Paper: Interactive Medical Image Segmentation Using Deep Learning

Summary

In this paper, the authors present a novel framework based on deep learning to deal with the problem of interactive segmentation. They use CNNs into bounding box and scribble-based segmentation pipeline. They further introduce image-specific fine-tuning which can either be supervised or unsupervised. One of their major goals is also to reduce the number of user interaction and user time than traditional interactive segmentation methods. They perform experiments on both 2-D segmentation and 3-D segmentation of multiple organs.

Due to the inherent challenge of medical images such as poor image quality and different imaging and segmentation protocols, the CNN based frameworks such as automated segmentation methods have not been remarkable. The alternative to this, which is interactive segmentation, performs a little better by integrating user's knowledge and considering application requirements. Such an interactive method should require a short user time as possible. Therefore, this is what initially motivated the authors to investigate combining CNNs with user interaction for medical image segmentation to achieve higher segmentation accuracy and fewer user interaction. The paper then aspires to tackle the three main challenges with using CNN for such problem domain: generalize to the previously unseen object class, adaptive to different test images, and fast inference and memory efficiency. To generalize to previously unseen objects, they propose a bounding-box based segmentation pipeline; to make CNN adaptive to different test images, they use image-specific fine-tuning; to tackle with speed and memory efficiently, they consider a balance among the receptive field, inference time, and memory efficiency. They call their network with this feature as BIFSeg.

For the 2D images, they adopted P-Net for bounding box-based binary segmentation. For the 3D images, they used an extension of P-Net and called it PC-Net. They used a grid search to find the important parameters in their network. To investigate the performance of their networks with the same bounding box, they compared P-Net with FCN and U-Net for 2D images and compared PC-Net with DeepMedic and HighRes3DNet for 3D images. All these frameworks are considered state-of-art in the field. BIFSeg was also compared with other interactive methods such as GrabCut and Slic-Seg for 2D segmentation and GeoS and GrowCut for 3D segmentation. The dice score between segmentation and the ground truth was used for quantitative evaluation.

As a result of their initial 2D segmentation based on P-Net, they found out that P-Net performs noticeably better than FCN and U-Net when dealing with unseen objects. For unsupervised refinement, the initial segmentation obtained by P-Net was refined by CRF, BIFSeg(-w) and BIFSeg without additional scribbles. They found out that BIFSeg achieves a larger improvement of accuracy from the initial segmentation. In the case of Supervised refinement with additional scribbles, BIFSeg achieves significantly better accuracy for the placenta, and previously unseen fetal lungs and maternal kidneys compared with P-Net + CRF and BIFSeg(-w). They also found out that supervised fine-tuning achieves 3-5 percentage points higher Dice than unsupervised fine-tuning. They found a similar result for 3D segmentation as well, where BIFSeg performed better than CRF and BIFSeg(-w) in all scenarios. In comparison with other interactive methods, they found out that their model achieves similar final Dice scores with significantly less user time.

Strengths

- The paper has a very good visual representation of the results along with a good explanation.
- The paper also presents their result utilizing the statistical tools such as p-value and probabilities which gives the user an idea about the statistical significance of the result.
- The paper is divided into a reasonable number of sections with a good description. There are also nice transitions between the sections.

Weaknesses

- Even though the gist of their architecture was fairly easy to follow; it was extremely hard to follow the equations that they included in the paper.
- The paper has some repeated information in between the sections. For example, the Contribution section and the end of the Introduction; similarly, the Discussion and Conclusion could have been made concise, since most of the information has already been elaborated in the previous sections.

Confusions

- What are the dilation parameters that they include in their CNN Models architecture?
- What is the isotropic vs. anisotropic receptive field? How does the anisotropic receptive field help reduce memory consumption compared to the isotropic receptive field?

Discussions

- What is a network-based uncertainty vs. scribbles-based uncertainty? How are they handled?
- What is the purpose of the concatenation layer in their model?
- Discuss Unsupervised vs. Supervised Image-Specific Fine-Tuning?