

A Project title

STUDIOS: Segmenting Tomograms Using Different Interpretation of Simplicity

B In a nutshell

3D tomograms (volumetric images resulting from computed tomography) acquired at synchrotron or laboratory facilities provide a view into the internal structure of matter. The need for accurately segmenting tomograms has not yet been met by the deep learning (DL) methods, even though DL has successfully solved most segmentation problems in computer vision. Contrary to an obvious suggestion, I claim that successful DL methods for segmenting photographs should *not* serve as inspiration for segmenting tomograms, as it will lead to large and data-hungry models. Instead, I suggest developing models which exploit the *simplicity* of tomograms.

The methods I intend to develop diverge from the current image segmentation paradigm, and cannot lean on the existing frameworks for DL. This makes method development difficult and time-consuming. But, if successful, the new methods will pave the way for future research in DL for tomograms.

C Research idea and context

DL methods achieve remarkable results on many segmentation problems in computer vision. When it comes to segmenting large 3D tomograms, e.g., tomograms captured at synchrotron facilities like ESRF or MAX-IV, the use of DL is still modest. For example, in scientific highlights for 2021, ESRF mentions *one* article [1] where DL is used for segmentation. They applied a freely available U-Net-like convolutional network to segment and quantify the nano-structure of nacre (mother of pearls). From a manually labeled subset of data they obtained satisfactory (not great) results (Fig. 1), concluding their work *demonstrates the potential of DL*. That DL struggles with this segmentation problem is frustrating, when considering that thresholding, the ultimately simplest segmentation approach, yields a reasonable starting point for nacre segmentation.

In [1] they suggest using larger networks. More labeled data to train the model is another suggestion heard when DL disappoints. This leads to high memory and computational requirements making DL models slow and difficult to use. I argue that the size of the network, and the amount of labeled data, are not the *only* reasons behind the somehow disappointing behaviour of DL when used on tomograms of materials with relatively simple structure.

Adding to frustration about the under-performance of DL, is the fact that a tomogram as in [1] is itself a remarkable accomplishment of world-class X-ray physicists using facilities that cost millions of euros to run. The very fact that the sample *was* scanned at ESRF witnesses that yet unanswered questions

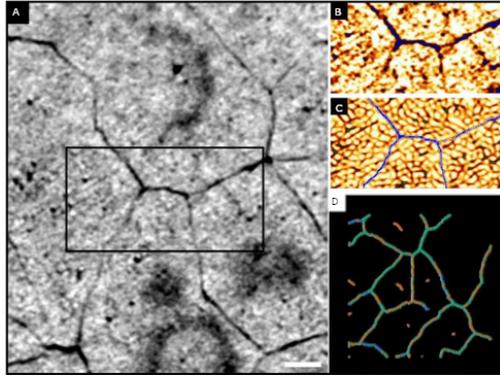


Figure 1: Figure taken from [1] with data (A), thresholding (B), manual labeling (C), and segmentation result (D).

are potentially hidden in the data. So why do we still struggle to use DL segmentation when analysing this precious data? Ironically, the uniqueness of the tomograms is yet another factor contributing to the problem, since pre-trained models have little use on unique data.

That DL is state-of-the-art for segmentation of photographs is indisputable [2]. Here, methods with many convolutional layers, hourglass (encoder-decoder) architecture, and skip connections (like U-Net) are popular choices. Such methods are available in open-source libraries provided by companies like Facebook and Google. The ease of use of DL methods led to enormous research activity which increased the segmentation accuracy to levels that were unthinkable before DL.

When developing methods for 3D segmentation, the logical choice is to start with what works well (and is available) for 2D. The adaptation from 2D to 3D is usually considered being a *generalization*. Therefore, methods are adapted for example by using 3D convolutions, increasing the complexity. This makes the models even more dependent on training data.

It is important to mention that 3D DL methods are developed mostly with a focus on medical imaging. Here, images often depict organs, and rather large standardized medical datasets are beginning to emerge. Also, medical images differ tremendously from tomograms in MS. As a result, the tools available to MS researchers today are, in a sense, *twice removed* from their original use case: those are methods for medical images which are in turn adapted from methods for photographs.

