Article

Patterns

Stabilizing deep tomographic reconstruction: Part A. Hybrid framework and experimental results

Highlights

- Deep reconstruction solution to the instabilities identified by a recent *PNAS* paper
- Analytic compressed iterative deep (ACID) method is for hybrid reconstruction
- ACID framework combines benefits from deep learning and compressed sensing
- ACID is accurate and stable, superior to sparsity-regularized reconstruction alone

Authors

Weiwen Wu, Dianlin Hu, Wenxiang Cong, ..., Hengyong Yu, Varut Vardhanabhuti, Ge Wang

Correspondence

hengyong-yu@ieee.org (H.Y.), varv@hku.hk (V.V.), wangg6@rpi.edu (G.W.)

In brief

We propose an analytic compressed iterative deep (ACID) framework for accurate yet stable deep reconstruction. ACID synergizes a deep reconstruction network trained on big data, kernel awareness from compressed sensing-inspired processing, and iterative refinement to minimize the data residual relative to real measurement. We anticipate that this integrative model-based data-driven approach will promote the development and translation of deep tomographic image reconstruction networks.

Wu et al., 2022, Patterns *3*, 100474
May 13, 2022
https://doi.org/10.1016/j.patter.2022.100474

CellPress

Patterns

CellPress
OPEN ACCESS

Article

Stabilizing deep tomographic reconstruction: Part A. Hybrid framework and experimental results

Weiwen Wu,[1,2,6] Dianlin Hu,[3] Wenxiang Cong,[1] Hongming Shan,[1,4] Shaoyu Wang,[5] Chuang Niu,[1] Pingkun Yan,[1] Hengyong Yu,[5,7,*] Varut Vardhanabhuti,[6,*] and Ge Wang[1,*]

[1]Biomedical Imaging Center, Center for Biotechnology and Interdisciplinary Studies, Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA
[2]School of Biomedical Engineering, Sun Yat-sen University, Shenzhen, Guangdong, China
[3]The Laboratory of Image Science and Technology, Southeast University, Nanjing, China
[4]Institute of Science and Technology for Brain-inspired Intelligence, Fudan University, Shanghai, China
[5]Department of Electrical & Computer Engineering, University of Massachusetts Lowell, Lowell, MA, USA
[6]Department of Diagnostic Radiology, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong SAR, China
[7]Lead contact
*Correspondence: hengyong-yu@ieee.org (H.Y.), varv@hku.hk (V.V.), wangg6@rpi.edu (G.W.)
https://doi.org/10.1016/j.patter.2022.100474

THE BIGGER PICTURE Tomographic image reconstruction with deep learning has been a rapidly emerging field since 2016. Recently, a PNAS paper revealed that several well-known deep reconstruction networks are unstable for computed tomography (CT) and magnetic resonance imaging (MRI), and, in contrast, compressed sensing (CS)-inspired reconstruction methods are stable because of their theoretically proven property known as "kernel awareness." Therefore, for deep reconstruction to realize its full potential and become a mainstream approach for tomographic imaging, it is critically important to stabilize deep reconstruction networks. Here, we propose an analytic compressed iterative deep (ACID) framework to synergize deep learning and compressed sensing through iterative refinement. We anticipate that this integrative model-based data-driven approach will promote the development and translation of deep tomographic image reconstruction networks.

Proof-of-Concept: Data science output has been formulated, implemented, and tested for one domain/problem

SUMMARY

A recent PNAS paper reveals that several popular deep reconstruction networks are unstable. Specifically, three kinds of instabilities were reported: (1) strong image artefacts from tiny perturbations, (2) small features missed in a deeply reconstructed image, and (3) decreased imaging performance with increased input data. Here, we propose an analytic compressed iterative deep (ACID) framework to address this challenge. ACID synergizes a deep network trained on big data, kernel awareness from compressed sensing (CS)-inspired processing, and iterative refinement to minimize the data residual relative to real measurement. Our study demonstrates that the ACID reconstruction is accurate, is stable, and sheds light on the converging mechanism of the ACID iteration under a bounded relative error norm assumption. ACID not only stabilizes an unstable deep reconstruction network but also is resilient against adversarial attacks to the whole ACID workflow, being superior to classic sparsity-regularized reconstruction and eliminating the three kinds of instabilities.

INTRODUCTION

Medical imaging plays an integral role in modern medicine and has grown rapidly over the past few decades. In the United States, there are more than 80 million computed tomography (CT) scans and 40 million magnetic resonance imaging (MRI) scans performed yearly.[1,2] In a survey on medical innovations, it was reported that "the most important innovation by a

considerable margin is magnetic resonance imaging (MRI) and computed tomography (CT)."[3] Over the past several years, deep learning has attracted major attention in medical imaging. Since 2016, deep learning has been gradually adopted for tomographic imaging, known as deep tomographic imaging.[4–9] Traditionally, tomographic reconstruction algorithms are either analytic (i.e., closed-form formulation) or iterative (i.e., based on statistical and/or sparsity models). Very recently with deep tomographic imaging, reconstruction algorithms have used deep neural networks (i.e., data driven).[10–12] This new type of reconstruction algorithm has generated tremendous excitement and promising results in many studies. Some examples are included in the recent review articles by Wang et al. and Chen et al.[13,14]

While many researchers are devoted to catching this new wave of tomographic imaging research, there are concerns about deep tomographic reconstruction, with the landmark paper[15] by Antun et al. as the primary example. Specifically, Antun et al. performed a systematic study[15] to reveal the instabilities of a number of representative deep tomographic reconstruction networks, including AUTOMAP.[16] Their study demonstrates three kinds of network-based reconstruction vulnerabilities: (1) tiny perturbations on the input generating strong image artefacts (potentially, false positivity); (2) small structural features going undetected (false negativity); and (3) increased input data leading to decreased imaging performance. These critical findings are warnings and at the same time opportunities of deep tomographic imaging research. Importantly, the study by Antun et al.[15] found that small structural changes (e.g., a small tumor) may not always be captured in the images reconstructed by the deep neural networks, but standard sparsity-regularized methods can capture these pathologies. It is worth noting that the issue of missing pathologies was one of the main concerns raised by radiologists in the fastMRI challenge in 2019.[17]

