Can We See the Bottom? A Look into Deep Learning-Based Image Reconstruction

Dominik Nickel, Ph.D.

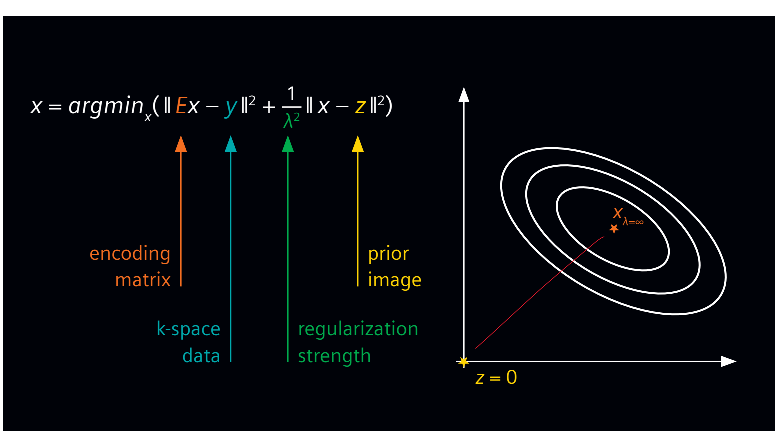
Principal Key Expert for Sequence and Reconstruction Techniques, Siemens Healthineers AG, Erlangen, Germany

In 2016, machine-learning pioneer Geoffrey Hinton suggested that "people should stop training radiologists now. It's just completely obvious within five years deep learning is going to do better than radiologists." Even though this has not happened so far and is unlikely to happen in the foreseeable future, the impact of deep learning on radiology and the way imaging is performed and interpreted is undeniable. For MR imaging (MRI), it has turned out that image reconstruction can benefit hugely from the inclusion of trained components and can generate better images from fewer data. Initial works on deep learning-based image reconstruction started around the time of Hinton's comment [1, 2]. With its capability to push MRI toward higher resolution, higher signal-to-noise, and/or shorter acquisition time, it only took a few years for the technique to transition from technical research into commercially available products that are widely used in clinical routine.

The fast clinical transition combined with the reputation of neural networks for being black boxes understandably raises questions and concerns about the loss or masking of relevant information and the hallucination of "beautiful" but false images. Users would like to have a better understanding of how the technique works, why it enables a significant performance improvement regarding image quality, and where its limitations are. Even though answering these questions about deep learning applications is generally difficult, it is important to be more explicit about the specific type of architecture being referred to. After various approaches were explored, the "physics-based" or "physics-informed" reconstructions have prevailed for generating an image from raw *k*-space data. This class of reconstructions is characterized by the fact that a conventional parallel imaging model is combined with trainable components that take care of image enhancement.

Physics-informed reconstructions

The idea of combining parallel imaging and image enhancement is not new. It is also a central aspect of Compressed Sensing (CS) as used in MRI. For Compressed



1 Illustration of regularized SENSE as a quadratic optimization problem, involving a data consistency term that relates the image x to k-space data y using an encoding matrix E involving sampling mask, Fourier transformation, and coil sensitivity maps. The data consistency term is balanced by a quadratic regularization involving a prior image z that is assumed to be zero. The definition of the regularization strength λ is such that the reconstruction provides images close to the prior for small values and close to the unregularized SENSE reconstruction for high values of λ. Therefore, the trajectory of solutions for varying regularization strength λ starts from the prior image and steps toward the minimum of the data consistency term.

Sensing, the reconstructed image is obtained by iteratively solving an optimization problem which includes a data consistency term based on a parallel imaging model and image regularization terms. Deep learning-based reconstructions (DLR) are often inspired by an explicit, iterative algorithm that is used for this kind of optimization. However, the connection to an optimization problem is removed in DLR, and some processing steps are replaced by trainable components. This is referred to as an unrolled architecture. Most notably, image regularizations are typically replaced by neural networks. As such, Compressed Sensing reconstructions may be considered a special case of a physics-informed deep learning reconstruction as they are in the space of allowed configurations that can be obtained through training. But there is a large space of possible configurations, such that the heuristically designed versions employed in Compressed Sensing are unlikely to be picked.

