

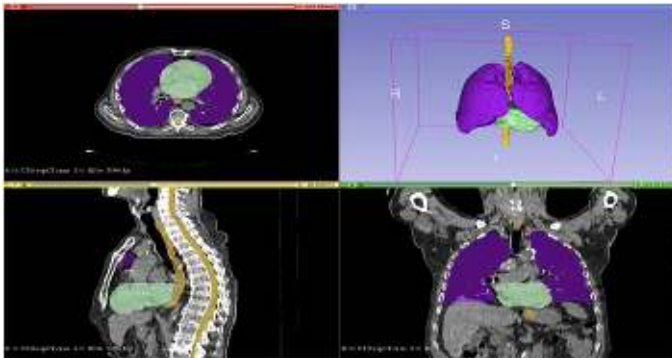
## Abstract

We are delivering the proof of concept for a tool, capable of automatic segmentation of organs in 3D medical images obtained by CT, using methods of computer vision and neural networks. Our proposed methods heavily relies on 3D Unet [1] with customized loss function using a combination of Dice-Sorensen score and Distance transform maps. Evaluation of our methods shows statistically significantly better results on almost all organs, then best performing model on the same dataset.

## Introduction

- Organ segmentation done on a daily basis by radiologists, mostly relying on interactive or semi-automatic methods.
- Required for planning radiotherapies, surgery planning, applying specific algorithms for searching pathological formations in tissues, etc.
- Fully automated method can rapidly reduce time and cognitive load required from experts, and speedup segmentation, allowing more people to be served.

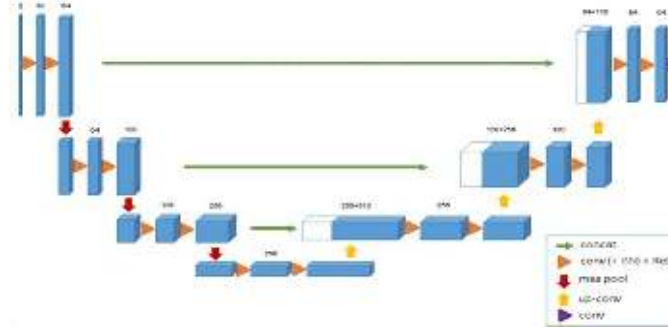
Image 1: Segmented organs in thorax, the expected output from proposed method



## Input Data Preprocessing and Augmentations

- Input images can dramatically vary across scanners both in voxel intensities for same scanned tissue and in resolution.
- Voxel intensities are normalized to *Hounsfield units(HU)*, where every scanner needs to be calibrated for both water and air. Transformation to HU is that as easy as applying the formula:
 
$$HU = 1000 \times \frac{\mu - \mu_{air}}{\mu_{water} - \mu_{air}}$$
- Since each organ has specially trained network for it, and uses 2-stage crude and then fine segmentation approach, using simple rescaling with interpolation for required resolution of 240x240x120 voxel per input image works fine.
- Annotated CT scans with multiple annotated data is hard to come by, so to have more data to train on, we are using data augmentations of 3 types:
  - Rotation of up to 10°
  - Introduction of physiological noise
  - Zoom

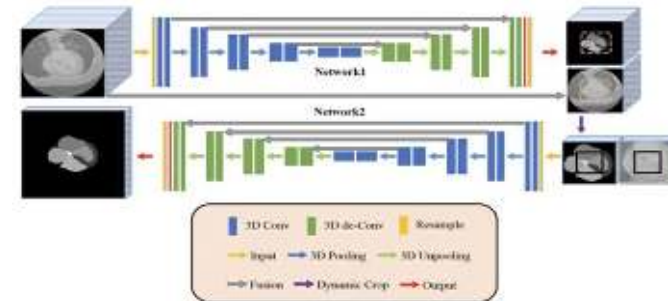
Image 2: 3D Unet architecture schematic [2]. Unet is type of Convolutional Neural Network, that shares features between compression and expansion paths, avoiding compression losses when upconvoluting.



## Method – 2 stage 3D Unet

- Segmentation is done in 2 steps:
  - Coarse segmentation with first Unet as region proposal
  - Fine segmentation with second Unet as refinement
- Each organ has 2 network experts specially trained just for it.
- Resulting multiorgan segmentation is a merge of all segmentations.
- First network works with resolution of 240x240x120 voxels, sometimes drastically reducing image quality, in exchange for having better context, using self-observed anatomical relations for segmentations.
- The network is motivated to pursue and learn anatomical relations with our custom build loss function (see section Loss function) based on dice score augmented with the gravity of mistakes
- Resulting segmentation is used to create a bounding box (with slight padding) and crop Region of Interest (ROI) from input data.
- The 2nd stage can work on crop of much higher resolution, preserving more details in the input. On another channel segmentation the from previous stage is also passed to the network.
- Since only ROI is passed in, anatomical relations are now irrelevant, but more important are the boundaries of the organ. So the pure Dice score is now enough as loss function.

Image 3: The 2-stage 3D Unet with the first stage serving for cropping the ROI from the input data [3]



## Loss function - Dice score augmented by distance

- Dice score tells the network how precise it is, however, if it makes a mistake, Dice score tells nothing about gravity of the mistake.
- False positive match 1 voxel from real border is labeled with the same gravity as falsely labeling liver as heart.
- Dice score is calculated as:  $DSC = \frac{2TP}{2TP + FP + FN}$
- To evaluate the gravity of mislabeling we prepared distance transform maps (DTM) from ground truth segmentations.
- Further away from ground truth, the higher values are in DTM (see Image 4)
- To get the gravity of the mistake just make Hadamard's product of prediction and DTM:  $\text{sum}(\text{prediction} \odot \text{dtm})$

Image 4: Distance transform map for right lung used for augmenting loss.



## Results

- We compared both 1-stage and 2-stage 3D-Unet with the best performing model results on the same dataset, provided via AAPM Thoracic Autosegmentation Challenge.
- Since the challenge was over before submission could be made, and test data remained secret, we couldn't perform pair statistical tests, but we had to settle for unpaired testing against the self-made test dataset.
- For each organ we performed either two sample t-test, if achieved Dice score passed both the normality and variance test, Welch two sample t-test if they didn't pass variance F-test, and Wilcoxon Rank-sum test if even Shapiro-Wilk normality test didn't pass. This switch logic was used for comparing best with 1-stage and best with 2-stage.
- Statistical testing between 1-stage and 2-stage was performed by paired testing if variance and normality conditions were met.
- 2-stage 3D Unet shows statistically significantly better Dice score results in 4 out of 5 tested organs, and with a spine being equally good challenge winner.

Table 1: Comparison of reached Dice score between best model so far on same dataset, 1-stage and 2-stage 3D Unet

Average Dice score	Right lung	Esophagus	Left lung	Heart	Spine
Best model on dataset	71.99	28.19	76.29	45.97	61.16
1-stage	82.27	35.66	71.67	80.69	69.69
2-stage	94.26	46.29	91.02	84.79	66.73
Significantly better than best/1stage	YES/YES	YES/NO	YES/YES	YES/YES	NO/NO

## Contact

Michal Kováčik  
Faculty of Informatics and Information Technologies STU  
Email: xkovacikm2@stuba.sk  
Website: vgg.fiit.stuba.sk

## References

- O. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation"
- O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation"
- C. Wang, T. MacGillivray, G. Macnaught, G. Yang, and D. Newby, "A two-stage 3D Unet framework for multi-class segmentation on full resolution image"