Historically, a debate, challenge, or crisis typically inspired theoretical and methodological development. In the context of tomographic imaging, there are several such examples. In the earliest days of CT reconstruction, analytic reconstruction received a critique that given a finite number of projections, tomographic reconstruction is not uniquely determined, meaning that ghost structures can be reconstructed, which do not exist in reality but are consistent with the measured data.[18] Then, this problem was solved by regularization, such as enforcing the band limitedness of the underlying signals.[19] Iterative reconstruction algorithms were initially criticized that image reconstruction was strongly influenced by penalty terms; in other words, what you reconstruct could be what you want to see. After selecting regularization terms and fine-tuning hyperparameters, these shortcomings were addressed. Hence, such algorithms have been made into clinical applications.[20,21] As far as compressed sensing (CS) is concerned, the validity of this theory is based on restricted isometry or robust null space properties.[22,23] The correct sparse solution will most likely be obtained under the assumed properties. However, these restricted isometry properties may not always be valid or verified in various applications such as few-view CT and fastMRI. In these cases, heuristically designed sampling patterns and empirically adjusted sampling parameters are often used to approximate an ideal random matrix-based data acquisition scheme so that a collected dataset is sufficiently informative.[24] In practice, encouraging results were widely reported in these relaxed applications of CS theory. Nevertheless, the sparsity constraint could be either too strong and smear features or too weak and result in artefacts. For example, a tumor-like structure could be introduced, and pathological vessels may be filtered out if the total variation is overly minimized, as demonstrated in purposely designed numerical examples.[25] Despite the limitations, multiple sparsity-promoting reconstruction algorithms are used on commercial scanners, with excellent overall performance.

The emerging deep tomographic imaging methods encounter challenges, as reported by Antun et al.[15] In addition to extensive experimental data showing the instabilities of several deep reconstruction networks, Gottschling et al. pointed out that these instabilities are fundamentally associated with the lack of kernel awareness[26] and are "nontrivial to overcome."[15] However, their experiments show that CS-inspired reconstruction algorithms worked stably, while their selected deep reconstruction network failed under the same conditions,[15] since CS-based algorithms use sparse regularization that has "at its heart a notion of kernel awareness."[26]

This article focuses on the feasibility and principles of accurate and stable deep tomographic reconstruction, demonstrating that deep reconstruction networks can be stabilized in a hybrid model with a CS module embedded and are superior to CS-based reconstruction alone. Specifically, to overcome the instabilities of the deep reconstruction networks, here, we propose an analytic compressed iterative deep (ACID) framework illustrated in Figure 1A. Given deep reconstruction network Φ and measurement data $\boldsymbol{p}^{(0)}$, an image can be, first, reconstructed, but it may miss fine details and introduce artefacts. Second, a CS-inspired module Θ enforces sparsity in the image domain,[27] with a loss function covering both data fidelity and sparsity (e.g., total variation,[28] low-rank,[29] dictionary learning[30]). Third, the forward imaging model projects the current image to synthesize tomographic data, which is generally different from the original data $\boldsymbol{p}^{(0)}$. The discrepancy is called a data residual that cannot be explained by the current image. From this data residual, an incremental image is reconstructed with the deep reconstruction network Φ and used to modify the current image aided by the sparsity-promoting CS module Θ. This process can be repeated to prevent losing or falsifying features. As a meta-iterative scheme, the ACID reconstruction process cycles through these modules repeatedly. As a result, ACID finds a desirable solution in the intersection of the space of data-driven solutions, the space of sparse solutions, and the space of solutions subject to data constraints, as shown in Figure 1B. Because this integrative reconstruction scheme is uniquely empowered with data-driven prior, ACID would give a better solution than the classic sparsity-regularized reconstruction alone; for details, see the method details section.

An important question is whether the ACID iteration will converge to a desirable solution in the above-described intersection of the three spaces (Figure 1B). The answer to this question is far from trivial. A deep learning network represents a non-convex optimization problem, which remains a huge open problem (see more details in the review by Danilova et al.[31]). The non-convex optimization problem in a general setting is of non-deterministic polynomial-time hardness (NP hardness). To

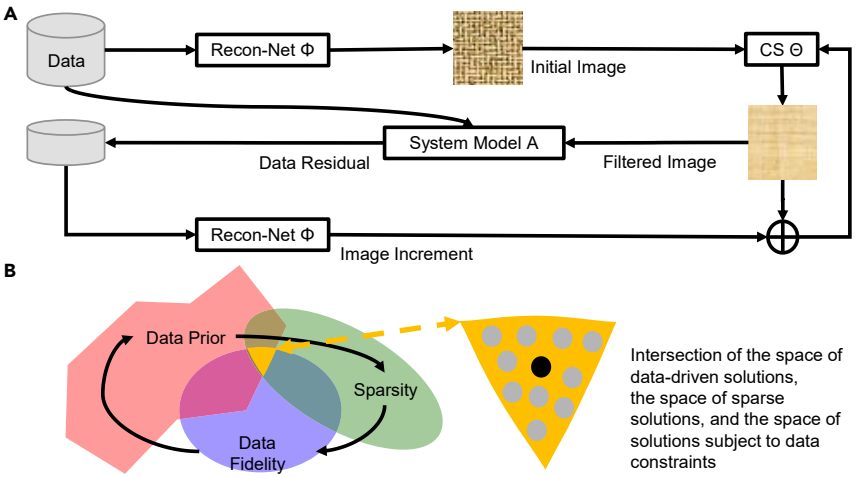


Figure 1. ACID architecture for stabilizing deep tomographic image reconstruction

(A) Initially, the measurement data are reconstructed by the reconstruction network Φ. The current image is sparsified by the CS-inspired sparsity-promoting module Θ (briefly, the CS module). Tomographic data are then synthesized based on the sparsified image according to the system model $\boldsymbol{A}$, and compared to the measurement data to find a data residual. The residual data are processed by the modules Φ and Θ to update the current image. This process is repeated until a satisfactory image is obtained.

