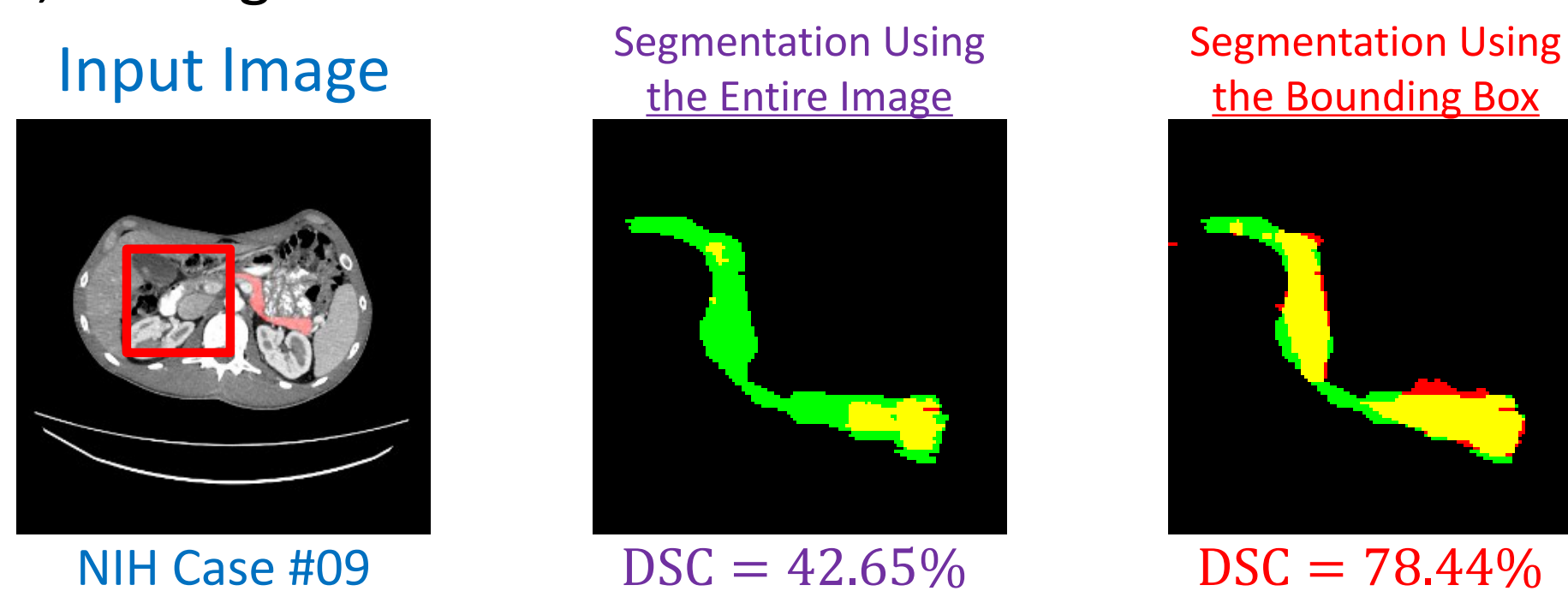
### Deep Learning Basics

Deep learning is the state-of-the-art solution for a wide range of image-based applications. It is based on the idea that a deep network can capture very complicated distribution in image space. The basic unit of a deep network is a **neuron**, i.e., a mathematical function for a specified purpose. Neurons with the same function form a **layer**, and a deep network is a *hierarchical* structure with many layers. Training a deep network involves propagating neural responses back and forth and updating network weights.



