**Cascaded 3D U-Net** — Multi-Organ Segmentation (Biomedical Image Segmentation)

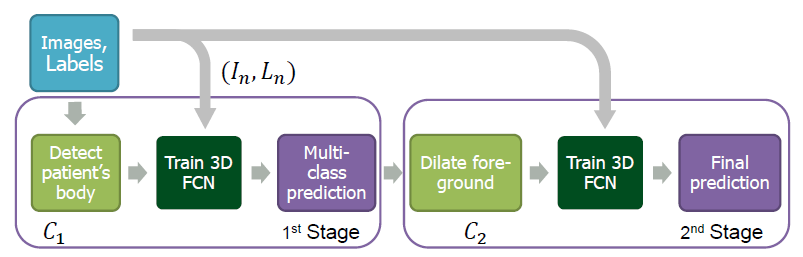
Two stages

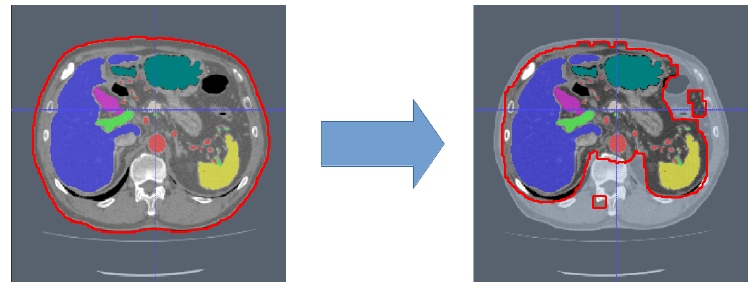
* A FCN Model to identify the organs of interest.
* FCN model is trained to have a detailed segmentation of the organs

# Outline

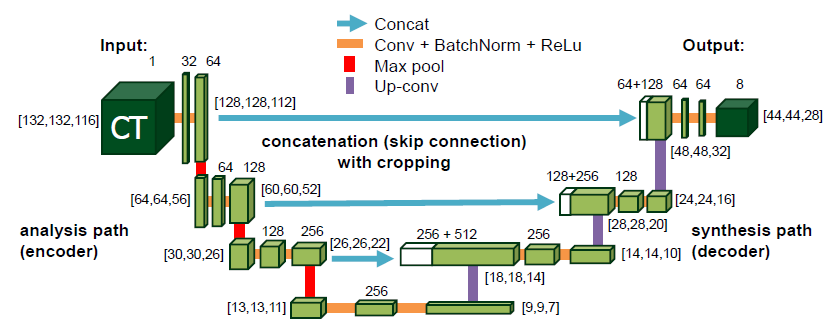
1. **Coarse-to-Fine Cascaded U-Net**
2. **Loss Function**
3. **Training & Validation & Testing**
4. **Comparison with State-of-the-art Approaches**

# ****Coarse-to-Fine Cascaded U-Net****

**Multi-stage cascaded training scheme**

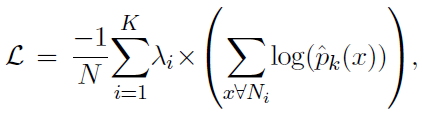
**C1 Candidate Region** (Red) input to Stage 1 3D U-Net (Left), **C2 Candidate Region** (Red) input to Stage 2 3D U-Net (Right)

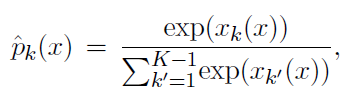
* The first-stage FCN sees around 40% of the voxels as **C1 Candidate Region**using only a simple mask of the body created by thresholding the image.
* **C2 Candidate Region is output from the first-stage FCN.**
* In second-stage FCN, the amount of the image’s voxels is reduced by around 10%.
* This step narrows down and simplifies the search space for the FCN.

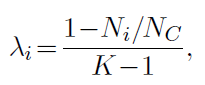
**3D U-Net**

* In this paper, the 2-stage FCNs used are 3D U-Net.
* (For details about 3D U-Net, Please feel free to read my review about it.)
* At the end of the network, the last layer contains a 1×1×1 convolution that reduces the number of output channels to the number of class labels (K=8, 7 organs plus 1 background) and a size of 44×44×28 of each channel.

