Generating details for FastMRI on extremely sparse inputs

Leonid Kulikov
Skolkovo Institute of Science and Technology
Moscow, Russia
leonid.kulikov@skoltech.ru

Ghaith Mqawass
Skolkovo Institute of Science and Technology
Moscow, Russia
ghaith.mqawass@skoltech.ru

Nicholas Babaev
Skolkovo Institute of Science and Technology
Moscow, Russia
nicholas.babaev@skoltech.ru

Anton Antonov
Skolkovo Institute of Science and Technology
Moscow, Russia
anton.antonov@skoltech.ru

Nazerke Sandibay
Skolkovo Institute of Science and Technology
Moscow, Russia
nazerke.sandibay@skoltech.ru

Elmira Volkova
Skolkovo Institute of Science and Technology
Moscow, Russia
elmira.volkova@skoltech.ru

ABSTRACT

Current work is dedicated to recovering not so medically meaningful images to more informative ones. There is an intersection between medicine, signals' processing and deep learning techniques. This field is very promising and is on cutting edge as extra way of getting diagnosis.

Github repo: https://github.com/theleokul/stylish-fastmri

1 INTRODUCTION

Magnetic Resonance Imaging or MRI for short is safe and quite informative way for patients and doctors of getting information about soft tissues of human body as well as functional and structural measurements, which leads to early detection and diagnosis of many diseases.

The main advantage of MRI approach is that it is noninvasive, however it is relatively slow compared to other imaging modalities. The total examination time can vary from 15 to 60 minutes. Patient should stay motionless during that period of time, otherwise different artifacts(tissue- or motion-related) will affect the resulting image and further diagnosis. Moreover, the relatively long scan times lead to high costs that limit the availability of MRI scanners. Therefore, fast acquisition and reconstruction are crucial to improve the performance of current MR scanners, which led in recent years to the development of techniques such as parallel reception, compressed sensing and multi-band accelerations. There are many different types of MR images that an MR scanner can produce. For instance, there are T1-weighted, T2-weighted, PD-weighted, Fluid-Attenuated Inversion Recovery (FLAIR), among others. To make things more complicated, there are sub-types of those types of images (e.g., T1-weighted images come in the flavors: MPRAGE, SPGR, etc.). Depending on the task, this information may be extremely useful because of the unique characteristics of each of these types and sub-types of images.

The acquisition of MRI images is done through consecutive reading-out of single lines in Cartesian k-space. Spatial encoding in MRI involves frequencies and phases, so it is naturally amenable to analysis by Fourier techniques. It was already mentioned, that long acquisition time is intrinsic to the MRI scanner. This period could be shortened by reducing the number of acquired lines in k-space, for example by undersampling the 2D or 3D k-space. However, this can contradict to the Nyquist criterion, resulting in aliasing and blurriness in the reconstructed images, rendering them unqualified for clinical purpose.

2 PROBLEM FORMULATION

In this work we focused on the problem of partially filled scan reconstruction/augmentation. Partially filled scans can be obtained orders of magnitude faster. x4, x8, x16, x32 accelerations were studied. Such acceleration rates don't come at a free cost, many fine-grained details aren't preserved in such scans, therefore the reconstruction algorithm should generate absent details which are consistent with available information.

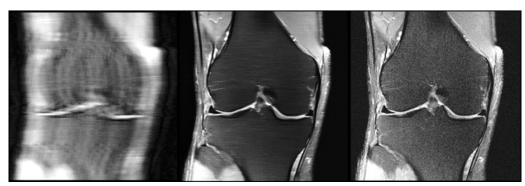

Figure 1: Comparison of reconstructed and fully sampled image

On extreme acceleration rates (x16, x32) acquired scans are extremely sparse and even SOTA methods fail to reconstruct fine details. We came up with the idea to tackle this challenging task and constructed the generative architecture extension for the SOTA approach [6] based on renowed StyleGAN [4].

It is crucial that the generated details has a lot in common with the reality, because of a reconstructed scan usage to address medical issues. For that purpose careful user study should be provided to completely approve or disapprove the solution. We haven't such resources, thus tackled this problem in a qualitative human perceptual sense. Basically we observed the fact that the generative component indeed allows to manifest thin edges while providing textures even for extremely sparse scans (x16, x32 acceleration rates). Though the details aren't so good in medical sense, we have provided the rigorous analysis of the results and explanations in the section with experiments.

