Review of 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation

Anonymous

08-02-2024

Reference

Özgün Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:2164893

Summary

This paper introduces a 3D extension of the 2D U-Net, replacing its 2D neural network components with their 3D counterparts while maintaining the original U-Net's shape and structure. The authors claim that the novelty lies in their utilization of deep networks for 3D segmentation tasks. They evaluate their model through two experiments: one with fully segmented expert volume annotations, and another with only sparse expert volume annotations of the Xenopus kidney. These experiments demonstrate both qualitative and quantitative success in their task, outperforming the 2D U-Net which in this case would process the volume slice by slice. The semi-automated results offer a promising approach to reduce the annotation-burden for experts.

Reasons to accept the paper

- The work is very novel as it is the first method proposed that uses deep volumetric neural networks for this segmentation task. They take into account recent advances in deep learning by integrating batch normalization to help stabilize training.
- The work is applied to interesting and complex data in the form of the Xenopus kidney which shows complex shape and texture.
- Expands from the previous 2D UNet work by working with not only fully annotated 3D data but also semi-annotated 3D data which is a more challenging task.
- Contains an ablation study in the results to validate the importance of batch normalization and the volumetric neural network components in their results.
- The author shares a link toward the code which allows us to potentially reproduce the claimed results.

Reasons to reject the paper

- The dataset is very small, only 3 samples, which makes it difficult to evaluate if this model would generalize to other samples or if the model would generalize to other tasks.
- The authors mention other related works but fail to compare their work to these methods on their dataset or any other dataset. Additionally, the author only proposes one metric (IoU) to determine the performance of their solution.
- The article would benefit from more details on how the semi-automated training is set up.
- It is unclear if the annotations are done by the non-medical expert authors or the medical expert authors. Furthermore, it is unclear if there was any sort of inter-annotator agreement between medical experts to produce these segmentation masks. This may lead to biased segmentation masks which could tend to simpler structures than the true structures.

Comments

Weak reject. Overall, the work presented in this paper is good. It proposes to use deep networks applied to a difficult problem with limited data, which is novel at the time of writing the article. However, the size of the dataset, the lack of metrics and method comparisons, and the lack of detail on the methodology and the data collection lead me to believe this work needs another iteration of writing and experimenting.

Major Comments

A few sentences explaining the motivation for this particular kidney segmentation task would be beneficial in getting the reader to engage with the paper. Furthermore, an explanation of confocal microscope imaging for those informed on medical imaging but unaware of the uses and particularities of this imaging technique would make the article more digestible.

As mentioned above, the comparison with the 2D UNet and the ablation study helps to highlight the importance of their contribution. However, I believe there is a missing paragraph giving more context about the semi-automated annotation experiments. Furthermore, all the models in the related works should be included in the comparison table in the results section. Additionally, I would like to see the evaluation section expanded to include more metrics other than IoU. Furthermore, a sentence mentioning how the data was annotated, as explained previously, would benefit the credibility of the results.

For the data, the author could take two avenues. Either they can continue in this limited data scenario but in this case, they should still add more test samples to their dataset. The other avenue would be to evaluate their method on other datasets such as BraTS or LiTS which have more samples and are a common benchmark for this type of tool.
BraTS : https://arxiv.org/ftp/arxiv/papers/2305/2305.19369.pdf
LiTS : https://arxiv.org/pdf/1901.04056v2.pdf

It would be nice to have an explicit discussion section outside of the results and conclusion section. As of now, it is unclear what is the potential future work and what are the limitations of this method.

Even if the data augmentation techniques such as elastic deformations are from a previous publication, it is necessary to have a summary of these techniques and a table with their hyperparameters in the appendix. Furthermore, a table in the appendix that resumes the different hyperparameters used to train all models would help the reader quickly find these important parameters to reproduce the results or to gauge what has worked well in the past for future work.

Minor comments

The breakdown of the introduction is bizarre with a section 1.1 for related work but no other subsection. Maybe review the structure so that there are two subsections, take out the related works from the introduction section or use a different type of heading.

For Section 3.2, it is common practice to include the loss function as an equation somewhere in the methodology section and to refer to it instead of just naming it. In my opinion, it would help readability if this were converted to an equation and referenced: "The IoU is defined as true positives/(true positives + false negatives + false positives).".

In the network architecture section, I appreciate the use of the figure. However, it would be nice to add a table for the tensor sizes of various inputs, hidden layers and output sizes instead of putting them in the text.

"Data augmentation is done on-the-fly, which results in as many different images as training iterations. We ran 70000 training iterations on an NVIDIA TitanX GPU, which took approximately 3 days." This sentence is difficult to read and should be reviewed. Additionally, it is misleading to say that there are as many images as training iterations as there is always a non-zero chance of having the same transformation parameters for an elastic deformation and the images are not different images but augmented images.