(B) Illustration of the solution space.

solve this problem with guaranteed convergence, practical assumptions must be made in almost all of the cases. These assumptions include changing a non-convex formulation into a convex formulation under certain conditions, leveraging a problem-specific structure, and seeking only a local optimal solution. Specifically, the Lipschitz continuity is a common condition used to facilitate performing non-convex optimization tasks.[32,33]

Given the theoretically immature status of the non-convex optimization, to understand our heuristically designed ACID system in terms of its convergence, we assume that a well-designed and well-trained deep reconstruction network satisfies our proposed bounded relative error norm (BREN) property, which is a special case of the Lipschitz continuity as detailed in part B, the theoretical part[34] of our current papers. Based on the BREN property, the converging mechanism of the ACID iteration is revealed in our two independent analyses.[34]

Here, we outline the key insight into the convergence of the ACID workflow. In reference to Figure 1, we assume the BREN property of a deep reconstruction network Φ, as characterized by the ratio being less than 1 between the norm of the reconstruction error and the norm of the corresponding ground truth (assuming a nonzero norm without the loss of generality); that is, the error component of the initial image reconstructed by the deep network Φ is less than the ground truth image in the L_2 norm. This error consists of both sparse and non-sparse components. The non-sparse component is effectively suppressed by the CS module Θ. The sparse errors are either observable or unobservable. The unobservable error is in the null space of the system matrix $\boldsymbol{A}$ and should be small relative to the ground truth image given the BREN property (the deep reconstruction network will effectively recover the null space component if it is properly designed and well trained). ACID can eliminate the observable error iteratively, owing to the BREN property. Specifically, the output of the module Θ is re-projected by the system matrix $\boldsymbol{A}$, and then the synthesized data are compared with the measured data. The difference is called the data residual due to the observable error component. To suppress this error component, we use the network Φ to reconstruct an incremental image and add it to the current image, and then refine the updated image with the CS module Θ. In this correction step, the desirable incremental image is the new ground truth image, and the BREN property remains valid as this step is a contraction mapping. In other words, the associated new observable error is less than the previous observable error, by the BREN property of the deep reconstruction network Φ. Repeating this process leads to the observable error diminishing exponentially fast (the BREN ratio less than 1). In doing so, the ACID solution will simultaneously incorporate data-driven knowledge, image sparsity preference, and measurement data consistency.

Note that, in a recent paper,[35] a two-step deep learning strategy was analyzed for tomographic imaging, in which a classical method was followed by a deep-network-based refinement to "close the gap between practice and theory" for that particular reconstruction workflow. The key idea is to use the null space network for data-driven regularization, achieving convergence based on the Lipschitz smoothness. We emphasize here that our analysis on the convergence of ACID is in a similar spirit.[34]

RESULTS

Given the importance of the recent study on the instabilities of some representative deep reconstruction networks,[15] the main motivation of our work is to stabilize deep tomographic reconstruction. Hence, our experimental setup systematically mirrored what was described by Antun et al.,[15] including datasets and their naming conventions, selected reconstruction networks, CS-based minimization benchmarks, and image quality metrics. As a result, the Ell-50 and DAGAN networks were chosen for CT and MRI reconstructions, respectively (details in the method details section and supplemental information). Both of those CT and MRI networks were subjected to the instabilities reported by Antun et al.[15] In addition to the system-level comparison, we performed an ablation study on the ACID workflow and investigated its own stability against adversarial attacks. For details about adversarial attacks, see Antun et al.[15] and our other paper.[34] The full descriptions of the original simulated cases of C1–C7, M1–M12 and A1–A4 are in the supplemental information, part I.

Stability with small structural changes

We demonstrated the performance of the ACID network with small structural changes. The Ell-50 network was used as a special FBPConvNet.[36] Figure 2 shows representative results in two simulated CT cases: C1 and C2 (the details can be found in the

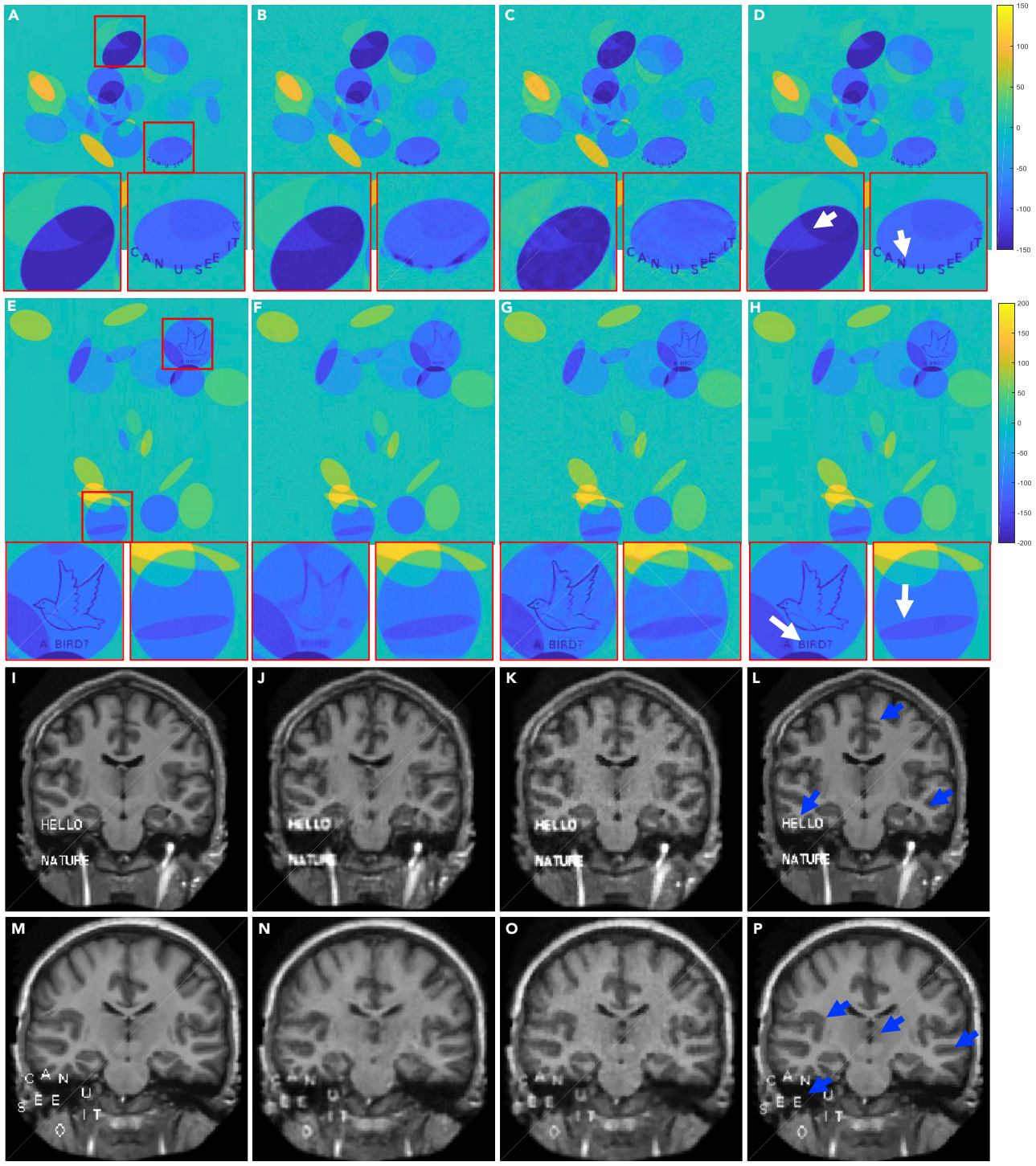


Figure 2. Performance of ACID with small structural changes in the simulated CT and MRI cases, respectively