Once an architecture is specified, the optimal parameters need to be determined. This is where the data-driven aspect of deep learning comes in. The most established approach is supervised training based on fully sampled acquisitions serving as ground truth. For typically thousands of such datasets, a desired target image is generated. Furthermore, inputs that mimic an accelerated acquisition are synthesized. Additional inputs such as coil sensitivity maps needed for the parallel imaging component may be provided. The network parameters are then obtained by established training methods that stochastically optimize the model parameters by comparing the network output to the desired target using a similarity metric or loss function. Training typically lasts multiple days on dedicated GPU servers. Once finished, the obtained configuration or model parameters can be exported for prospective use on a scanner. Prospective execution has a numerical demand similar to an iterative reconstruction and is usually done on integrated GPUs.

A conceptionally intuitive network architecture

To gain insight into the performance of deep learning-based image reconstruction, an architecture that allows direct connection to parallel imaging and image-based denoising is presented in the following.

Starting off with conventional parallel imaging, a regularized SENSE [3] reconstruction is illustrated in Figure 1. Already in this reconstruction, an additional regularization term is introduced that allows to condition the inversion of the linear problem when the parallel imaging acceleration becomes too large for the available coil configuration.

At this point, it is worth mentioning the *g*-factor, which is often used to describe the additional local noise amplification in parallel imaging. It is sometimes suggested that the *g*-factor is purely determined by the acquisition and its acceleration. This is not correct because, for instance, a zero-padded Fourier transformation corresponds to ideal noise propagation and a *g*-factor of '1'. Of course, this is with unacceptable aliasing, but artifacts are not reflected in the *g*-factor. Any practical parallel imaging reconstruction therefore includes a regularization that helps to limit the noise amplification. The definition of the regularization strength λ in Figure 1 is such that it is also an upper bound on the noise amplification or *g*-factor. In that sense, it is interpretable, which will become relevant below. At the same time, λ is the step size on the trajectory from the prior image toward the unregularized SENSE reconstruction. Figure 2 shows a comparison of two reconstructions with and without regularization using aggressive retrospective acceleration. It is apparent that a regularization is required to reduce noise amplification, i.e., to restrict the *g*-factor.

Given a test dataset that is later also used for more involved networks, a first experiment is to evaluate metrics between regularized SENSE reconstructions and expected targets from the fully sampled acquisition as a function of

2 Unregularized SENSE reconstruction **(2A)** and regularized SENSE reconstruction with $\lambda = 4$ **(2B)** from retrospectively subsampling a fully sampled acquisition **(2C)**. The retrospective acceleration factor is 2 × 2 in AP and HF direction, deliberately chosen to be too aggressive for the given coil setup.

the regularization strength. This is presented in Figure 3. Interestingly, the best results are not achieved for the unregularized SENSE, but depending on acceleration and chosen metric for regularization strengths on the order of 5. Determining the optimal regularization this way may already be considered a machine learning reconstruction – possibly the simplest conceivable one, and with unimpressive performance.

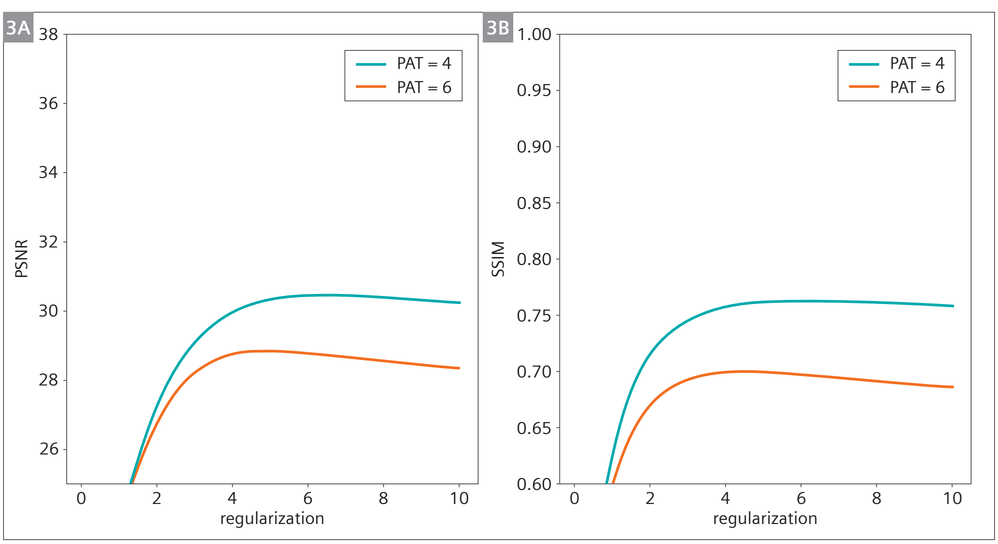
For deep learning-based reconstruction, image-to-image networks are typically included for the enhancement of intermediate images. In the simplest case, this may be a denoising network after a parallel imaging reconstruction. A more general approach that alternates between parallel imaging reconstruction and deep learning-based enhancement is depicted in Figure 4.

