💬 COMMENTARY

# Subtle pitfalls in the search for faster medical imaging

Kai Tobias Block[a,1] 🔵

The rise of machine-learning (ML) techniques over the last decade has impacted almost every scientific field to some extent. Medical imaging and radiology are no exception. Since radiologists are, in essence, recognizing patterns in medical images, it seems reasonable to apply ML models for supporting—or even fully automating—such image interpretation tasks. However, while the vision of using ML in radiology continues to spark excitement, clinical real-world applications turned out to be more challenging than initially anticipated and, up to now, ML techniques for automated disease detection have not found the way into routine practice. Nevertheless, for specific subproblems, ML techniques have proven to be very powerful.

One prominent example is the segmentation of structures on medical images, a problem that has led to the widely used U-Net model. Another example is image reconstruction, i.e., the formation of images from the raw signals that imaging devices acquire. Seminal papers from multiple groups have shown that ML models can be utilized to improve both the visual image quality and diagnostic value (1), meaning that pathologies can be better identified. Motivated by the high excitement about ML, this has drawn many researchers to work on the reconstruction problem, including members from the medical-imaging field as well as from adjacent disciplines. Within a short time, publications on ML-based image reconstruction have skyrocketed, and several dedicated conferences have been organized.

Such enthusiasm, on the one hand, has been a blessing for our field because it brings many fresh ideas and different thinking into medical imaging. On the other hand, it has also started a competitive race for the "best" reconstruction model, measured by various metrics for the reconstruction accuracy. This has led to an increasing number of works that claim exceptional reconstruction performance based on conducted numerical experiments. However, when looking closely at the methodology, it often turns out that the datasets utilized for validation do not accurately reflect the signals received by imaging devices. Hence, such results look impressive on paper, but they do not hold true when implementing the techniques for real-world use.

In many cases, validation datasets have been synthesized from publicly available data sources such as image repositories, which inherently introduces assumptions and simplifications of the signal-generation process. This is commonly referred to as "inverse crime" because the same approximations are used when synthesizing and when reconstructing the data. Consequently, it remains unnoticed that an algorithm may perform much worse if used with real-world data. While this topic has been the subject of numerous discussions at conferences when presenters showed "inverse crime," the impact on the estimated reconstruction performance has typically been described only in qualitative or anecdotal terms.

In PNAS, Shimron et al. (2) present a comprehensive investigation of this effect, in which they have analyzed multiple scenarios of improper use of validation data generated from image libraries. This shows, in a quantitative and very illustrative manner, how large of a mistake one can make if not paying enough attention to the evaluation of reconstruction algorithms, and they reveal two concrete examples of how easily incorrect conclusions can be drawn by accidentally utilizing image libraries that have undergone postprocessing steps that are not directly visible to the human eye, namely zero-padding interpolation in Fourier space and JPEG image compression.

Shimron et al. (2) selected magnetic resonance imaging (MRI) as a representative example in their paper, one of the most widely used and most fascinating modalities in radiology. Unfortunately, MRI is also one of the most complex imaging techniques, which sometimes leads to misconception of the acquired imaging signal. To put Shimron et al.'s work into context, it is helpful to review how the imaging data are generated.

## Signal Generation in MRI

In MRI, the patient is positioned in a strong magnetic field (1.5 or 3 T). By creating a short time-varying electromagnetic field using a waveform generator and radiofrequency (RF) coil that surrounds the patient, protons of the tissue get excited to a higher energy state if the frequency of this "RF pulse" coincides with the Larmor frequency, $\omega = \gamma \cdot B$, which is proportional to the magnetic field strength B ($\gamma$ is the gyromagnetic ratio of protons). After stopping the RF pulse, a signal can be measured (also with Larmor frequency) that is induced in the RF coil by the protons as response to the prior excitation. The phenomenon is called "magnetic resonance" and is only detectable for a short moment until the protons relax back to the lower energy state. This procedure is repeated sequentially in MRI to obtain signals from the tissue. However, this, by itself, would not allow looking inside the body.

Spatially resolving the signal is possible by switching an additional magnetic field (after the RF excitation) that creates a linear variation of the magnetic field strength in one spatial

direction. This additional field is called "gradient field" (due to the linear relationship between space and field strength), and it can be generated in the X, Y, and Z direction using three electrical coils built into the MRI device. Because the frequency of the induced MR signal is proportional to the field strength that the protons experience, the signal that the protons induce in the RF coil suddenly becomes dependent on the location of the protons. For example, when switching a gradient field in the Z direction, protons at the location $z = -10$ cm induce a lower frequency than protons at $z = 0$ cm, while protons at $z = +10$ cm induce a higher frequency. By analyzing which frequencies are contained in the received RF signal, it is then possible to conclude from which locations these signal components originated. This can be achieved by performing a Fourier transformation (or spectral analysis) of the received signal. In other words, by switching gradient fields that create a linear relationship between space and resonance frequency, it becomes possible to measure the Fourier transform. Nobel Prize winner Paul Lauterbur proposed this brilliant idea in 1973 (3).

The acquired MRI signal, therefore, corresponds to the patient's Fourier transform, which can be traversed by switching gradient fields with varying amplitude along X, Y, and Z. When switching the gradients in such order that the Fourier space (or "k-space") is sampled on a grid, an image can be reconstructed by digitizing the signal, arranging the values on a matrix, and calculating the fast Fourier transform (FFT).