# 2. Loss Function





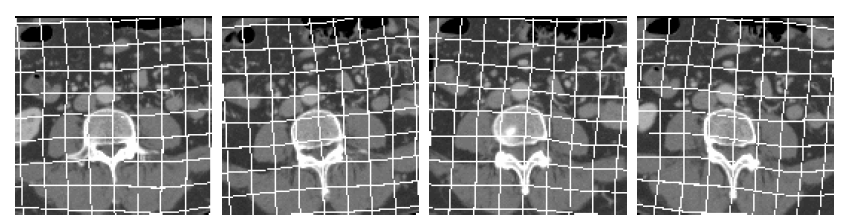


* Ni are the number of voxels within one class in Ln.
* NC is the number of voxels within candidate region C1 or C2.
* λi is the value based on Ni and NC, i.e. based on the occurrence frequency of the class. And summing all λiequals to 1.
* Simply speaking, the loss is the **weighted voxel-wise cross-entropy loss**.

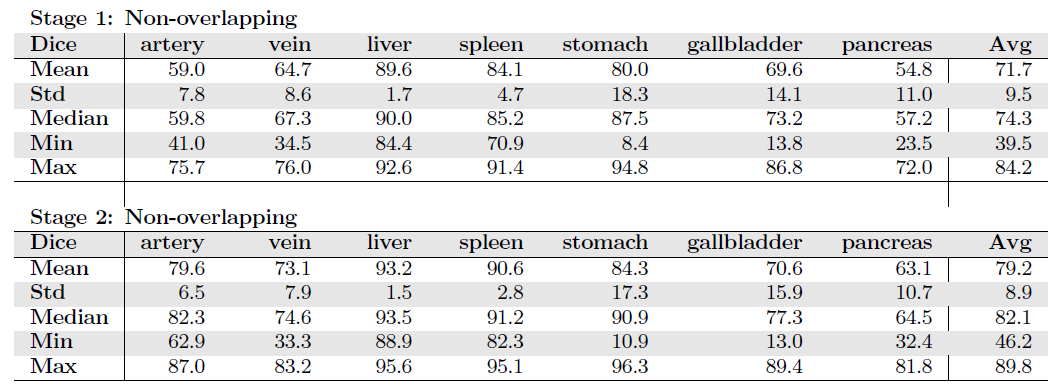
# 3. ****Training & Validation & Testing****

## 3.1 Training & Validation

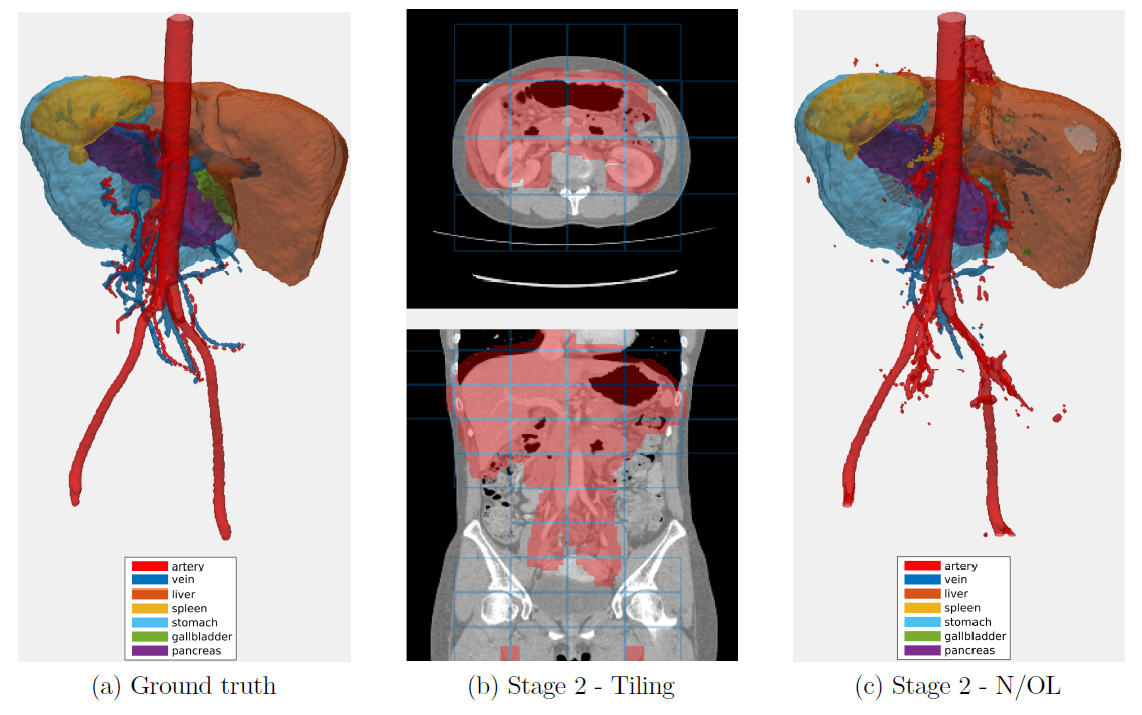
* **Dataset**: 331 contrast-enhanced abdominal clinical CT images.
* Each CT volume consists of 460 — 1177 slices of 512×512 pixels. The voxel dimensions are [0.59–0.98, 0.59–0.98, 0.5–1.0] mm.
* A random split of 281/50 patients is used for training and validating the network.