The basic idea is to have regularized SENSE reconstructions where the regularization strength steps from a current prior image toward the unregularized SENSE reconstruction. The step sizes are then determined through training, and the network estimates the next prior image from the result of the previous SENSE reconstruction. So the task of the network appears constrained and the conventional parallel imaging reconstruction is a subset of the training architecture. Another advantage is that there is no requirement to have a larger number of iterations to ensure convergence. This is ensured by using the regularized SENSE reconstruction as an update mechanism.

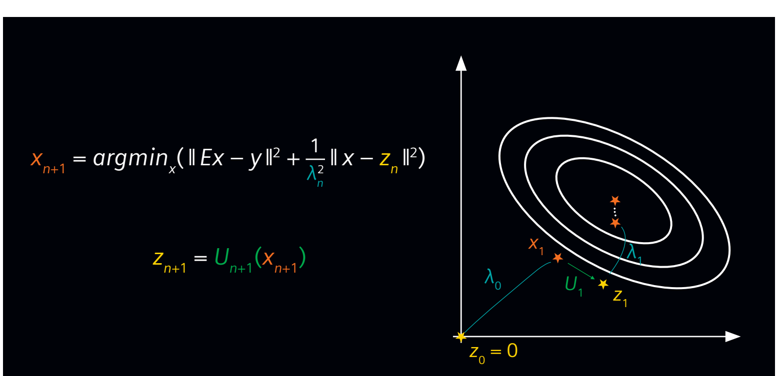
To evaluate the potential of such an architecture, an unrolled network with 6 iterations and conventional U-nets [4] was set up and trained using about 5,000 training pairs from about 500 acquisitions. On the same test datasets used in Figure 3, the finally obtained reconstruction achieved PSNR values of 37.7 dB and 36.5 dB, as well as SSIM values of 0.93 and 0.92 for a retrospective acceleration of 4 and 6, respectively. This outperforms the conventional reconstruction drastically. In Figure 5, an example image is shown on the left, which visually comes close to the ground truth.

Most notably, the noise amplification in the center of the image and associated with an increased *g*-factor seems to be absent. This is a rather general behavior observed for unrolled, deep learning-based reconstructions. Some insight can be obtained when looking at the regularization strengths at each iteration. These are stated in Table 1.

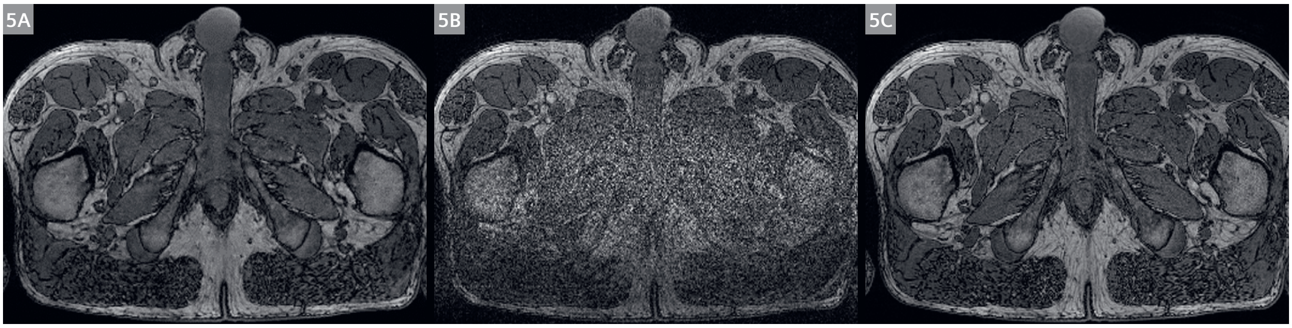
It is apparent that the training leads to a parameterization where the first iterations perform a weakly regularized parallel imaging reconstruction whose output is taken to estimate the following prior using a neural network.



