Stereology as weak supervision for medical image segmentation

Building large medical imaging datasets for image segmentation is a challenging task due to manual outlining. In this work, we explore the use of stereology to cut the costs of annotation.

Introduction

Stereology

Through the stereology, the average volume density volume of a target structure can be estimated accurately from the sum of areas on planar sections distanced by “T”. Such areas can be in turn estimated using a test grid consisting of a set of points “p” arranged in a rectangular or hexagonal lattice, each with an associated surrounding area “A”. By counting the number of test grid points “P” falling into the structure of interest, the volume of the structure is estimated as “V = T*A*∑P”.

U-net

The most frequently used CNN architecture for medical image segmentation is U-net, published by Ronneberger et al, and its derivatives.

In this work, we adopt the point counting method as weak supervision of an image segmentation task and demonstrate that training a U-net with a point counting mask incurs in limited decrease in segmentation performance, which is out-weighted by the opportunity of annotating volumes at a much lower cost.

Methods & Experiments

In this work a 3D U-Net is trained using both a dense segmentation target as well as targets generated from grids at different point densities on the same data. The focus was on hexagonal grids, to maximize the chances of overlap with rounded biological shapes. The horizontal distance between two adjacent grid points is denoted as stride. Grids are superimposed on each segmented slice and their offset is randomized per slice. Grid points are then manually classified whether or not they overlap with the structure of interest. At each training iteration, a dense ground truth is dynamically generated by padding each grid point on its 2D plane by a variable amount. During inference, the scan is subdivided into consecutive slabs spaced by one slice. An average of all predictions at each voxel is assumed as the output,

Results

We have shown that weak supervised stereology based learning methods achieve accuracy within a few percent of full supervised methods for all grids.

Performance measured by dice coefficient obtained using a dataset composed of 320 scans and a grid of stride 5 exceed the ones obtained using a dense target on 100 scans. Using large strides (greater than 9), along with randomizing padding are a key factor for achieving good segmentation performance.

We have sought to create a tool which improve precision workflow for tagging large medical image data sets.