Name: Ge, Jiaxin
SID: 58153795
UCInetID: gej4

Midterm write-up

Methods

There are 342 prostate images that have been used to train three different CNN models and there are 69 images used to validate those models.

All three models are based on U-net architecture. Model 1 uses the standard 2D U-net architecture. It has 10 layers and uses convolution operations and same padding to subsample the training data into segmentation with local features to make the network deeper. Then it uses convolutional transpose to reverse the convolution operation and upsample features. Finally, it concatenates between each contracting and expanding layer to ensure glocal features that also can be learned. The network uses RELU as activation function and uses batch normalization. Model 2 uses the exact same architecture as Model 1, but the input images are different. The images are pre-processed by other training models and crop from 256 x 256 resolution to 128 x 128 resolution. Model 3 is based on the inputs from Model 2. Different from the standard U-net architecture, Model 3 uses a hybrid network in which it uses 3-slice 3D input to predict every desired 2D slice output. In the contracting layer, the model uses strided convolutions with same padding in xy direction as well as it uses valid padding in the z direction. The model also uses residual connection to prevent unstable activation for some neurals.

The three models use Adam as the optimizer in which the learning rate is set to 2e-4. The batch size is set to 12 and uses cross entropy as loss function. Approximately, 12 iterations are required for convergence. During the training period, the value for loss function decreases and converges.

In the evaluation session, I use the mean, median, 25 percentile and 75 percentile for both transitional and peripheral zones in order to evaluate model accuracy.

Results

Both training and validation cohort statistics of the three models are conducted. The performance of the models are progressional from model 1 to model 3. More details are shown in the tables below.

Training cohort statistics

	T TRANSITIONAL ZONE mean	T TRANSITIONAL ZONE median	T TRANSITIONAL ZONE 25th-tile	T TRANSITIONAL ZONE 75th-tile	T PERIPHERAL ZONE mean	T PERIPHERAL ZONE median	T PERIPHERAL ZONE 25th-tile	T PERIPHERAL ZONE 75th-tile
model 1	0.757282	0.767931	0.722961	0.807674	0.901161	0.914520	0.884082	0.926927
model 2	0.799555	0.814830	0.771032	0.843021	0.912978	0.919799	0.900713	0.932697
model 3	0.802659	0.814742	0.773855	0.843967	0.916865	0.926121	0.904203	0.938459

Validation cohort statistics

	V TRANSITIONAL ZONE mean	V TRANSITIONAL ZONE median	V TRANSITIONAL ZONE 25th-tile	V TRANSITIONAL ZONE 75th-tile	V PERIPHERAL ZONE mean	V PERIPHERAL ZONE median	V PERIPHERAL ZONE 25th-tile	V PERIPHERAL ZONE 75th-tile
model 1	0.701912	0.721722	0.657567	0.767067	0.889729	0.900158	0.872128	0.914436
model 2	0.716508	0.737924	0.650525	0.795677	0.887891	0.896518	0.874417	0.915683
model 3	0.723521	0.750314	0.656455	0.794174	0.892764	0.903859	0.877962	0.922896

Discussion

With the same hyperparameters and iterations, the results of the three models are expected because the three models are progressional complex. Model 2 uses less noisy and more detailed data compared to model 1, whereas model 3 uses more complicated techniques in the architecture. The performance improvement between model 1 and model 2 is approximate 5% and that between model 2 and model 3 is less than 1%. The segmentation performance within the peripheral and transitional zones affected equally in all models, because they were both solving classification problems based on all models so their performance should be similar.