Overview of the solution:

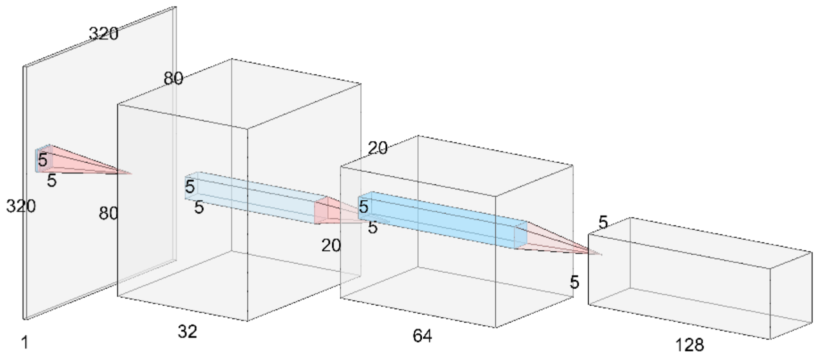
In this project we considered some reconstruction architectures such as U-Net, deep residual network, Transformers, SRGAN (a version of GAN), custom CNNs.
The chosen network was U-Net with skip connections, detailed as follows.
The input first passes through a sampling layer and then through our U-Net: the encoding phase involves sequential application of encoder blocks, each consisting of a convolutional layer ($kernel\ size = 5$, $stride = 2$), normalization, ReLU and max pooling ($kernel\ size = 2$), In the bottleneck, the input undergoes max pooling, a $1 * 1$ convolution, and interpolation, followed by concatenation + convolution to reduce the features. The decoding phase sequentially applies decoder blocks with up-sampling, convolutional layers ($kernel\ size = 5$, $stride = 2$), normalization, ReLU, incorporating skip connections, and finally, an interpolation produces the output image. We have added a visualization of the network with further explanations, also we will refer to the model by the name we gave it – GO_RecoNet.
Why We think that this was a good choice for us? U-Net is highly favored for medical image reconstruction due to its effective encoder-decoder structure with skip connections, which excel at capturing both high-level and local contextual information which is especially crucial in medical analysis as can be seen here.
Unlike ResNet, which focuses on layer-wise residual learning, U-Net's symmetric expanding path allows for precise localization and reconstruction of fine details from low-quality inputs, making it very good for tasks requiring exact structural replication. Its skip connections, which bridge the encoder directly to the decoder can significantly enhance information flow and help to preserve spatial hierarchies, often lost in deeper networks like deep residual networks.
While transformers are adept at capturing global dependencies, they might neglect in learning local details which in our opinion is a very crucial part in MRI reconstruction.
Additionally, transformers might require substantially higher computational resources, which can be an obstacle to us with our given time and resources, In our opinion U-Net can be more resource-efficient because it can be very adapted to the task.
SRGAN aims to enhance perceptual realism but may compromise clinical accuracy, introducing undesirable artifacts through adversarial training. U-Net minimizes this issue by being trained to closely match the ground truth, thus maintaining high fidelity to original medical images. It can also offer more stable and predictable training compared to the often-challenging training dynamics of GANs.
Custom CNNs can be specifically tailored but require extensive experimentation to optimize, whereas U-Net provides a reliable, proven framework with less need for customization, making it particularly advantageous when working within time constraints. It strikes an optimal balance between extracting deep features and utilizing surface-level information, and might avoid unnecessary complexity and computational burdens as can be partially seen here.

The chosen loss criteria for the GO_RecoNet model is MSE, to ensure both accurate and high-quality MRI reconstructions. MSE loss focuses on pixel-wise accuracy, crucial for detailed medical images. At first, we experienced using both MSE and PSNR as loss functions, combining them through tuning of coefficients. We took the PSNR loss as $\frac{1}{PSNR}$ to be able to minimize it along the MSE. The use of the combined loss criteria did not yield superior results compared to using MSE alone. This can be explained because PSNR is inherently dependent on MSE and using both doesn't provide more information.

