Undersampled Retinal OCT Image Reconstruction

Xiao Tang, Holly Wen, Wade Wei
Department of Biomedical Engineering
Duke University
Durham, NC 27705
xt39@duke.edu yw425@duke.edu zw213@duke.edu

Abstract

Optical coherence tomography volume scanning has a trade-off between scanning speed and low-axis resolution because more turning points has more flyback time. In order to increase the scanning efficiency, real-time OCT has few turning during scanning process, leading to quite poor low-axis resolution. In this project, we apply retinal OCT images to a fully dense U-Net neural network and reconstruct images from the downsampled data. Our pre-processing includes randomly cropping, denosing and contrast enhancement. Compared with other existing interpolations, our model has similar results with 50% downsampling rate but a better denoising effect. It's possible to increase the low-axis resolution but maintain the same or even better fast-axis resolution in OCT.

1 Instructions

Optical coherence tomography (OCT) produces depth-resolved retinal images for diagnosis and monitoring eye disease. This imaging technique uses low-coherence light to acquire cross-section images of retina based on interferometry. After decades of development, OCT can provide three-dimensional data in real-time with faster light source and novel design.

The 3D volume scan of OCT is a series of parallel B-scans along the low-axis. Recently, research groups have combined the OCT with surgical operating microscopes, achieving more than 10 frames per second in volume scan. However, limited by the flyback time of the scanner, real-time volume scan has a trade off between the scanning speed and low-axis resolution. Thus, the en-face view of the volume scan has anisotropic resolutions, which requires a learning curve for surgeons to recognize the features. In order to increase the low-axis resolution, the number of B-scans for every volume scan increases but each B-scan is downsampled. Therefore, we plan to use deep neural network training to reconstruct the downsampled images, which increases the low-axis resolution and hold the same fast-axis resolution at the same time.

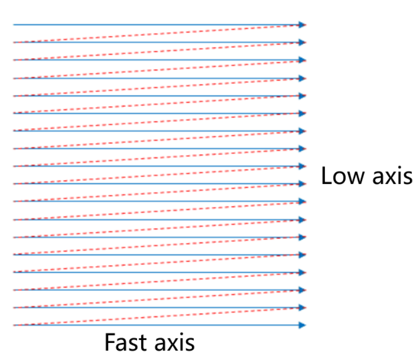

Figure 1: Conventional scanning mode. Note that the red dots are related to the inefficient scans

The goal of this project is to train the model to reconstruct the downsampled B-scans. Firstly, we randomly crop the images to fit the size of the model. Then we use several physical layers to denoise, enhance the contrast and undersample the training images. After training, the model is used to reconstruct the retinal B-scans after artificial down-sampling process. Our results…

2 Related Work

In recent two year, some networks like IKC, USRNet and DPSR were proposed to reconstruct super-resolution images. Hossein et al. present a nonlocal weighted sparse representation method for reconstruction of retinal optical coherence tomography images. They use the subsampled data to train the model with efficient denoising and interpolation algorithms and mainly focus on the denoising effect on the reconstructed images.

In this project, we use the fully dense U-Net network, first proposed by Guan et al. to train the artificially undersampled data. The fully dense U-Net model is used by Anthony et al. to reconstructed the undersampled photoacoustic microscopy images and performs better compared with convolutional neural network architectures. Based on their work, we try to apply retinal OCT data and reconstruct the images.

3 Methods

3.1 Deep learning network Fully dense U-net

The details of our regenerative model can be found at this link: (PDF) Reconstructing undersampled photoacoustic microscopy images using deep learning including the loss function: To summarize, this network utilizes the fully dense U-Net which integrates dense blocks in the expanding and contracting paths of the U-net. In a word, the FD U-net allows each layer within a dense block to be concatenated with the outputs of all the previous layers, which makes sure that each layer only needs to learn progressively that either augment the prior layer or diversify the collective feature set. Another thing worth noting is that the model we used has two modifications from the original model: Relu activation has been replaced by Elu activation The max-pooling layers were replaced with a 1x1 convolution block and a 3x3 convolution block with a stride of 2.

3.2 Data preparation

The OCT dataset we used is the retinal OCT images from Paul Mooney. The dataset consists of 227 OCT images with the size of 512x496. The dataset was randomly divided into roughly 80% training, 10% validation and 10% testing. The OCT images were thresholded with a 3x1 median filter and a 1x3 median filter to remove the noise and artifacts in the images. We also performed contrast enhancement and grayscale intensity rescaling. The training part of the dataset was applied to the optimization of the stochastic gradient descent algorithm and the validation dataset was intended to gather the best model information according to the validation metrics, but it was not used extensively for our purpose.

3.3 Data augmentation

In order to fit our data to the model which only takes in 128x128 pixels, we cropped our input images to 128x128 patches. To properly implement the cropping, we used a random crop in the tensorflow toolbox to augment and standardize the images. In this fashion, a standard subimage with 128-by-128 pixels at a random location within the original fully-sampled data was gathered. For our purpose, 5 random crops per iteration was performed on each image.

