Generative Models: Project 2

Hongli Lin (kag52rer), Michele Paterson (mipat104),
Sandro Jose Leal Miudo (salea100), Sven Klein (svkle100)

January 15, 2025

1 Choice of Methods

In this project, we used a Generative Adversarial Network (GAN) to generate fibre orientation maps (FOM). For the specific choice of model, we used the basic variant from the lecture with gradient ascent for better gradients. This choice was made since it was the easiest to implement. Addtionally, we used the StyleGAN2 model to generate images from the FOMs. This choice was made because it is a very powerful model that can generate high-quality images and its ablity to catch the different scales of details in the images. Due to time constraints, we did not implement the model ourselves but used the implementation from the StyleGAN2 repository. However other models such as Stable Diffusion, Flows would probably also acheive better results but with the cost of more computational power. For image generation, we used two different methods. The first method generates images from the target distribution unconditionally (task 3). The second method uses the corresponding transmittance maps as conditional inputs into our model (task 4). This requires a different model architecture, as we need to model to somehow infer the FOM from the given transmittance map information. For this, we decided to use the Pix2Pix method. This choice was made again because it is easy to implement, but also because we already had paired transmittance maps and FOMs, so Pix2Pix should perform well compared to other methods.

2 Problems

One major problem we had was that training would collapse. It turned out that this was due to our discriminator being too good. That would lead to it correctly classifying every image and the generator could never improve. We solved this by making several changes to the discriminator. We reduced the amount of layers and also added dropout layers to it. These changes lead to a more balanced training where the losses of the generator and the discriminator were more balanced. Another issue is that with some probability a tile from the FOM gets sampled where there is no brain tissue on it. This is usually because the cropped tile is too close to the edge of the image. This leads to the tile being completely black (FOM) or white and black (transmittance). Therefore generating a completely black tile will

always fool the discriminator. This could lead to a mode collapse. We acknowledged the problem and highly recommend implementing countermeasures like for example minibatch discrimination. We use a more simpler approach by removing the black and white tiles from the dataset with a certain threshold: during sampling, we check if the median value of the image is below and above a certain threshold. If it is below the threshold we discard the image and sample a new one. This approach is not perfect but it is a simple solution to the problem. Last but not least, the tile size chosen could also be a problem. As we have little knowledge about cells and their structures, we don't know what the best tile size would be. We chose 64×64 as a compromise between computational cost and detail. However, some bigger tiles could include larger structures and therefore be more informative.

3 Results

3.1 Unconditional Generation

By a visual inspection, we can tell that the model learned at least something about the target distribution. If not filtered, it generates the already mentioned black images as well as images that look similar to sampled patches from the FOMs. Since we don't know very much about the brain and its cell structures we can't really comment on if the generated images contain realistic structures or not. We also calculated the FID using the inception network and the validation set and achieved an FID of 4.18. This is not a very high score, but it also shows that we at least somewhat captured the data distribution. Meanwhile since the inception network was not trained on FOMs its features probably are not very meaningful. Beside the nature of InceptionV3 is not medical, the input of the model is 299x299x3,and the deepest layer should have 2048 features, our generated data is 64x64x3, which is needed to be rescaled to 299x299x3 to be fed into the model, meaning the original image have much less information than the model is trained on. This may cause the model to not scale well to our data (75 x 75 is the minimum size for InceptionV3. Ideally, we would use a model that is familiar with FOMs as input. We thought of using the discriminator for that, but since the generator is specifically trained to fool the discriminator this also would not lead to a very accurate score.

3.2 Conditional Generation

For this task, we have access to target images so we can perform an analysis by direct comparison. Just from a visual comparison, we can already tell that the generated images are not close to the real images at all. They lack a lot of details and color. This is confirmed by calculating the mean squared error (MSE). We calculate an MSE of 1340 with a standard deviation of 1903. This indicates the predictions are mostly wrong. However, we look at the structural similarity index we get a result of 0.45 with a standard deviation of 0.35. This score measures structural similarity on a scale of $[-1, 1]$. So it detected at least some similarities between the images. This indicates that the model can reason about the structure of the FOM from the transmittance map, but struggles to predict the correct orientation of the nerve fibres.