Four phantoms with structural changes are reconstructed by ACID and competing techniques.

(A) The original image of CT case C1 with 2 magnified regions-of-interest (ROIs).

(B–D) Ell-50, CS-inspired, and ACID results, respectively, from (A).

(E –H) Counterparts of C2. Each CT dataset contains 50 projections. The image structures marked by white arrows show the advantages of our ACID in terms of CT imaging.

(I) The original image of MRI case M1.

(legend continued on next page)

supplemental information, part I). To examine the degrees of small image structure recovery allowed by all of the reconstruction methods, some text, the contour of a bird, and their mixture were used to simulate structural changes in CT images. It is observed in Figures 2A–2H that the proposed ACID network provided a superior performance owing to the synergistic fusion of deep learning, CS-based sparsification, and iterative refinement. In this case, ResNet-50 served as the deep network in the ACID workflow.

It can be seen in Figures 2A–2H that CS-inspired reconstructions produced better results than the ResNet-50 network. This is consistent with the results reported by Antun et al.[15] The CS-based reconstruction approach retained the structural changes. The text and bird were still identifiable in the CS-inspired reconstruction but became unclear in the ResNet-50 results. In contrast, the text "CAN U SEE IT" and bird were well recovered using our ACID network. While the contour of the bird was compromised in the CS reconstruction, ACID produced better image quality than the CS method.[37] In terms of edge preservation, the ResNet-50 reconstruction gave better sharpness overall than the corresponding CS reconstruction. Furthermore, ACID corrected the structural distortions seen in the ResNet-50 and CS results. A similar study was performed on MRI with small structural changes, as shown in Figures 2I–2P. Since DAGAN[38] was used as a representative network by Antun et al.,[15] we implemented it for this experiment. The text was added to brain MRI slices (M1 and M2 cases; more details in supplemental information, part I). Figures 2I–2L show the M1 results reconstructed from data subsampled at a rate of 10%. It is difficult to recognize the phrase "HELLO NATURE" in the DAGAN reconstruction. The structures were effectively recovered by the CS method but with evident artefacts due to the low subsampling rate. In addition, the edges of "HELLO NATURE" were severely blurred, as was the text. However, our ACID network produced excellent results with the clearly visible words. To further show the power of ACID with small structural changes, another example (M2) in Antun et al.[15] was reproduced as Figures 2M–2P. The text "CAN U SEE IT" was corrupted by both DAGAN and CS, rendering the insert hard to be read. Again, the text can be easily seen in the ACID reconstruction. Indeed, compared with the DAGAN and CS results, the ACID reconstruction kept sharp edges and subtle features. The reconstructed results of M2 (similar to the DAGAN results in Antun et al.,[15] but with different subsampling rate and pattern) also support the superior performance of ACID.

In brief, ACID exhibited superior stability with structural changes over the competitors, as quantified by the peak signal-to-noise ratio (PSNR), structural similarity (SSIM), normalized root-mean-square error (NRMSE), and feature similarity (FSIM) in Table 1. In all of these cases, ACID consistently obtained the highest PSNR and SSIM scores indicated by boldface font.

Stability against adversarial attacks

A tiny perturbation could fool a deep neural network to make a highly undesirable prediction,[15] which is known as an adversarial attack.[39,40] To show the capability of the ACID approach against adversarial perturbations, the simulated CT (cases C3 and C4) and MRI (cases M3 and M4) reconstructions under such perturbations are given in Figure 3. Figures 3A–3D show that ResNet-50 network led to distorted edges, as indicated by the arrows. Although the CS reconstruction had a stable performance against tiny perturbations, these distortions could not be fully corrected, with remaining subsampling artefacts. In contrast, this defect was well corrected by ACID. It is observed in Figures 3F–3I that the artefacts marked by the arrows induced by perturbation distorted the image edges in the ResNet-50 reconstruction. This could result in a clinical misinterpretation. Although these artefacts were effectively eliminated in the CS reconstruction, CS-related new artifacts were introduced. Encouragingly, the corresponding edges and shapes were faithfully reproduced by ACID without any significant artefacts. In addition, the text "CAN YOU SEE IT" was completely lost in the ResNet-50 reconstruction. In contrast, our ACID results preserved the edges and letters. The worst MRI reconstruction results from tiny perturbations were obtained by DAGAN, as shown in Figures 3L and 3Q. Compared with DAGAN, the CS-based reconstruction provided higher accuracy, but still failed to preserve critical details such as edges, as shown in Figures 3M and 3R. However, our ACID network overcame these weaknesses. Table 1 summarizes the quantitative evaluation results.

To demonstrate the ACID performance in a practical setting, more experiments were performed in these CT and simulated MRI cases with noisy data. The CT reconstruction results were obtained in the case C5, generated by adding Gaussian noise to the C1 data. Also, the reconstruction results were obtained in the experiments on M5 and M6, generated by adding Gaussian noise to the M1 and M2 datasets, respectively. With the original networks (including ResNet-50 and DAGAN) and CS methods, the image edges and other features were notably blurred. However, all of the features including the embedded words were well recovered by ACID as shown in Figure 4. It is observed that ACID gave better quantitative results than the competitors. Specifically, ACID suppressed image noise more effectively than the CS-based reconstruction method, even though the network was not trained for denoising. The quantitative results are also given in Table 1.