3.4 Downsampling Procedure

We manually downsampled our fully-sampled images in order to mimic the undersampling encountered in the tradeoffs in retinal OCT imaging data acquisition. For our purpose, we downsampled the x-axis by a ratio of 2:1 to imitate the increasing artificial scanning step size in x-direction behavior in OCT data acquisition. The missing pixels can be added back according to the downsampling ratios.

We used the zero-fill method in response to that. To elaborate on this method, the missing pixels were replaced with a constant value of zero.

3.5 Train the Network

All the networks use the Adam algorithm with a batch size of 16 for our purpose. The models were trained for 150 epochs.

4 Results

4.1 Reconstructed images

Downsampled and reconstructed Images are provided below.

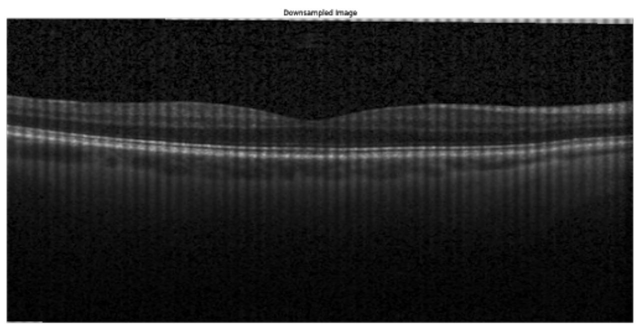

Figure 2: Downsampled image

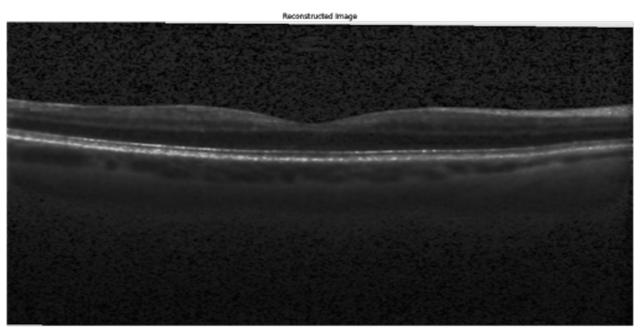

Figure 3: Reconstructed image

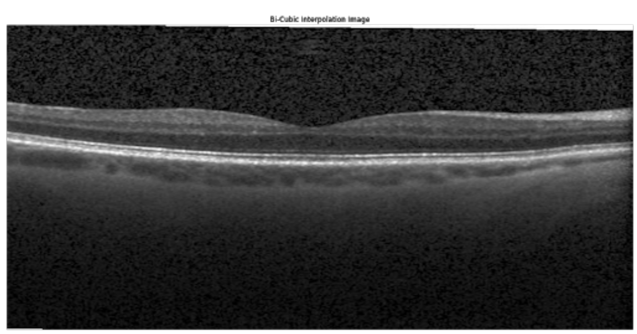

Figure 4: Bicubic

Our method was able to reconstruct downsampled images. Comparing with bicubic interpolation, reconstructed image quality was less improved and much more denoised.