One limitation of the MRI principle is that traversing the Fourier space takes time because the gradient fields cannot be switched instantaneously. In the initial years, device manufacturers have therefore strived to achieve faster and faster switch times using more powerful amplifiers and improved coil designs. However, eventually a limit was reached where further acceleration would put the patient at risk, as rapidly changing magnetic fields can trigger unpleasant and potentially dangerous nerve stimulation.

## Faster Speed with Smart Reconstruction

At this point, activities for MRI acceleration shifted to the image-reconstruction side. The idea is to skip sampling steps in Fourier space, resulting in shorter scan time but incomplete datasets, and to compensate for the missing data using intelligent reconstruction algorithms. This explains the strong interest in applying ML techniques for the reconstruction problem: Advanced algorithms can not only create "nicer-looking" images but they can make MRI faster, lower examination costs, and create a better patient experience because the time during which patients must hold still or hold breath is shortened.

Reconstruction from incomplete data is possible by incorporating "a priori" information into the algorithm, i.e., by taking advantage of assumptions that can be made upfront about the image. Mathematically, this is done by formulating the reconstruction as an inverse problem (instead of direct FFT use) and by introducing regularization or penalty terms that steer the algorithm toward a solution that is plausible based on the a priori information. Thus, the aim is to find an image that fits to the acquired (but incomplete) Fourier samples and that minimizes the penalty functions. A solution

can then be found using iterative numerical optimization, e.g., with the conjugate-gradient method.

The first generation of such methods, known as compressed sensing (4), used hand-picked and rather simplistic functions for the a priori knowledge. One example is the total variation (TV), implying that medical images are piecewise constant to some degree. Another example is the wavelet transform, implying that medical images are wavelet-compressible (whereas artifact patterns caused by incomplete sampling are not). These first-generation approaches are now available in clinical MRI devices.

Recently, it has been shown that ML techniques can be used to train and learn regularization functions instead of relying on basic hand-picked functions. It makes sense that training a high-capacity ML model on a large amount of reference data improves the ability to distinguish spurious imaging artifacts (such as aliasing or noise patterns) from the true object, and initial results have been very promising (1). This new generation of algorithms also introduces additional degrees of freedom, e.g., by allowing the imposed prior knowledge to vary during the reconstruction process.

However, many questions are still unanswered, including what the best model architecture for such application is and to what extent the learned prior knowledge generalizes, i.e., if a single trained model can be applied for different applications or if separate models need to be trained for each body region, examination type, or device type. This makes image reconstruction a highly interesting research topic, and it is not surprising that numerous groups have decided to work on the problem.

## The Simulation Pitfall

Unfortunately, some research groups do not have access to MRI devices, typically if they are not affiliated with a radiology department, which makes it difficult to get hold of raw MRI data. Such datasets are rarely shared because they can contain confidential patient information. As a workaround, data are often simulated, both for the model training and evaluation. Here, a common pitfall is to generate simulation data by taking images (e.g., from public libraries), performing an inverse FFT, and using the values as a substitute for the MRI signals.

While MRI devices, indeed, measure the Fourier transform, the devil is in the details. As part of the devices' processing pipeline, many additional (and nonlinear) operations are applied before final images are exported, including filters for distortion correction, intensity normalization, and noise reduction. Moreover, MRI data are complex-valued (i.e., objects have a locally varying phase), but only the magnitude value is exported. Modern MRI devices also use multiple receive RF coils with different complex-valued sensitivity profiles, which are combined during the reconstruction. Such properties are neglected when synthesizing data from images. Because the acquisition is not instantaneous, real MRI signals can also be affected by data inconsistencies from motion (e.g., the beating heart), blood flow, and signal relaxation. Furthermore, physical constraints exist for the acquisition. For example, it is not possible to "jump" in Fourier space (it must be traversed along trajectories). There is also a

"memory" effect of the signal because the acquisition is often faster than the protons' relaxation time, making the sampling order relevant. Finally, the digitalization of the signal, i.e., the discrete sampling of the continuous Fourier transform, plays a role as well because it leads to effects such as Gibbs ringing and periodic copies in image space.

If these aspects are ignored, oversimplified simulation data are created that do not reflect the complexity (and impurity) of real-word MRI data. When additionally using data sources that have undergone postprocessing steps to reduce the information content, as illustrated in Shimron et al.'s paper (2), reconstruction results are obtained that are biased and just too good to be true. Therefore, the work of Shimron et al. should be seen as a wake-up call to be vigilant when evaluating new MRI reconstruction methods. Various helpful resources exist. Several initiatives have now published libraries with anonymized real-word MRI data, such as the fastMRI project (5), and multiple Bloch simulation tools have been released (6), which can simulate the MRI signals accurately. Such offerings should be taken advantage of to prevent further "inverse crime."

1. K. Hammernik *et al.*, Learning a variational network for reconstruction of accelerated MRI data. *Magn. Reson. Med.* **79**, 3055–3071 (2018).
2. E. Shimron, J. I. Tamir, K. Wang, M. Lustig, Implicit data crimes: Machine learning bias arising from misuse of public data. *Proc. Natl. Acad. Sci. U.S.A.* **119**, 10.1073/pnas.2117203119 (2022).
3. P. C. Lauterbur, Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature* **242**, 190–191 (1973).
4. M. Lustig, D. Donoho, J. M. Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.* **58**, 1182–1195 (2007).
5. F. Knoll *et al.*, fastMRI: A publicly available raw k-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning. *Radiol. Artif. Intell.* **2**, e190007 (2020).
6. Z. Cao *et al.*, Bloch-based MRI system simulator considering realistic electromagnetic fields for calculation of signal, noise, and specific absorption rate. *Magn. Reson. Med.* **72**, 237–247 (2014).