3 LITERATURE REVIEW

MAGNETIC Resonance Imaging (MRI) is a widely applied imaging modality, with excellent soft tissue contrast and high spatial resolution. MR images are essential for clinical assessment of soft tissue as well as functional and structural measurements, which leads to early detection and diagnosis of many diseases. However, MRI is relatively slow compared to other imaging modalities. The total examination time can vary from 15 minutes for knee imaging to an hour or more for cardiac imaging. The long acquisition time is intrinsic to the scanner and physics properties of MRI. For the majority of scans performed in clinical practice, this acquisition is done through consecutive reading-out of single lines in k-space. These readouts are constrained by physical limitations of the hardware, the contrast generating principle, and human physiology. The scanning time could be shortened by reducing the number of acquired lines in k-space, i.e. by undersampling the 2D or 3D k-space. However, this could violate the Nyquist criterion, resulting in aliasing and blurriness in the reconstructed images, rendering them unqualified for clinical purpose. Compressed Sensing (CS) and Parallel Imaging are the most common solutions for acceleration by undersampling, while maintaining image quality. Compressed Sensing, the focus of this paper, introduced by [2] and [5], leverages the fact that MR images can be compressed in some domain, restoring the missing k-space data through an iterative reconstruction algorithm [8]. Parallel Imaging uses multiple receive coils that provide an additional signal encoding mechanism, allowing to reduce the number of necessary k-space lines to reconstruct an image, thus partially parallelizing the data acquisition [3].

The FastMRI challenge organized by Facebook led to the development of some state of the art techniques for speeding up the process of MR images reconstruction. [7] developed a modular cross-domain neural network XPDNet and applied it to solve the challenge. This approach is still the best performer in PSNR on the fastMRI leaderboard for both knee and brain at acceleration factor 4. In their model the authors do not use a k-space correction network because it would be very heavy in memory use. However, the authors introduced a refinement network, initially estimated from the lower frequencies of the retrospectively under-sampled coil measurements. This sensitivity maps refiner was a simple U-net model.

Another SOTA work [1] investigated a new approach based on compressed sensing. Authors applied patterned masks on the Fourier transform of the original slices and used the masks proposed by the fastMRI challenge. The model consisted of two main components: Pix2Pix as an image-to-image translation method to correct the artifacts introduced by the k-space undersampling and SRGAN to upscale the low-resolution images to their original size. The results showed SSIM of 0.956 at x16 acceleration factor. However, in our work we want to investigate another model and tackle the problem using different approach.

4 DATASET

The fastMRI dataset consists of two types of scans: knee and brain(neuro). Each type has the training, validation, and masked test sets. Also, multi-coil and single-coil tracks with k-space information are provided. In magnetic resonance imaging, measurements are acquired in the Fourier domain through these coils. So-called k-space is a 2D or 3D complex-valued space representing spatial frequencies in the MRI. Accelerating MRI is a critical medical imaging problem and reducing the number of k-space measurements is a standard way of speeding up the examination time. K-space also allows reconstructing, filtering based on frequency value. Coil measurements are transformed into spatial information using Inverse Fourier Transform. Each element in the k-space matrix represents information for each individual sampled frequency. The spatial location in the matrix with respect to the center describes the direction and the individual spatial frequency component. The magnitude and the angle of the complex number stored represents the amplitude and the phase of the frequency signal, i.e. points in k-space near the center represent low frequency signals whereas those farther from the center represent high frequency signals. It worth noting that low frequency signals make up the contrast in the image, while high frequency signals form the edges. Undersampling low frequency signals reduces the contrast of the image, while undersampling high frequencies reduces edges. The undersampled k-space data are used to construct blurry images.

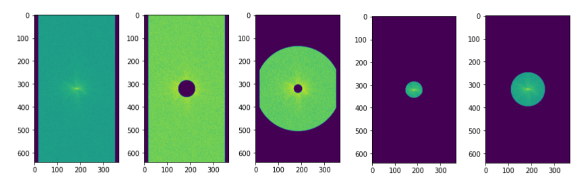

Figure 2: K-spaces

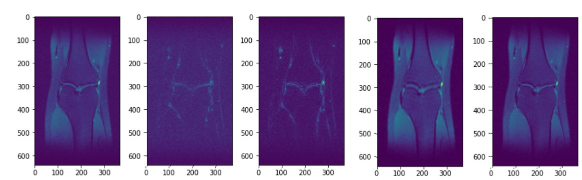

Figure 3: Corresponding images

The resultant images can be reconstructed back using Fourier Transformation. For these purposes built-in functions were used.