Stability with more input data

Intuitively, a well-designed reconstruction scheme is expected to increase its performance monotonically as more input data become available. It was pointed out by Antun et al.[15] that the performance of some deep reconstruction networks, such as ResNet-50 and DAGAN, degraded with more input data, which is certainly undesirable. To evaluate the performance of ACID with more input data, cases C1, C2, M1, and M2 were analyzed. The numbers of views in the CT cases were set to 10, 20, 30, 50, 60, 75, 100, 150, and 300, and in the simulated MRI cases, the subsampling rates were set to 1%, 5%, 10%, 20%, 30%, 40%,

(J –L) The DAGAN, CS-inspired, and ACID results, respectively, from M1.
(M–P) Counterparts of M2. The subsampling rate of MRI is 10%. The display windows for C1, C2, M1, and M2 are [−150 150]HU, [−200 200]HU, [0 0.7], and [0 1], respectively. The blue arrows demonstrate that our ACID provides much clearer image edges as well as finer structures. The difference images are provided in supplemental information, part III.B.

Table 1. Quantitative analysis results in the experiments

CT	Cases	C1	C2	C3	C4	C5	
PSNR	Ell-50	31.80	31.49	34.02	29.57	25.57	
	CS	32.62	31.81	33.52	30.44	22.49	
	ACID	40.86	38.78	38.76	36.02	31.14	
SSIM	Ell-50	0.922	0.953	0.924	0.882	0.651	
	CS	0.951	0.954	0.933	0.944	0.769	
	ACID	0.995	0.993	0.990	0.987	0.901	
NRMSE	Ell-50	0.0120	0.0108	0.0124	0.0168	0.0236	
	CS	0.0088	0.0084	0.0113	0.0112	0.0216	
	ACID	0.0039	0.0046	0.0071	0.0079	0.0140	
FSIM	Ell-50	0.960	0.981	0.955	0.938	0.868	
	CS	0.964	0.971	0.949	0.955	0.877	
	ACID	0.987	0.998	0.991	0.992	0.947	
MRI	Cases	M1	M2	M3	M4	M5	M6
PSNR	DAGAN	29.59	29.16	28.22	27.55	29.33	29.06
	CS	30.91	30.23	29.83	29.33	29.58	29.35
	ACID	37.59	34.91	34.18	32.23	34.76	32.69
SSIM	DAGAN	0.923	0.896	0.877	0.851	0.907	0.906
	CS	0.951	0.941	0.936	0.926	0.933	0.929
	ACID	0.981	0.977	0.966	0.941	0.946	0.942
NRMSE	DAGAN	0.0338	0.0350	0.0402	0.0419	0.0353	0.0358
	CS	0.0285	0.0308	0.0322	0.0342	0.0332	0.0340
	ACID	0.0133	0.0188	0.0201	0.0260	0.0192	0.0249
FSIM	DAGAN	0.969	0.953	0.949	0.935	0.964	0.956
	CS	0.977	0.967	0.968	0.958	0.972	0.965
	ACID	0.990	0.985	0.977	0.969	0.976	0.970

and 50%. Figure 5 shows that the performance of Ell-50 decreased with more projections than what were used for network training, being consistent with the observation by Antun et al.[15] In contrast, ACID performed better with more views in terms of PSNR and SSIM. Similarly, the performance of DAGAN decreased when more data were collected at sampling rates higher than those used for DAGAN training, which agrees with the conclusion on DAGAN in the work of Antun et al.[15] However, ACID produced better reconstruction quality in terms of PSNR and SSIM. The performance of ACID substantially improved with more input data, indicating that our ACID generalizes well to more input data, similar to the CS methods.

Ablation study on ACID

ACID involves deep reconstruction, CS-inspired sparsification, analytic mapping, and iterative refinement. To understand the roles of these algorithmic ingredients, we evaluated their relative contributions to the ACID reconstruction quality. Specifically, we reconstructed images using the three simplified versions of ACID by removing/replacing individual key components. The three versions include (1) improving the initial deep reconstruction with CS-inspired sparsification without iteration, (2) replacing deep reconstruction with a conventional reconstruction method, and (3) abandoning the compressed sensing constraint. Figure 6 shows that each simplified ACID variant compromised the ACID performance significantly.

Comparison with classic iteration-based unrolled networks

Based on our experiences, we believe that the use of deep learning as a post-processor or an image-domain data-driven regularizer in a classic iterative reconstruction algorithm, as was suggested by Wang,[4] is inferior to ACID that leverages the power of deep learning from the data space to the image space, since once an image is reconstructed using a classic method, some clues in the data space may be lost for deep learning-based reconstruction. It is mainly because the classic iterative reconstruction cannot take full advantage of data-driven prior, even if a deep learning image denoiser is used, such as in ADMM-net.[41] Different from existing iteration-based unrolled reconstruction networks that only use deep learning to refine an intermediate image already reconstructed using a classic iterative algorithm, ACID reconstructs an intermediate image with a deep network trained on big data and through iterative refinement. To highlight the merits of ACID, the classic ADMM-net was chosen for comparison.[41] The ADMM-net was trained on 20% subsampled data, with a radial sampling mask while the other settings are the same as that in Yang et al.[41] Figures 7A–7C show that ACID achieved the best-reconstructed image quality, followed by ADMM-net and DAGAN sequentially. The phrase "HELLO NATURE" was blurred in the DAGAN reconstruction but became clearer in the ADMM-net reconstruction. However, the artefacts due to subsampling remain evident in the ADMM-net