### Motivation & Approach

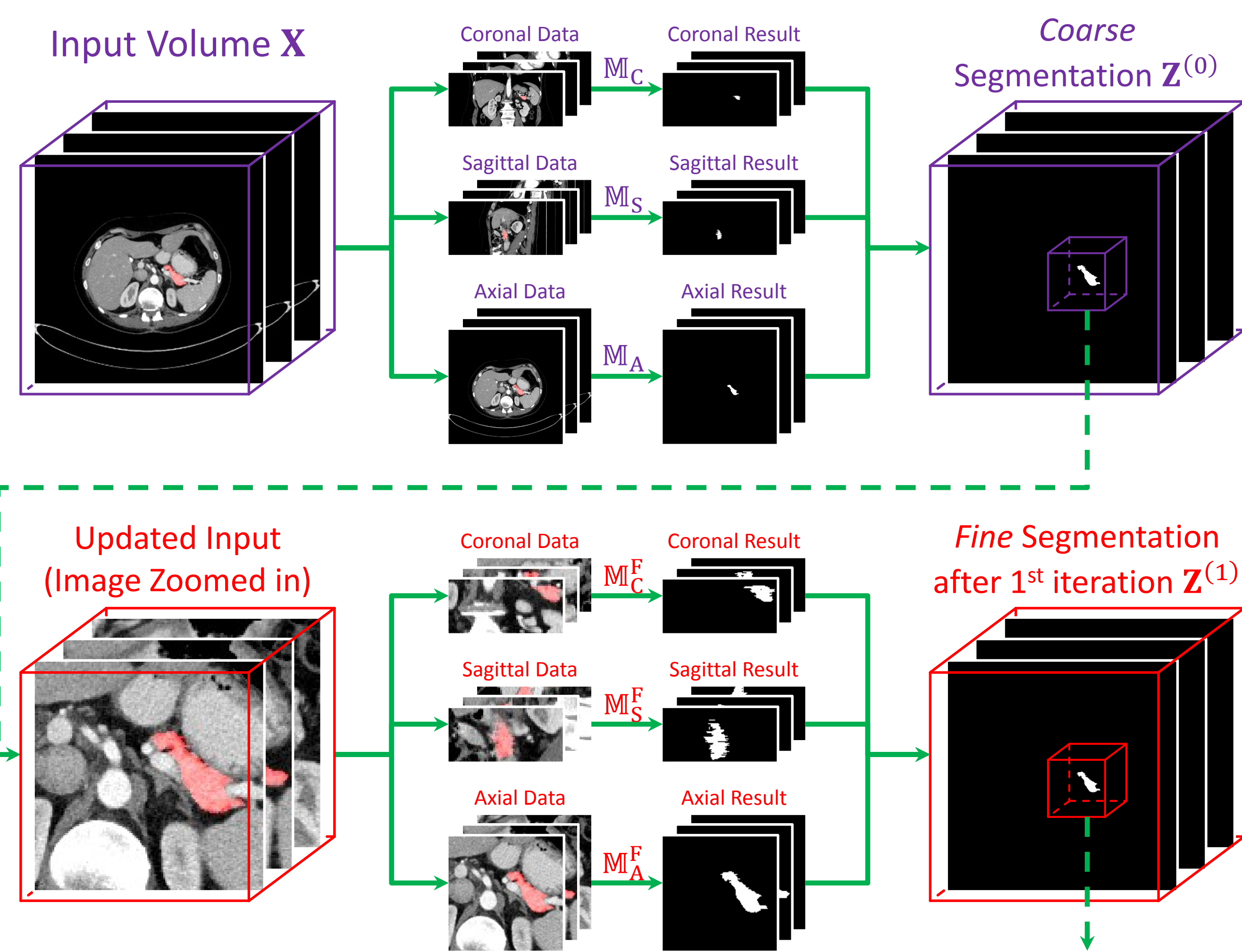
Deep networks for segmentation (e.g., FCN [6]) are often less accurate on small targets such as the pancreas, arguably because the background region contains random noise which distracts the network. However, if we focus on a smaller region around the target, the segmentation becomes much more accurate.



Our idea is to shrink the input region according to the rough segmentation results. Thus, this problem can be formulated as an optimization problem in which the segmentation mask appears in both input and output. This is a **fixed-point model** and can be solved via a **coarse-to-fine** iteration.

In the **training** stage, we train two sets of networks for coarse and fine segmentation, respectively. In the **testing** stage, we perform iteration (see the right figure) until convergence.

### Input Volume X



## EXPERIMENTAL RESULTS

### Quantitative Results

We evaluate our approach on the NIH pancreas segmentation dataset, and report average accuracy and standard deviation over 82 cases.

We also compare either fixed or flexible conditions for the iteration to terminate. In overall, a large fraction of improvement comes from the first iteration, and iteration improves the worst case significantly.

	Mean DSC	# Iterations	Max DSC	Min DSC
Coarse Segmentation	75.74 ± 10.47	—	88.12	39.99
After 1 Iteration	82.16 ± 6.29	1	90.85	54.39
After 2 Iterations	82.13 ± 6.30	2	90.77	57.05
After 5 Iterations	82.11 ± 6.09	5	90.75	62.40
After 10 Iterations	82.25 ± 5.73	10	90.76	61.73
After $d^{(t)} > 0.90$	82.13 ± 6.35	1.83 ± 0.47	90.85	54.39
After $d^{(t)} > 0.95$	<b>82.37 ± 5.68</b>	2.89 ± 1.75	90.85	<b>62.43</b>
After $d^{(t)} > 0.98$	82.28 ± 5.71	7.35 ± 2.88	90.78	61.94
After $d^{(t)} > 0.99$	82.28 ± 5.72	9.87 ± 0.73	90.77	61.94
Best of All Iterations	82.65 ± 5.47	3.49 ± 2.92	90.85	63.02
Oracle Bounding Box	83.18 ± 4.81	—	91.03	65.10

We also show that the reported average accuracy is competitive among the state-of-the-arts.