5 METRICS

To evaluate our model performance and compare it with the base model we chose SSIM and PSNR metrics. SSIM calculates the Structural Similarity Index between 2 given images which is a value between -1 and +1. A value of +1 indicates that the two images are very similar while the value of -1 indicates that the 2 images are very different. Sometimes these values are adjusted to be in the range of [0,1], while the extremes hold the same meaning. PSNR calculates the Peak Signal-to-Noise ratio between the maximum possible power of an image and the power of corrupting noise that affects the quality of its representation. The higher this index is the better as it means that the quality of the image is high and not so corrupted by noise.

Given that G is the ground truth image and R - reconstruction the metrics equations can be expressed as follows:

$$SSIM(G,R) = \frac{(2\mu_G\mu_R + C_1) + (2\sigma_{GR} + C_2)}{\left(\mu_G^2 + \mu_R^2 + C_1)(\sigma_G^2 + \sigma_R^2 + C_2\right)}$$

where μ is the mean of the corresponding image, σ stands for the standard deviation.C_1, C_2 are some variables to stabilize the division with weak denominator.

$$PSNR(G,R) = 10log_{10}\left(\frac{MAX_G^2}{MSE(G,R)}\right)$$

6 METHODS

6.1 Base model

As a base model we considered an Adaptive-CS-Network architecture[6]. This architecture won the FatsMRI competition in 2019. The architecture is based on a standard for this type of tasks Unet model. It includes data consistency, which can be formulated as:

$$e = F^T(MFx_i - My)$$

In original paper it is defined as "soft data consistency" and allows the network to evaluate the reliability of the acquired data and potentially compensate errors in the coil combination.

6.2 Proposed model

The proposed model is based on the model Adaptive-CS-Network, but has some significant differences.

6.2.1 StyleGan influence. AdaIn operations were added to base model. AdaIn operation is defined as:

$$AdaIN(x_i, y) = y_{s,i}\frac{x_i - \mu(x_i)}{\sigma(x_i)} + y_{b,i}$$

6.2.2 Conditioning. We used conditioning for extracting some meaningful texture features before using the main model. For this were used VAE encoder based on the pretrained ImageNet MobileNetV2 and MappingNet, which consisted of 3 linear layers.

6.2.3 Proxy. In order to tune conditioning component we used Encoder based on the pretrained MobileNetV2 as well. Out goal was to teach conditioning component extract feature vector which are to close to features which were extracted from ground truth images by proxy component.

6.3 Discriminator

The discriminator architecture was taken from Image-to-Image Translation with Conditional Adversarial Networks paper published by berkley AI Research in 2018. It includes standard Conv2d-Batchnorm-leakyReLu blocks with adaptive average pooling in the end.

7 EXPERIMENT

The algorithm was implemented using Pytorch framework by Facebook. All the hyperparameters that were used during training are presented in our github repository.

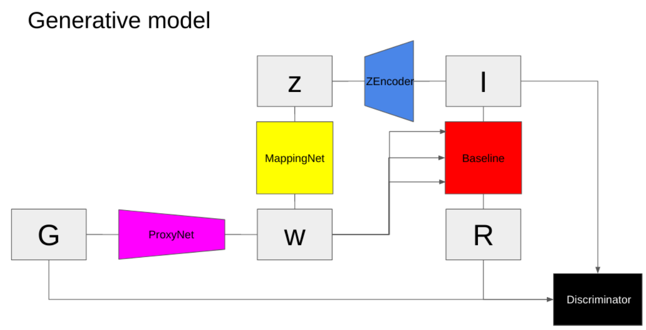

Figure 4: Proposed architecture.

Table 1: Metrics

MODEL	PSNR	SSIM
BASE	17.72	0.31
GENERATIVE	16.71	0.25

We have implemented two models: the base model and the model with the generative extension similar to StyleGAN. Both models were configured in a way to meet 16 Gb memory requirement of our resources and be comparable and trained them for 9 epochs each on single coil knee train dataset by FastMRI. We have applied the default preprocessing used for FastMRI UNet baseline approach except the normalization part which was replaced by simple mapping to zero - one domain.

Loss included several components: reconstruction (l1), adversarial (non-saturating with logits), kl (Kullback–Leibler divergence for the first estimation of texture features) and texture (l1). The texture component measures discrepancy between obtained texture features and proxy texture features which are received from the frozen pretrained on ImageNet MobileNetV2 network. The usage of such proxy is due to the fact that it constructs very expressive features from the images and allows to guide the texture generator part in the direction of effective and concise feature representation. We understand that there is a gap between MRI and natural images domains and accepted such assumption to simplify the training procedure. One can also unfreeze the proxy network to let the network choose the effective feature representation by it's own, though we haven't check that case.