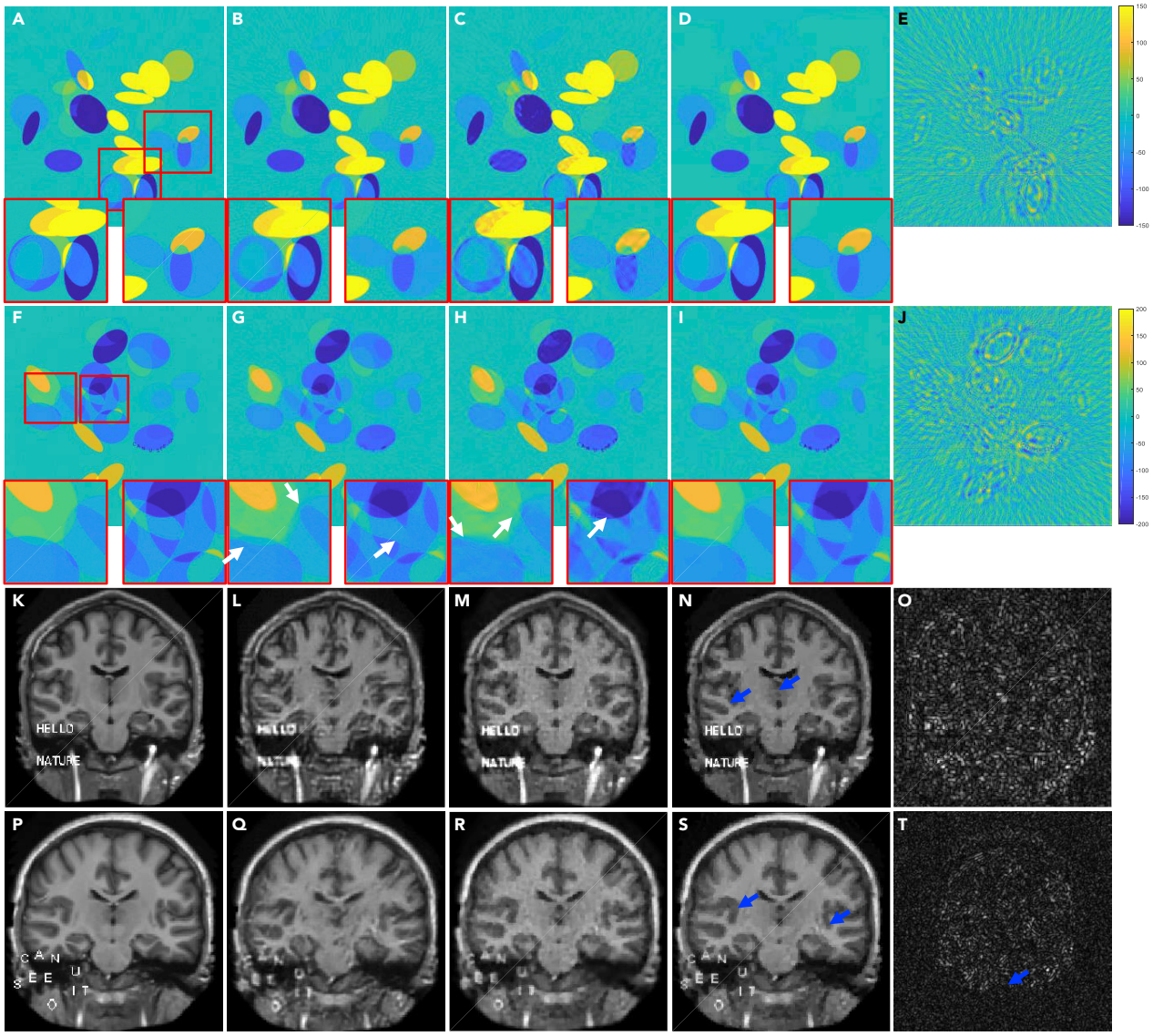


Figure 3. Performance of ACID against adversarial attacks coupled with structural changes in the simulated CT and MRI cases

(A) The ground truth CT image in the C3 case with 2 magnified ROIs (window [−80 80]HU); (B)–(D) are EII-50, CS, and ACID reconstructions; and (E) shows the adversarial sample (window [−5 5]HU). (F)–(J) are the counterparts in the C4 case (display window for (F)–(I) is [−150 150]HU and the display window for (J) is [−5 5]HU). The image structures indicated by white arrows show the advantages of our ACID in terms of CT imaging against adversarial attacks.

(K) The ground truth MRI image in the M3 case (normalized to [0, 0.7]).

(L–N) DAGAN, CS, and ACID reconstructions.

(O) The adversarial sample (window [−0.05 0.05]).

(P–T) The counterparts of MRI in the M4 case. The blue arrows demonstrate that our ACID provides much clearer image edges as well as finer structures against adversarial attacks. The difference images can be found in supplemental information, part III.B.

reconstruction. In Figures 7D–7F, there are strong artefacts in the reconstructed image by DAGAN. However, the image quality from ADMM-net is better than that of DAGAN, such as in terms of edge sharpness. The image edges and features in the ACID images are overall the best, as shown in Figure 7. To quantify the performance of these techniques, the PSNR and SSIM were computed, as shown in Figure 7.

The ACID flowchart can be unfolded into the feedforward architecture. However, such an unfolded reconstruction network (similar to MRI-VN[42]) could still be subject to adversarial attacks, if kernel awareness is not somehow incorporated. Given the current graphics processing unit (GPU) memory limit, it is often impractical to unfold the whole ACID (up to 100 iterations or more) into a single network.

The unrolled reconstruction networks show great deep tomographic performance. For example, they can reconstruct high-quality images from sparse-view measurements. However, there are at least three differences between ACID and the unrolled reconstruction networks, such as MetaInv-Net.[43] First, large-scale trained networks, such as DAGAN[38] and EII-50[36] could

