main commit id: 75b0c54ebbb0b70a3b3809bd37c4f5401559238a

Gradient Accumulation:
I was able to implement the gradient accumulation technique in the source code of Voxelmorph so that multiple training images could be considered before applying the gradients to update the weights. However, we must note that while gradient accumulation has helped us solve the batch training problem, it still does not help us increase the size of the training image.

Early Stopping:
As a result of GA, I was also able to adapt early stopping, which was not possible before due to the stochastic nature of the gradient descent (batch_size = 1), and the loss values fluctuated greatly, making it difficult to decide when to stop the training. Consequently, the number of epochs before the training ends is now about 10-15 epochs as opposed to the 300 epochs I did before. The results are still good enough and comparable to the mmi error obtained with a model trained to 300 epochs.

Janelia Dataset and Validation Data:
I included the Janelia data in the training along with DataSetGood, DataSetMedium, and DataSetRandom, resulting in a total of 660 training samples.
With this large amount of data, I was also able to prepare a validation set and include the validation measure in the voxelmorph source code, which is not present in the original. The decision to terminate early is based on this validation loss value. My breakdown of the data is as follows:
Training data : 560 (out of 660), which is ~85%.
Validation data : 100 (out of 660), which is ~15%

Learning Rate:
I tried increasing the learning rate from 1e-4 to 1e-3 and to 1e-2 to make sure that the learning rate was not too small, causing only a very small update of the weighting in each epoch, which could inadvertently cause a premature stop. I found that 1e-3 was fine, but 1e-2 was too high a learning rate and the training error increased.

Auxiliary Information:
I have looked through the source code for voxelmorph's semi_supervised training method - auxiliary information through segmentations. I think I have a good understanding of how we can integrate our auxiliary information (landmarks) into the training so that the network learns better.

Conclusion:
Finally, I would like to draw your attention to the fact that the results I had presented to you in the initial days were the result of a 300 epoch training of the network. Therefore, it is very possible that the model byhearted to specifically fit to the trained images. However, as I mentioned in Gradient Accumulation and Early Stopping, the results now after training only 10-15 epochs (decided by early stopping) also give comparable results to the model trained with 300 epochs.