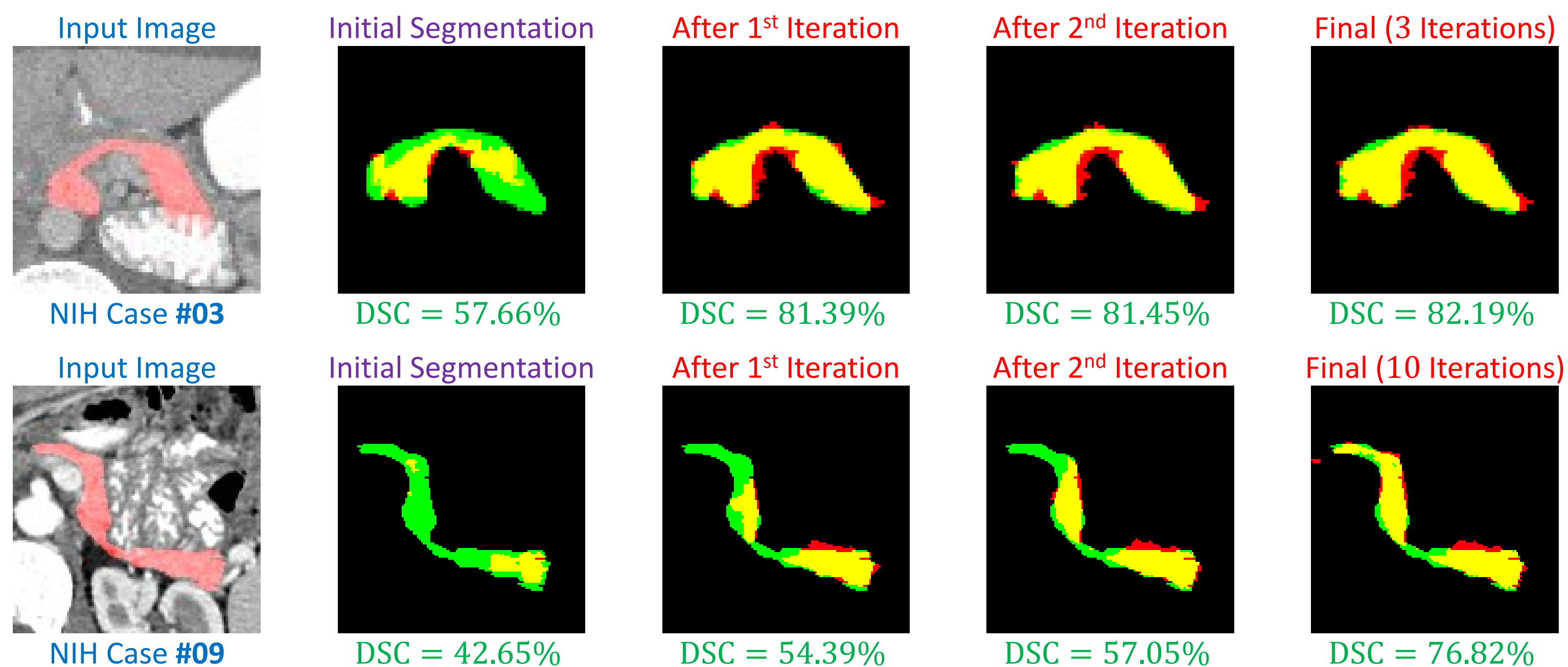
	Mean DSC	Max DSC	Min DSC
Coarse-scaled Segmentation	75.74 ± 10.47	88.12	39.99
Fine-scaled Segmentation	<b>82.37 ± 5.68</b>	<b>90.85</b>	<b>62.43</b>
Fine-scaled Segmentation + stat-fusion *	<b>84.06 ± 4.16</b>	<b>91.46</b>	<b>71.76</b>
Roth <i>et al.</i> , MICCAI'15	71.42 ± 10.11	86.29	23.99
Roth <i>et al.</i> , MICCAI'16	78.01 ± 8.20	88.65	34.11
Roth <i>et al.</i> , arXiv'17 *	81.27 ± 6.27	88.96	50.69
Cai <i>et al.</i> , MICCAI'17 *	82.4 ± 6.7	90.1	60.0

\* indicates a result which is available after this paper was submitted to MICCAI'17.

**Stat-fusion** is an advanced algorithm to combine the prediction results from *coronal*, *sagittal* and *axial* views. We used majority voting throughout this work.

### Visualization

We visualize segmentation with respect to iteration # to show the improvement of our approach.



### Convergence

In general, our approach converges very well.

- The average inter-iteration DSC is **0.8338** after the 1<sup>st</sup> iteration, and **0.9767** after the 10<sup>th</sup> iteration.
- Under the terminating condition of  $d^{(t)} > 0.95$ , the testing processes of 73 out of 82 cases end in 2 or 3 rounds, and only 3 of them require 10 or more rounds.

### Conclusions & Discussions

In this work, we propose a fixed-point model for coarse-to-fine pancreas segmentation from abdominal CT scan. We verify that shrinking the input region via rough prediction helps a lot in small organ segmentation. Based on 2D networks to deal with 3D image data, we achieve the state-of-the-art performance on a public dataset. Our approach generalizes well onto other small organs such as the *duodenum*.

However, there are still open problems to explore.

- ✓ Is the coarse-to-fine idea only useful in medical images, or can it be generalized to other natural images?
- ✓ Is it possible to design an end-to-end framework for the coarse-to-fine approach?