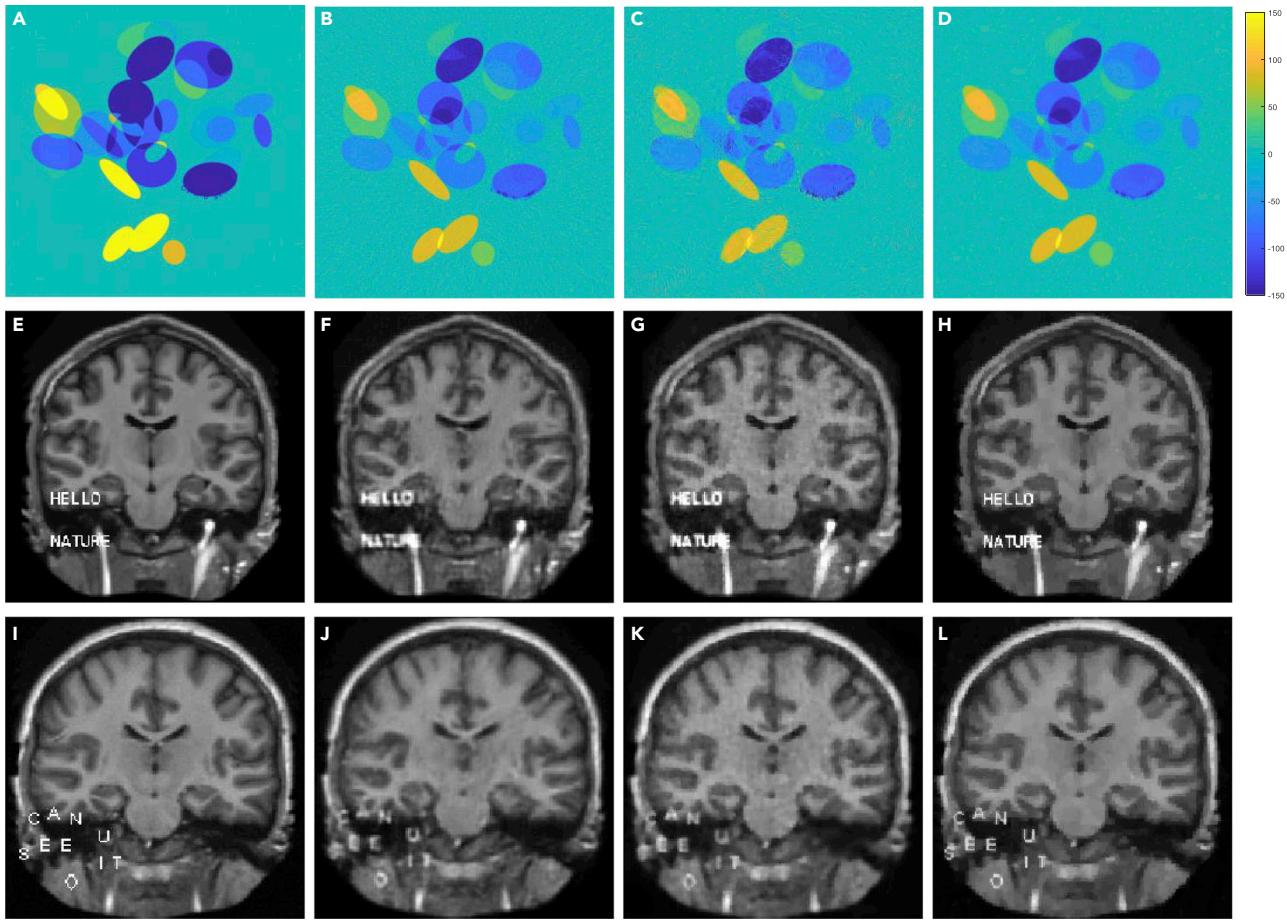


Figure 4. Reconstruction results in the simulated C5, M5, and M6 cases

(A–D) The ground truth, EII-50, CS, and ACID results on C5.

(E–H and I–L) The ground truth, DAGAN, CS, and ACID results on M5 and M6, respectively.

be incorporated into the ACID framework as demonstrated in this study. However, if the ACID scheme is unrolled into a feedforward network, only small subnetworks could be integrated—in other words, an unrolled ACID network could only use relatively light networks such as multiple layer convolutional neural networks (CNNs).[44,45] These sentences are not self-contradictory, because the ACID scheme is not an unrolled network. In fact, an unrolled network has a number of stages, each of which consumes a substantial amount of memory. Hence, the total size of the required memory is proportional to the number of stages. In contrast, ACID is computationally iterative, and thus the same memory space allocated for an iteration is reused for the next iteration. Therefore, a large-scale network can work with ACID, but when the ACID scheme is unrolled, only a lightweight network can be used for ACID reconstruction. Second, the size of images reconstructed by an unrolled network is typically small. For example, the input image consists of small pixels for ADMM-net,[41] MetaInv-Net,[43] LEARN,[44] and AirNet,[46] limited by the memory size of the GPU. The reconstructed low-resolution results could not satisfy the requirement of many clinical applications, especially for CT imaging tasks. Also, the unrolled networks were commonly designed for two-dimensional (2D) imaging, including the MetaInv-Net,[43] and they are difficult to use in 3D imaging geometry, since the memory increment is proportional to the number of stages. Third, it has not been intended by others to incorporate the theoretically grounded sparsity regularization module in such an unrolled architecture. This could be due to the fact that some needed operations (e.g., the image gradient L_0-norm[47]) could not be effectively implemented with compact feedforward networks, which demanded big data and could not be easily trained. Nevertheless, ACID can stabilize these unrolled networks. For example, Figure 8 demonstrates the results using ACID with a built-in model-based unrolled deep network (MoDL).[45] MoDL performed well with structural changes but suffered from adversarial attacks.[15] Synergistically, ACID with MoDL built in produced excellent image quality.

Adversarial attacks to the ACID system

As demonstrated above, ACID can successfully stabilize an unstable network. Then, a natural question is whether or not the whole ACID workflow itself is stable. To evaluate the stability of ACID in its entirety, we generated adversarial samples to attack the entire ACID system, with representative results in Figure 9.

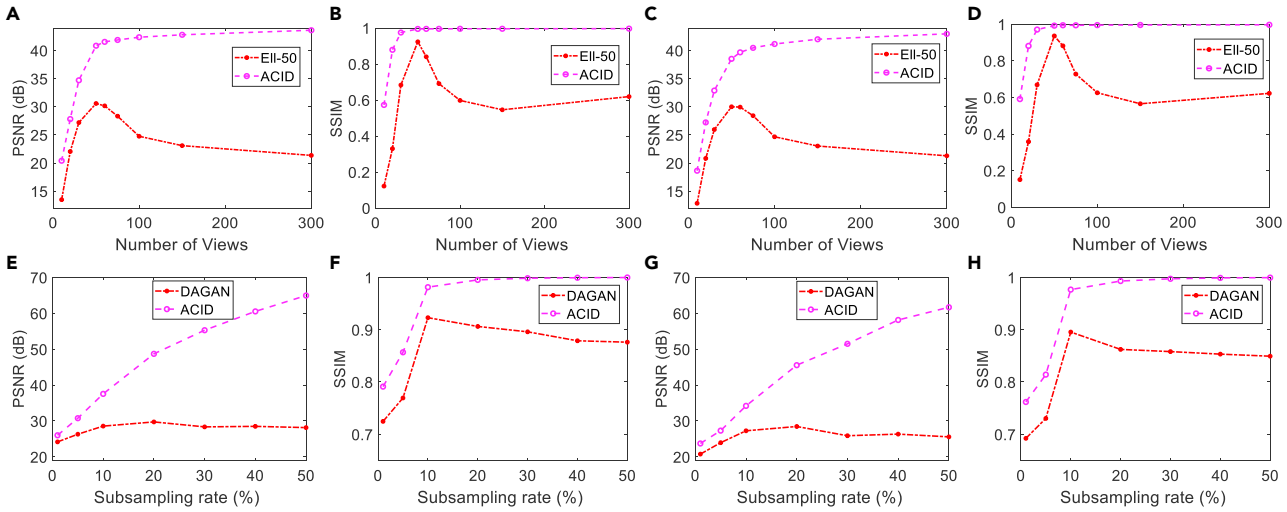


Figure 5. Performance of ACID with more input data

(A and B) and (C and D) contain the PSNR and SSIM curves with respect to the number of views in cases C1 and C2, respectively.