We got the loss curves presented on figures 5, 6. Metrics calculated on validation subset of single coil knee dataset are presented in the table 1. The generative model got lower metrics, yet these metrics are not fully agreed to the human perception. Therefore some visual reconstruction examples are presented on figures 7, **??** as well.

We can see on both examples that the generative model indeed tends to generate some kind of a texture even for very sparse inputs, but the texture obviously isn't of appropriate quality. The texture blobs are too big and almost not correspond to the ground truth. Yet the "styled" reconstructions are definitely more texturized comparing to the base approach.

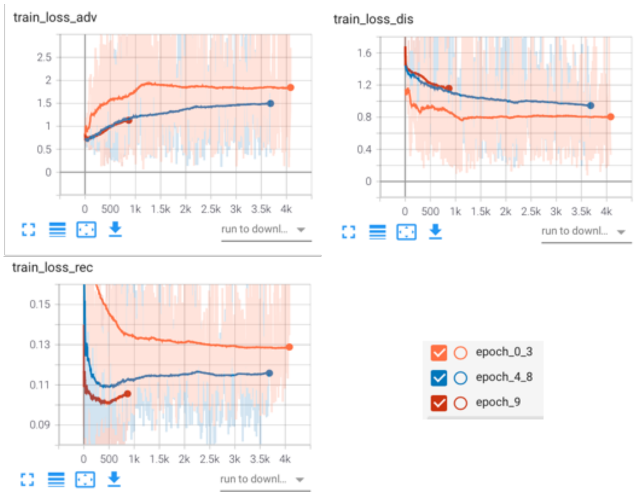

Figure 5: Training losses for the base model. *dis* - non-saturating discriminator loss, *adv* - non-saturating generator loss, *rec* - l1 generator loss.

Figure 6: Training losses for the generative model with StyleGAN-like extension. 9 epochs.

8 RESULTS

In this work we have designed and implemented the new semi-generative architecture for FastMRI. We got sharper images with the generative model comparing to the baseline, yet the quality of the details do not truly correspond to the ground truth. We attribute this discrepancy to the naive approach of texture modelling. Original StyleGAN architecture uses only one vector for the texture which is a potential bottleneck for learning different textures. Also the texture vector is applied globally (through AdaIN layers) to an intermediate tensors in non-spatially adaptive way. Spatial adaptiveness is important here, because different regions may correspond to different textures.

We conclude that this semi-generative design is quiet promising, yet in it's very naive realization doesn't bring us to a feasible quality. Further research of modelling many texture vectors and their spatially-adaptive fusing (e.g. via attention) is required to take out the final verdict.

Figure 7: Example 1. Top to bottom: input, base model, base model with generative extension and groundtruth. Left to right: x4, x8, x16, x32 accelerated reconstructions.

REFERENCES

[1] Aleksandr Belov, Joel Stadelmann, Sergey Kastryulin, and Dmitry V. Dylov. 2021. Towards Ultrafast MRI via Extreme k-Space Undersampling and Superresolution. arXiv:2103.02940 [cs.CV]

[2] D. L. Dohono. 2006. Compressed sensing. *IEEE Transactions on Information Theory* 52, 4 (2006), 1289–1307.

[3] M. B. Scheidegger K. P. Pruessmann, M. Weiger and P. Boesiger. 1999. SENSE: sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine* 42, 5 (1999), 952–961.

[4] Tero Karras, Samuli Laine, and Timo Aila. 2019. A Style-Based Generator Architecture for Generative Adversarial Networks. arXiv:1812.04948 [cs.NE]

[5] D. Donoho M. Lustig. 2007. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine* 58, 6 (2007), 1182–1195.

[6] Nicola Pezzotti, Sahar Yousefi, Mohamed S. Elmahdy, Jeroen van Gemert, Christophe Schülke, Mariya Doneva, Tim Nielsen, Sergey Kastryulin, Boudewijn P. F. Lelieveldt, Matthias J. P. van Osch, Elwin de Weerdt, and Marius Staring. 2020. An Adaptive Intelligence Algorithm for Undersampled Knee MRI Reconstruction. arXiv:2004.07339 [eess.IV]

[7] Zaccharie Ramzi, Philippe Ciuciu, and Jean-Luc Starck. 2020. XPDNet for MRI Reconstruction: an Application to the fastMRI 2020 Brain Challenge. arXiv:2010.07290 [eess.IV]

[8] R. Constable E. Haacke P. Lauterbur Z.-P. Liang, F. Boada and M. Smith. 1992. Constrained reconstruction methods in MR imaging. *Reviews in Magnetic Resonance in Medicine* 4, 2 (1992), 67–185.