Reproduce Distributed Learning Networks for Medical Imaging and Investigate the Performance in Edge Scenarios

Video Presentation: https://www.youtube.com/watch?v=uy_4VosHzUQ

Miao Zhang [miaoz18@Stanford.edu]

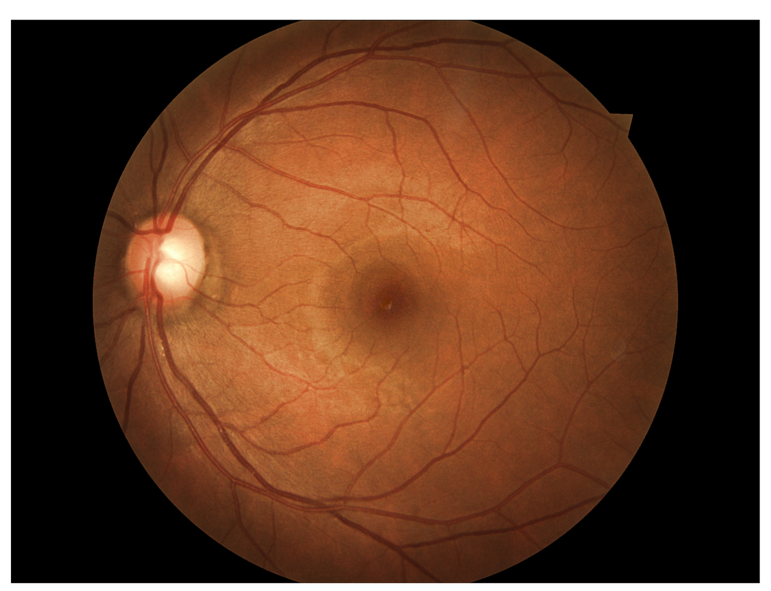
ABSTRACT

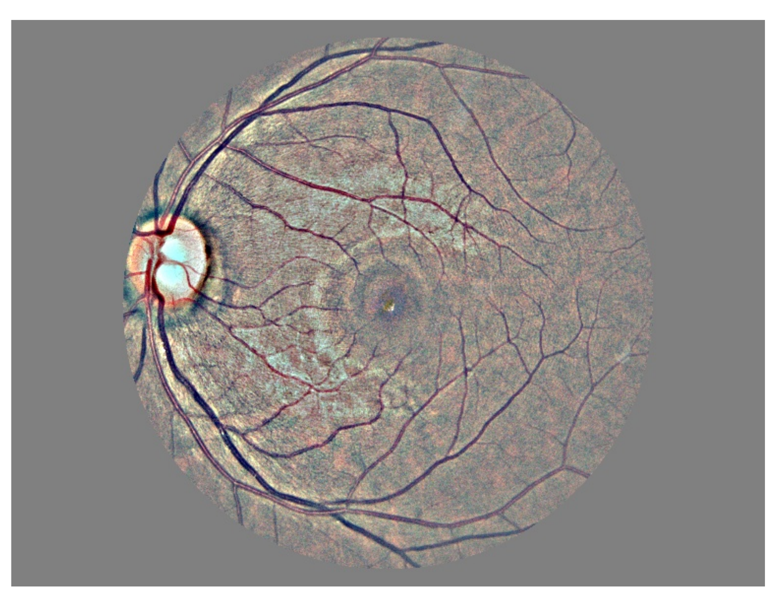
Data collection is a challenge for large scale machine learning, especially in healthcare area where privacy and regulation of data sharing need to be addressed carefully. Federated Learning aims to build incentive models on distributed and unstructured data for data transparent ecosystems. I implemented the technique that trains the deep neural network on 4 separate client sites with no data centralization but only the communication of model weights, and obtained the result with test accuracy around 74\%. I investigated the difficult scenarios that happens in real life including data quality and class imbalance.

