Name: Ge, Jiaxin
SID: 58153795
UCInetID: gej4

Final write-up

Methods

There are 53298 individual images that have been used to build three different CNN models, in which 29649 used to train the models and the remaining used to validate their performances. All three models use convolutional neural networks (CNN) architecture as baseline, building other techniques above to optimize the performance with different loss function and matrices.

There are three models in my final project. The first model uses a standard CNN architecture for binary classification. This classification model has seven layers and the last layer flattens all the convolutional feature maps in order to get a classification result. Since the pneumonia infection can be subtle to find in the images, the model uses two convolution operations in each layer to make the features deeper. The function uses standard softmax cross-entropy, in which softmax can standardized the logit score and cross-entropy can eliminate small errors.

The second model is a sole segmentation model, based on U-net architecture with other features. It 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 same padding can subsample the training data into segmentation with local features to make the network deeper, while the valid padding can remain geometric features in the z direction.It also uses residual connection to prevent unstable activation for some neurals. In the expanding layer, the model uses convolutional transpose to reverse the convolution operation and upsample features. It concatenates between each contracting and expanding layer to ensure global features that also can be learned.

The third model uses both a classification and a segmentation loss function to help regularize the network. The client object is overloaded to create a custom mask for class weights and masks. The model also uses a standard segmentation U-net architecture, with several strided-1, strided-2 and convolutional transpose operations. It also uses residuals layers to skip some connections in order to prevent vanishing gradients. It has an extra flattened layer for classification, with a weighted pixel-wise softmax cross-entropy loss function.

All three models use Adam as the optimizer in which the learning rate is set to 2e-4. During the training period, the values for loss function are prune to converge. Since we're building a global classifier, we need to collapse our final result into a binary prediction, no matter if it is a classification model or a segmentation model. We use accuracy to evaluate the performance of each model. Moreover, Model 2 and Model 3 use the same threshold for evaluation, which is 1000.

Results

Both training and validation cohort statistics of the three models are conducted. All the accuracy meets the requirement, which is >= 85%.The performance of the models are progressional from model 1 to model 3. More details are shown in the tables below. Note that the median, 25th-tile and 75th-tile are the data from model.predict() for each model. Mean is the accuracy for each model.

Training cohort statistics

	mean	median	25th-tile	75th-tile
model 1	0.876	0.0	0.0	1.00
model 2	0.886	81.0	0.0	5044.75
model 3	0.867	314.0	9.0	4145.75

Validation cohort statistics

	mean	median	25th-tile	75th-tile
model 1	0.863	0.0	0.0	1.0
model 2	0.884	0.0	0.0	4009.0
model 3	0.850	189.0	7.0	2984.0

Discussion

The global binary classification results show that all the performances are above 85%. The performance of model 2(segmentation) is the best, but with less than 1% to 2% better accuracy compared with the other two models. The results of model 2(segmentation) and model 3(dual loss) should be expected to be higher than model 1(classification), but actually their performances have shown to become similar. The reason might be because model 1(classification) uses multi-convolution operations in each layer so it creates featured details and high frequency patterns given the subtle pneumonia infection subtle performance. Overall, we can conclude that multi-convolution operations can optimize the performance of these models.