The question is: can segmentation of tomograms be achieved better if using *different* DL methods than those already established for segmentation of photographs? I argue that this is possible if carefully exploiting the *simplicity* of 3D tomograms. I am certain that we can develop DL methods targeted for tomograms which would require significantly less training data and computation. If efficiently implemented and made publicly available, such methods would have

a tremendous impact on 3D imaging.

D Proposed method of solution or concept

Despite being 3D, tomograms are, in certain aspects, significantly simpler than photographs. Photographs capture the appearance of objects which may change dramatically depending on the viewing angle and lighting conditions. Photographs, being projections, exhibit perspective, and occlusions. In contrast, voxel intensities of 3D tomograms encode the attenuation of X-rays in matter, which is directly linked to locally stable material properties. I denote this property *local stability*. Furthermore, 3D tomograms are reconstructions of 3D space, precisely capturing the shape (but not pose) of the imaged objects. I denote this property *geometric consistency*. Those two aspects are ignored when methods for photographs are generalized to 3D tomograms. Instead, I propose to develop DL segmentation methods that exploit the properties of tomograms.

My first hypothesis is that due to the local stability, small neighbourhoods extracted around points in tomograms carry more information needed to encode segmentation classes than in photographs. (In photographs, patches from a window of the car and a window of the house may look very similar.) This means that architectures suitable for segmenting tomograms should rely more on (not necessarily deep) rotationally invariant convolutions and less on downscaling and upscaling (which is used to encode long-range interactions).

My second hypothesis is that due to the geometric consistency of tomograms, encoding topological and morphological constraints will reduce the dependency on training data. This means that models for segmenting tomograms will benefit from different methods for representing the segmentation than the customary voxel grid. An alternative is a surface mesh or neural representation. Such representations are starting to emerge for segmentation [4, 3] but are still focused on medical data and still make use of large U-Net-like models.

E Major gains and obstacles

The main obstacle to this project lies in a substantial amount of code that needs to be written. As mentioned, the current development in computer vision is largely fueled by open-source DL frameworks from big commercial companies (PyTorch by Facebook and TensorFlow by Google). This makes it easy for researchers to address the type of problems the provided frameworks are developed for: problems involving photographs. And when addressing other problems (like tomograms) it is still easier to use an existing framework.

As we aim to develop *different* models, we will need to supplement the existing frameworks with a considerable amount of software developed in-house. This is a relatively high effort-wise investment considering how easy it is to use existing frameworks.

Once we produce an initial framework for DL segmentation of tomograms,

and demonstrate its potential, it will be easier for other researchers to cross this effort-wise threshold and contribute to the development of efficient methods for segmenting tomograms. We plan to assess our models on parameters including: size of the volumes processed, time used for training and inference, amount of labelled data needed.

Ultimately, the goal is to develop an easy-to-use, easy-to-train, simple and efficient DL segmentation framework and make it accessible to the research community. This will revolutionize the application of 3D imaging for quantitative studies of micro-structure. The impact may resemble what happened in computer vision with the use of DL methods becoming a standard, allowing us to solve previously unsolvable problems.

F Appropriateness

In research, as in many other situations, it is advisable to stand on the shoulder of giants. This research idea suggests the opposite: following the path which we know is difficult. For a conventional program, this might be too controversial and bold.

Also on a personal level, pursuing this research idea requires high commitment. Any image analyst aiming at fast success, is career-wise much better off by doing the opposite of what this project suggests. Following the flow (using the provided frameworks, and solving problems measured by benchmarks) is not only an easier way to good results, but also an easier path to highest-ranking conferences (which is where image analysts want to publish).

G Probable objections

An objection is: “If making simpler models can solve the complex problems, why has it not been done?” The answer is that the models I propose to develop are only simpler in some aspects. In other aspects (geometric models, computational efficiency, annotation, and visualization in 3D) the challenges are still large.

References

- [1] Maksim Beliaev et al. Quantification of sheet nacre morphogenesis using X-ray nanotomography and deep learning. *J. Struct. Biol.*, 2020.
- [2] Minaee et al. Image segmentation using deep learning: A survey. *IEEE T-PAMI*, 2021.
- [3] Ashwin Raju et al. Deep implicit statistical shape models for 3d medical image delineation. *arXiv preprint*, 2021.
- [4] Udaranga Wickramasinghe et al. Voxel2mesh: 3d mesh model generation from volumetric data. *MICCAI*, 2020.