4.2 Loss function

Graphs of various parameters vs. epoch

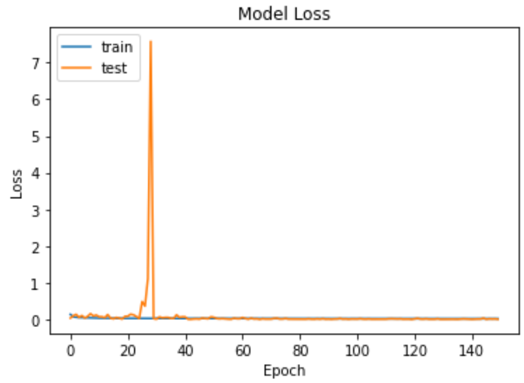

Figure 6a: Model Loss

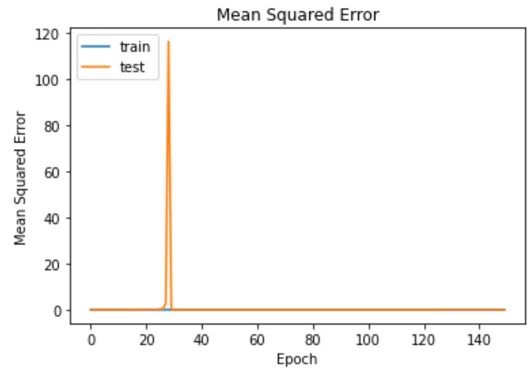

Figure 6b: MeanSquaredError

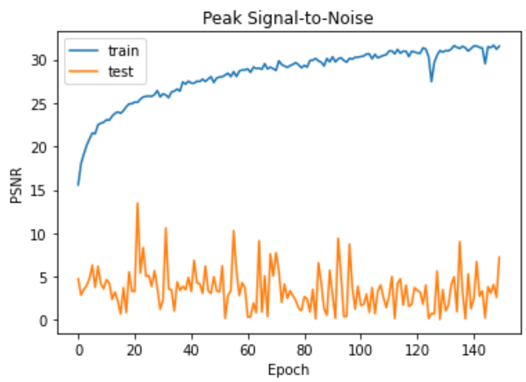

Figure 6c: SNR

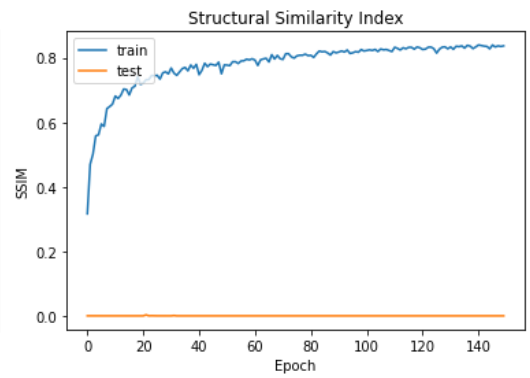

Figure 6d: Similarity

Parameters of train dataset look great; for test dataset, there's this weird spike. The structural similarity index is high in training samples but nearly zero in testing samples. It might be related to the randomly selection of the testing samples. A better comparison function needs to be introduced.

5 Discussion

First, we look at parameters to assess our model, including model loss, mean absolute error, mean squared error, KL Divergence, peak SNR and structural similarity index. For the train dataset, model loss drops after every epoch. After 150 epochs, mean squared error becomes very low. This meets our expectation that this FD-UNet model will be able to rescontruct a downsampled image set.

However, parameters for the test set was very strange. There's this weird spike for both model loss and mean square error (Figure 6a-b). What's more weird is this similarity index, denoting the similarity of trained result with actual result. It should increase as epoch increases; however, test stays at 0 all the time (Figure 6c-d). Possible error could come from bad train/test slicing: we are using 80/10/10 for train/valid/test, but that might be not too ideal. Meanwhile, test dataset could contain a lot of noise images with a large percentage, which therefore affects the performance of the actual test dataset that we feed into the model. On the other hand, there could be some coding errors within our model that stops our test from performing well.

Afterwards, we displayed our reconstructed image. Comparing with downsampled image (Figure 2), the reconstructed image was able to recover most of the information that the original possess. However, it does not have better visualization than a more mature model, bicubic interpolation. One advantage that our model has a better denoising effect compared with other models. In general, we do believe that our model has a promising effect on data cleaning, which is very helpful for image reconstruction as well.

One thing to note is that for our downsampling process, we tried a 50 percent downsample ratio. This results in a promising reconstructed image. Since we do not have much time, this is the only ratio we tried. For a future project plan, we could implement more sampling ratios to see how our model would perform. Another way to play with our model is by introducing another direction. So far, downsampling happens on one direction; it would be interesting to try on the other direction too, as well as on both directions (x,y). We can explore more of this idea in the future.

Moreover, we explored using both 128 by 128 image size and the original size without cropping. This starts from the observation that random cropping introduces a lot of background noises, which displays as black pixels on the raw images. These noises would affect our model performance. It turns out both image sizes do not have a large difference on our model performance. Also, for both image sizes, test dataset similarity is also always 0 as epoch increases. We are hoping to do more center cropping to remove more back noise for a better performance.

6 Conclusion and Future Work

Using the FD U-Net, we successfully reconstructed the downsampled OCT data to its original form, overcoming the trade-offs in OCT data acquisition while generating smooth, de-noised OCT image samples.

Since we focus on developing a regenerative model for OCT data specifically, a specialized data pre-processing pipeline could be developed for OCT images, addressing the concerns we encountered while applying the network to the current preprocessed data. Such pipeline could entail segmenting the volume of interest(Between layer ILM and PLE), threshold out unwanted information, applying further de-noise techniques, and adjust contract-enhancement ratio etc.

References

[1] Pan, H., Yang, D., Yuan, Z., Liang, Y. (2020). More realistic low-resolution OCT image generation approach for training deep neural networks. OSA Continuum, 3(11), 3197-3205.

[2] Abbasi, A., Monadjemi, A., Fang, L., Rabbani, H. (2018). Optical coherence tomography retinal image reconstruction via nonlocal weighted sparse representation. Journal of biomedical optics, 23(3), 036011.

[3] DiSpirito III, A., Li, D., Vu, T., Chen, M., Zhang, D., Luo, J., ... Yao, J. (2020). Reconstructing undersampled photoacoustic microscopy images using deep learning. arXiv preprint arXiv:2006.00251.

[4] Keller, B., Draelos, M., Tang, G., Farsiu, S., Kuo, A. N., Hauser, K., Izatt, J. A. (2018). Real-time corneal segmentation and 3D needle tracking in intrasurgical OCT. Biomedical Optics Express, 9(6), 2716-2732.

[5] Halupka, K. J., Antony, B. J., Lee, M. H., Lucy, K. A., Rai, R. S., Ishikawa, H., ... Garnavi, R. (2018). Retinal optical coherence tomography image enhancement via deep learning. Biomedical optics express, 9(12), 6205-6221.