**Various examples of plausible random deformation**

* **Data Augmentation**: Smooth B-spline deformations, random rotations.
* 200,000 iterations in the first stage and 115,000 in the second.

**Dics Similiary Score on Validation Set**

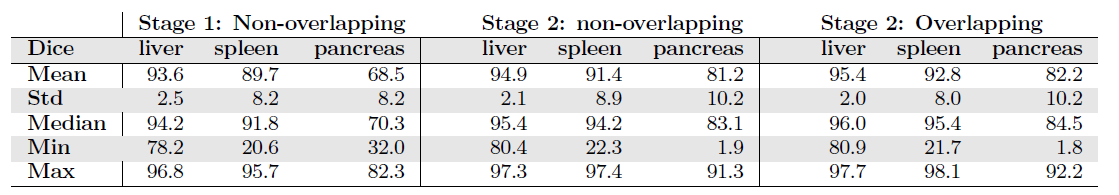
* With 2 stages, dice score is improved from 71.7% to 79.2%.
* On average, a 7.5% improvement in Dice scores per organ is achieved.

**Example of the validation set**

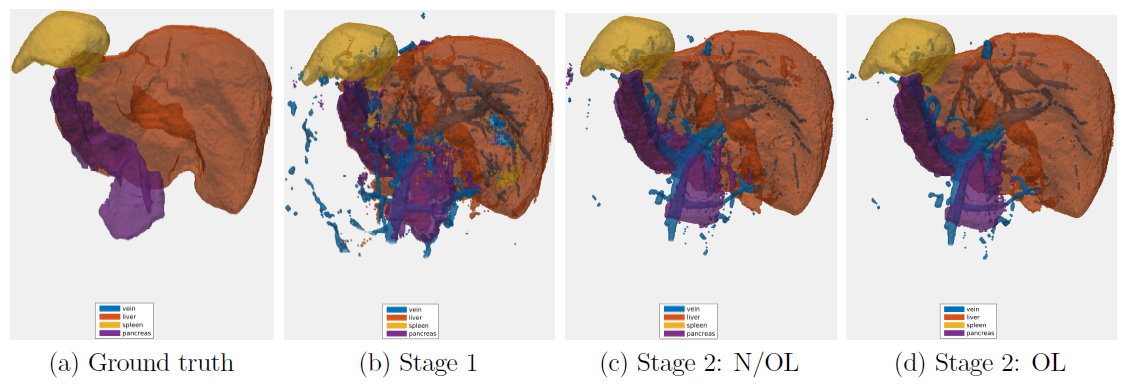
* The segmentation is quite good as shown above.

## 3.2 Testing

* **Dataset**: It originates from a different hospital, scanner, and research study with gastric cancer patients. 150 abdominal CT scans were acquired.
* Each CT volume consists of 263 — 1061 slices of 512×512 pixels. Voxel dimensions are [0.55–0.82, 0.55–0.82, 0.4–0.80] mm.
* The pancreas, liver, and spleen were semi-automatically delineated by three trained researchers and confirmed by a clinician.

**Dics Similiary Score on Test Set**

* The dice similarity score is improved at stage 2.
* With overlapping, the score is even higher.

**Example of the test set**