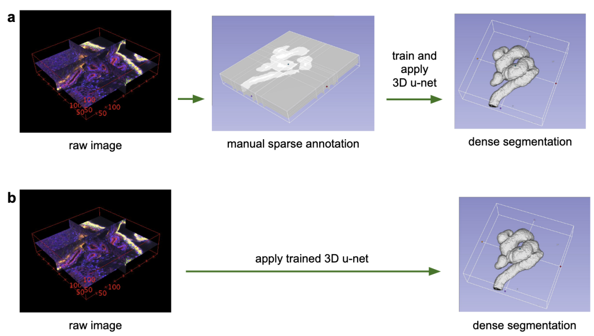
3D UNet

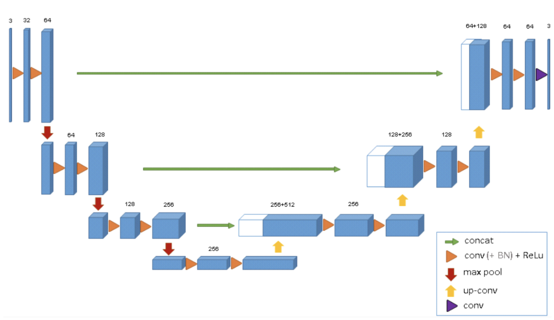
A Preprint

Authors suggest a deep network that learns to generate dense volumetric segmentations, but only requires some annotated 2D slices for training.



each layer contains two 3 × 3 × 3 convolutions each followed by a rectified linear unit (ReLu), and then a 2 × 2 × 2 max pooling with strides of two in each dimension. In the synthesis path, each layer consists of an upconvolution of 2 × 2 × 2 by strides of two in each dimension, followed by two 3 × 3 × 3 convolutions each followed by a ReLu.

Volumetric Segmentation with the 3D U-Net



Authors also avoid bottlenecks by doubling the amount of channels before pooling.(If you reduce the spatial dimensions without increasing the number of feature channels, you risk losing too much information for the network to perform its task effectively. The network might not be able to "represent"the important features from the input in such a compressed state.)

The first set of experiments focused on a semi-automated segmentation scenario. Here, the goal is for the network to produce a complete, dense 3D segmentation of a volume when given only a few manually annotated 2D slices from that same volume. To quantitatively assess this, they performed a 3-fold cross-validation on all manually annotated slices from their three samples. The results demonstrated that the 3D U-Net could achieve a high average Intersection over Union (IoU) of 0.863, significantly outperforming an equivalent 2D implementation and showing the clear benefit of leveraging the full 3D context. They also showed that the segmentation accuracy steadily improved as more annotated slices were provided, proving the model's ability to learn effectively from minimal user input.

The second set of experiments evaluated a fully-automated segmentation scenario. In this setup, the network was trained on two of the sparsely annotated kidney volumes and then tasked with segmenting the third, entirely unseen volume. This tested the model's ability to generalize its learned knowledge to new data. The experiments showed that the 3D architecture provided a substantial performance gain compared to a 2D network, confirming its suitability for building a representative training set from sparsely labeled data to automate segmentation on a larger scale.