3 Average peak signal-to-noise ratio (**3A**, PSNR) and structural similarity index measure (**3B**, SSIM) of test datasets with retrospective subsampling with acceleration 4 (petrol) and 6 (orange) as a function of the regularization strength.



4 Illustration of an unrolled deep learning-based reconstruction that starts from a vanishing prior imaging ($z_0 = 0$) and then performs iterations comprising a regularized SENSE reconstruction followed by the estimation of a new image prior using a neural network U_n. The number of iterations is fixed. The regularization strengths for each iteration and the model parameters of the neural networks form the trainable parameters. The output of the final network is taken as the reconstructed image. The diagram on the right sketches a trajectory of the iterations through the image domain with indicated ellipsoidal hypersurfaces of the data consistency term.



5 Prediction of an unrolled network with 6 iterations **(5A)**, and regularized SENSE reconstruction with λ = 4 **(5B)** from retrospectively subsampling a fully sampled acquisition **(5C)**.

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
7.05	7.05	3.91	1.85	1.11	0.91

Table 1: Regularization strengths for each iteration of a trained network with 6 iterations.

Therefore, the reconstruction quickly comes close to solutions with acceptable data consistency. In the latter iterations, the steps toward the unregularized SENSE solution become smaller. Here, the conditioning of the data consistency starts to become more relevant. Areas with small noise propagation are also better conditioned, such that the step sizes affect the image more and make sure that the image is consistent with the data consistency. At the same time, this also restricts the solution in regimes with worse conditioning, and the image regularization has more influence there. Therefore, the last iterations mostly help to improve the image quality in the less well-conditioned regions, and the noise appearance in the final reconstruction is much more homogenous.

The presented behavior therefore gives a reasonable explanation for the significantly improved performance of unrolled deep learning-based reconstructions compared to pure post-processing approaches. Alternating between data consistency based on a parallel imaging model and deep learning-based image enhancement allows to iteratively optimize the output of the reconstruction. This is also a reasonable explanation for why the approach generalizes surprisingly well, or rather why it is less dependent on the image content seen during training, as the architecture is highly incentivized to be consistent with a physical model with freedom mostly for the ill-conditioned domains. This freedom seems to be in local image enhancement such as denoising and edge preservation, meaning that topics like hallucination have so far not surfaced for physics-informed deep learning-based image reconstructions. The magical aspect that remains unexplained is the performance of the individual neural networks for image enhancement. This is a general aspect of deep learning and is outside of the scope of this paper.

Validation of deep learning-based image reconstruction

Even though the performance of deep learning-based image reconstruction can be explained to some extent, it needs to be validated with care. Currently, the transition into clinical routine is focused on improving established sequence types and contrasts. Therefore, clinical evaluations that use qualitative and quantitative metrics for comparison to the established clinical sequences are typically used to establish evidence [5].

References

1 Hammernik K, Klatzer T, Kobler E, Recht MP, Sodickson DK, Pock T, et al. Learning a Variational Network for Reconstruction of Accelerated MRI Data.
Magn Reson Med. 2018;79(6):3055–3071. Epub 2017 Nov 8.

2 Schlemper J, Caballero J, Hajnal JV, Price AN, Rueckert D. A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction.
IEEE Trans Med Imaging. 2018;37(2):491–503. Epub 2017 Oct 13.

3 Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI.
Magn Reson Med. 1999;42(5):952–62.

4 Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Navab, N, Hornegger, J, Wells, W, Frangi, A (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III. Springer International Publishing. pp. 234–241.

5 Afat S, Ursprung S, Herrmann J, Brendlin A, Werner S, Lingg A, et al. Deep Learning-based reconstruction in clinical MRI – How we do it in Tübingen. MAGNETOM Flash. 2024;87(2):10–16.

Contact

Dominik Nickel, Ph.D.
Siemens Healthineers AG
SHS DI MR RCT CLS BODY
Karl-Schall-Str. 6
91052 Erlangen
Germany
marcel.nickel@siemens-healthineers.com