Therefore we used only MSE as a loss function.
Overall, we try to ensure that the reconstruction will be both precise and visually pleasing.

In summary, The U-Net architecture was chosen for its effective encoder-decoder structure with skip connections, capturing both high-level and local contextual information, crucial for medical image analysis. Its symmetric expanding path allows for precise localization and reconstruction of fine details, ensuring high fidelity to original images. The skip connections enhance information flow and preserve spatial hierarchies, making U-Net ideal for MRI reconstruction tasks. The MSE loss criteria ensures pixel-wise accuracy and overall visual quality, making the model robust to noise and adaptable to specific requirements, resulting in precise and visually pleasing reconstructions. The network is visualized as follows.

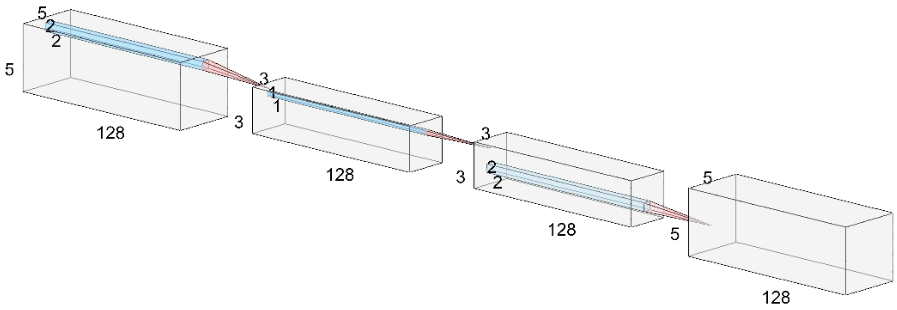
Encoding part:



The initial encoding phase involves the sequential application of an encoder block twice. Each encoder block compromises the following operations:

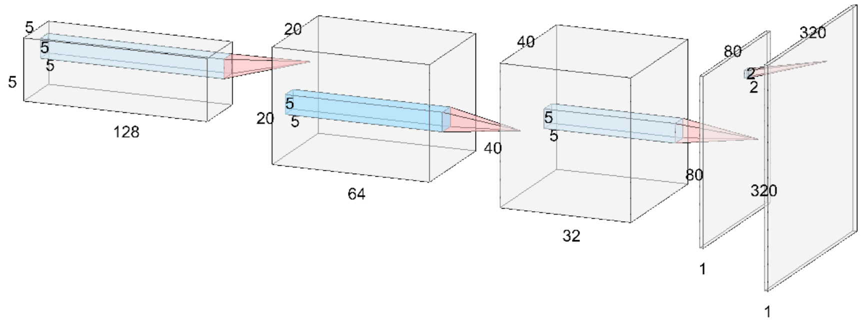
- Convolutional layer: $kernel\ size = 5$, $stride = 2$.
- Normalization layer.
- ReLU activation function.
- Max pooling layer: $kernel\ size = 2$.

Bottleneck:



The input undergoes max pooling, followed by a $1 * 1$ convolution and an interpolation operation. The resulting output is concatenated with the bottleneck input.
Subsequently, a convolutional layer is applied to the concatenated result to reduce the number of features: $256 \rightarrow 128$.

<u>Decoder</u>:



The decoding phase involves the sequential application of a decoder block twice. Each decoder block compromises:

- Up-sampling layer.
- Convolutional layer: $kernel\ size = 5, stride = 2$.
- Normalization layer.
- ReLU activation function.

Following each decoder block, a skip connection is incorporated (not shown in the graph).
Consequently, a convolutional layer is applied after each concatenation to reduce the feature dimension.
Finally, an interpolation operation is performed to produce the final output image.