PROBLEM STATEMENT

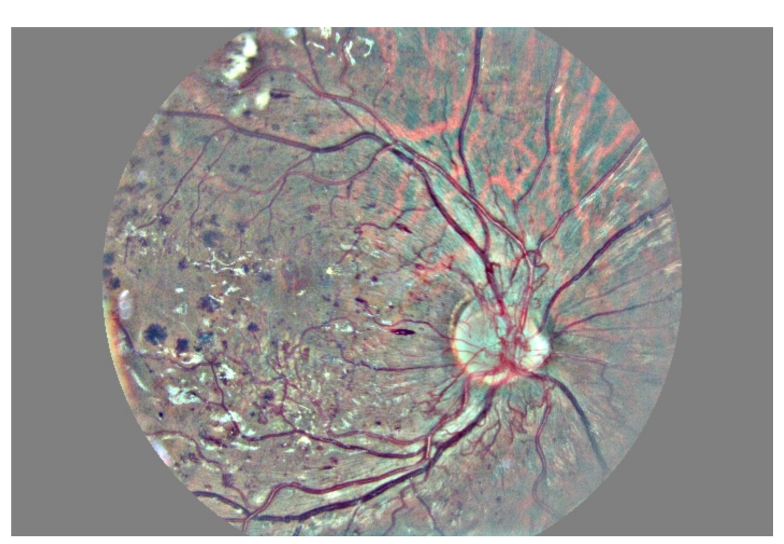
Diabetic Retinopathy Detection

Rating: 0

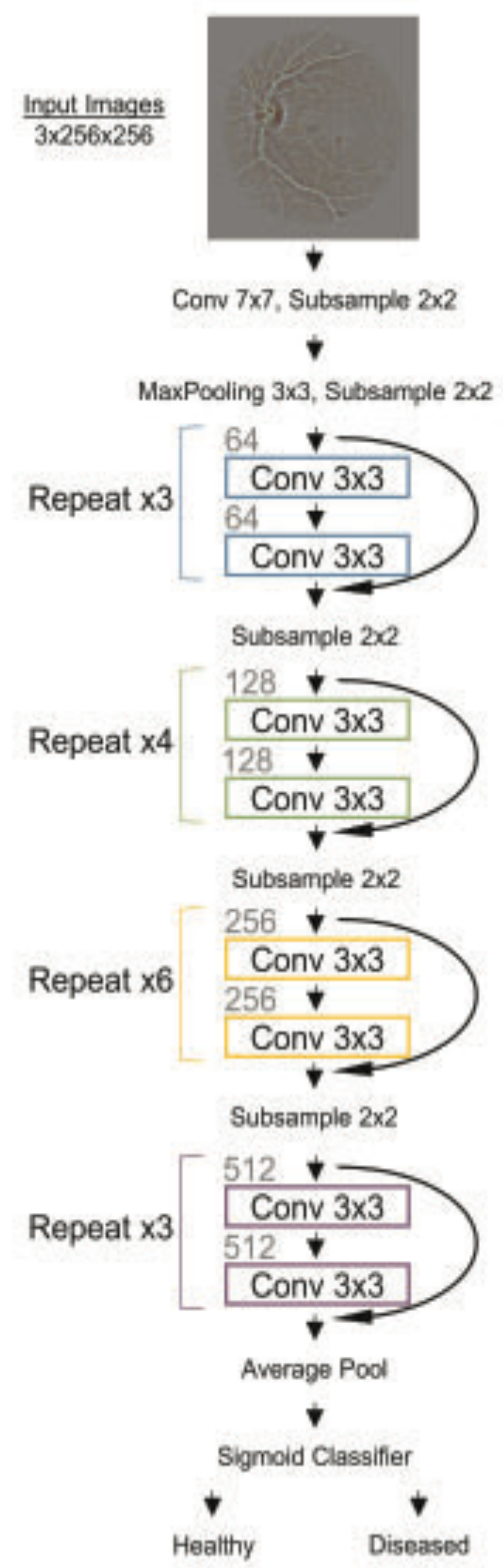




Rating: 4



GENERAL APPROACH



A 34-layer residual network architecture pretrained on ImageNet to do the images classification, shown in the left figure. The input shape was (256, 256, 3) and the output layer was a 2-class sigmoid layer. The weights of the network were optimized via a stochastic gradient descent algorithm with a mini-batch size of 32. The objective function used was binary cross-entropy. The learning rate was set to 0.001 and momentum coefficient of 0.9. The learning rates by multiplying by 0.25 when the same training images were used to train the neural network 20 times with no improvement of the validation loss.

DISTRIBUTED LEARNING

In each averaging round, the server model weights were distributed to local institutions to train models for a predetermined number of epochs, then the server collected and averaged the output model weights of all institutions and updated the server model state.

EXPERIMENTAL RESULTS

One Institution with data heterogeneity

	Testing accuracy (%) (Image quality variability)	Testing accuracy (%) (Class imbalance variability)
Skip the variable institution	72.9	72.9
Not skip the variable institution	71.5	70.2

Three Institutions with data heterogeneity

	Testing accuracy (%) (Image quality variability)	Testing accuracy (%) (Class imbalance variability)
Non-weighted averaging	70.7	68.3
Institution skipping	62.6	62.6
Weighted averaging	70.9	70.1

CONCLUSIONS AND FUTURE WORK

(1) Introducing data heterogeneity to institutions impacts the network performance, and the network is more vulnerable to the class imbalance variability than the image quality variability.
(2) Skipping the institution with heterogeneity helps mitigate the performance loss in scenarios where only a small number of institutions have data heterogeneity.
(3) Weighted averaging is an attractive alternative to mitigate performance loss of multiple institutions heterogeneity scenarios.

The future work includes more detailed analysis for the deployment of distributed learning in real-world medical imaging settings, for example, to investigate mores types of edge scenarios and better distributed learning methods.