Results Summary (PSNR - Mean and std) across train and test sets:

- Train set:

Drop rate	Learned mask (Mean $\pm$ Std)	Non-Learned mask (Mean $\pm$ Std)
0.2	32.60985 ± 1.39275	28.64482 ± 1.2367
0.4	31.8118 ± 1.36219	26.48024 ± 1.24633
0.6	30.66749 ± 1.38163	24.07635 ± 1.21372

- Test set:

Drop rate	Learned mask (Mean $\pm$ Std)	Non-Learned mask (Mean $\pm$ Std)
0.2	31.32835 ± 2.03575	27.231 ± 1.63954
0.4	30.64634 ± 1.95406	25.00568 ± 1.57326
0.6	29.35582 ± 1.76597	22.55603 ± 1.48975

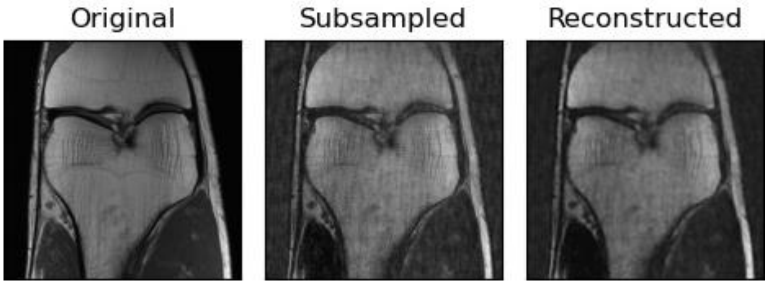
Analysis of the results: As we can see the average PSNR with the learned mask on the train set was ~ 31 and ~ 30 on the test set. Without the learned mask it is ~ 26 and ~ 25 respectively. The standard deviation with the learned mask on the train set is ~ 1.3 and ~ 1.2 on the test set. Without the learned mask it is 1.9 and 1.5 respectively.
Underfitting/Overfitting: An indicator to overfitting would be a relatively high performance on the training set and significantly poorer performance on the test set, as we can see there is only a slight decrease in the average PSNR with or without the learned mask, which is typically a practical scenario so we cannot state an obvious overfitting.
An indicator to underfitting would be getting poor training performance alongside similarly poor test performance. Typically, PSNR values ranges are:

- Below $20dB$ - poor quality.
- $20dB - 30dB$ - fair to good quality.
- $30dB - 40dB$ - good to excellent quality.
- More than $40dB$ – excellent quality.

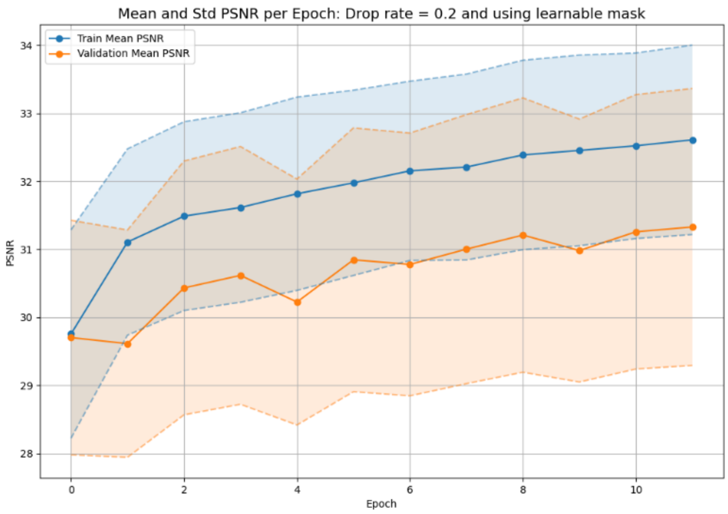
Hence, we would say that the model slightly underfits.

Visualization of results:

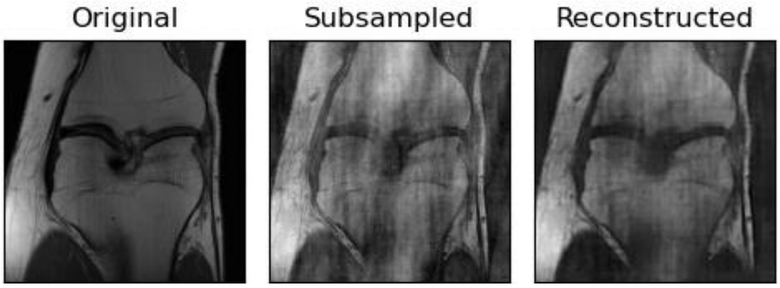
- Drop rate = 0.2 with a learnable mask:



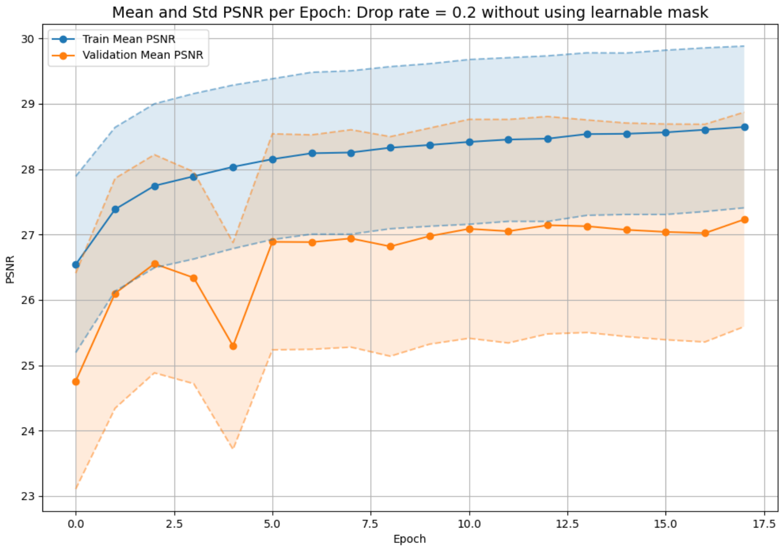
Training graph:



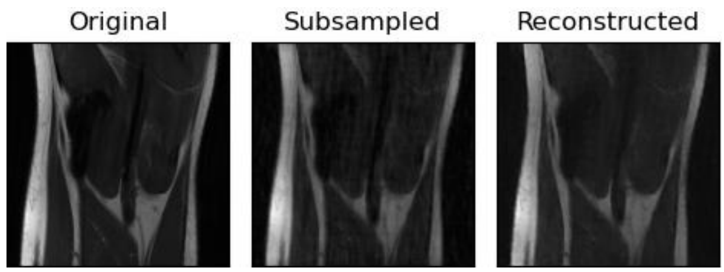
- Drop rate = 0.2 without a learnable mask:



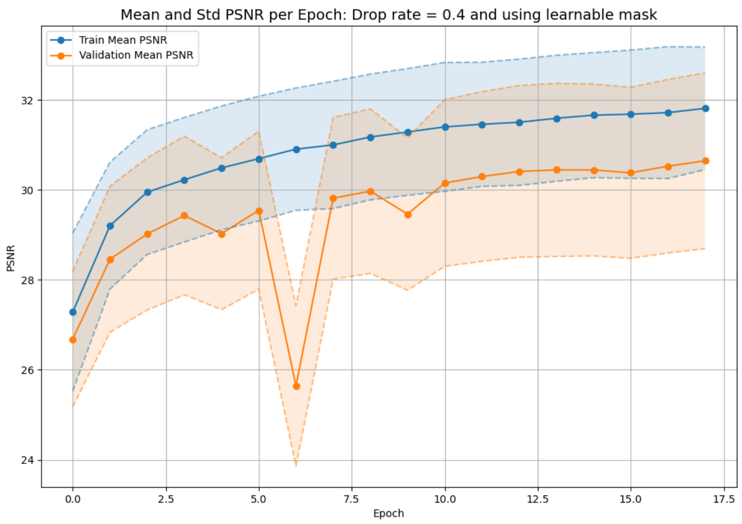
Training graph:



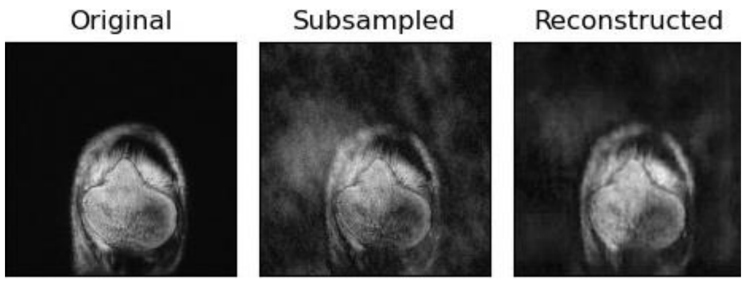
- Drop rate = 0.4 with a learnable mask:



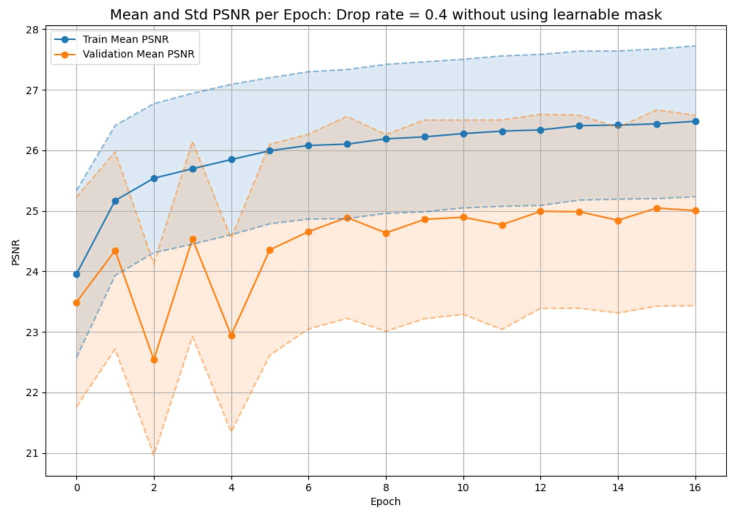
Training graph:



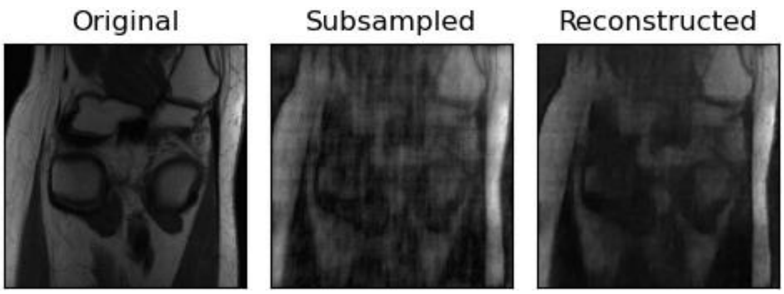
- Drop rate = 0.4 without a learnable mask:



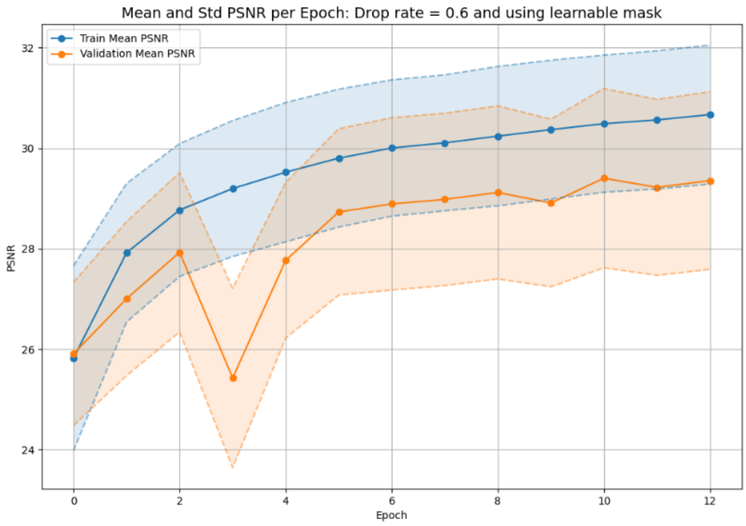
Training graph:



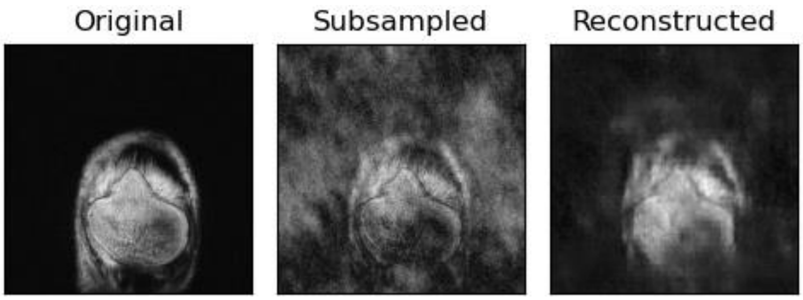
- Drop rate = 0.6 with a learnable mask:



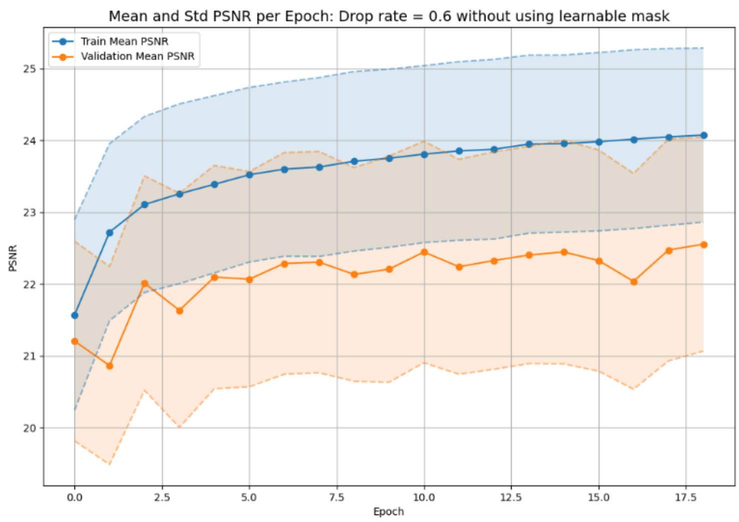
Training graph:



- Drop rate = 0.6 without a learnable mask:



Training graph:



Explanation of the results:

From the results we can conclude that learned masks consistently outperform non-learned masks across all dropout rates in both training and testing datasets. This demonstrated the superiority of the learned mask against the random one. This likely stem from the adaptive selection process of the learned masks, which prioritizes essential features over random dropping.
As the dropout rate increased from *0.2* to *0.6*, the PSNR values decreased for both mask types. The reason for this is obvious, higher drop rate makes the reconstruction more difficult. Also, this decrease was more pronounced in non-learned masks, indicating their lower robustness to data loss.
The architecture's complexity and the optimization techniques like the choice of loss function, using an advanced optimizer, regularization, dropout, batch size and more, all significantly influenced the performance. Effective handling of dropout and experimenting with different loss functions were especially crucial in maintaining high PSNR values for both learned and non-learned masks.

Assuming we had a year to work on this project (Future work) we will suggest experimenting with the following ideas:

1. Transformers – explore the use of vision transformers and Swin transformers to improve global dependency capture. Hybrid Models like U-Net with transformers or U-net with GANs can be used to leverage both structural fidelity and perceptual quality.
2. Hyperparameter tuning can be further researched, we can implement advanced techniques like Bayesian optimization or neural architecture search to find optimal hyperparameters. We can also perform a more extensive cross validation.
3. Improved loss function can be searched, we can conduct experiments with perceptual loss or some variation of multi-scale loss.
4. Mixed precision training may be used to speed up the training process and allow for handling larger batch sizes without compromising performance. Employing curriculum learning techniques, starting with simpler tasks, and gradually increasing complexity, can improve the model's learning efficiency and performance, ensuring it can handle complex MRI reconstruction tasks effectively.
5. Comprehensive testing would be conducted on a wide range of datasets, including different types of MRI scans, to ensure the model's robustness and versatility. Regular benchmarking against state-of-the-art methods would be performed to validate the model's performance and continuously improve its capabilities, ensuring it remains competitive and effective in medical image reconstruction.