(E and F) and (G and H) are the same type of curves with respect to different sub-sampling rates in cases M1 and M2, respectively.

Because ACID involves both deep reconstruction and sparsity minimization in the iterative framework, the adversarial attack mechanism is more complicated for ACID than that for a feedforward neural network. (See part B[34] of this article for details on the adversarial attacking method that we used to attack ACID.) Using this adversarial method, C6, C7, and M10–M12 images were perturbed to various degrees, being even greater in terms of the L_2-norm than what were used to attack individual deep reconstruction networks. Our ACID reconstruction results show that the structural features and inserted words were still clearly

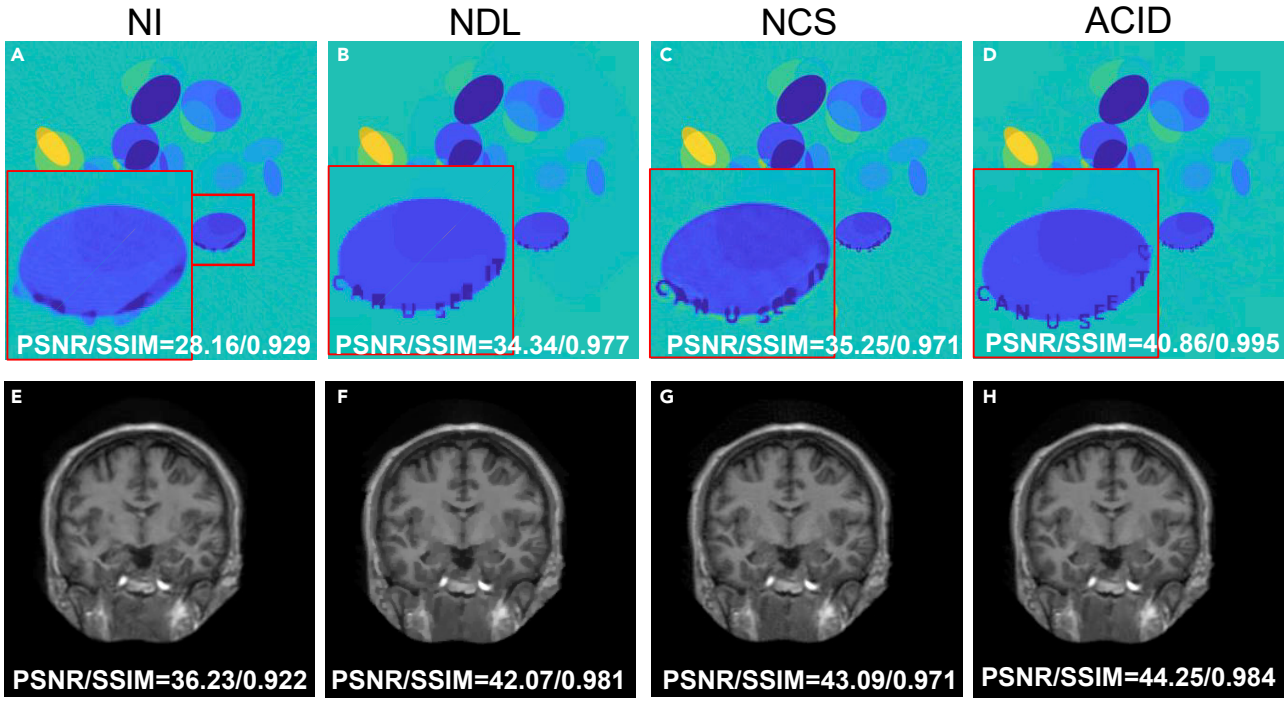


Figure 6. ACID ablation study in terms of visual inspection and quantitative metrics in the cases C1 and M7

NI denotes the reconstructed results by ACID without iterations (K = 1). NDL and NCS denote ACID without deeply learned prior and CS-based sparsification, respectively.

(A–D) These panels represent the reconstructed results by NI, NDL, NCS, and ACID in the C1 case.

(E–H) The reconstructed results by NI, NDL, NCS and ACID in the M7 case.

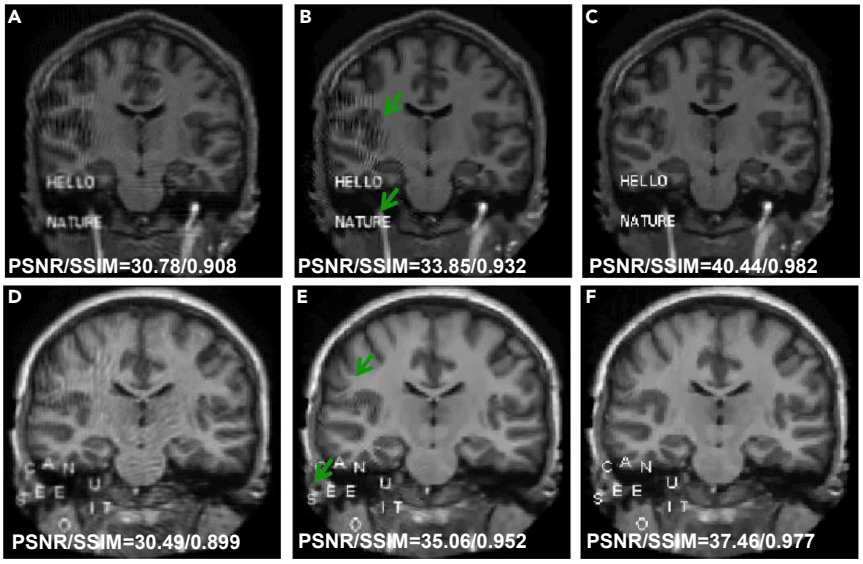


Figure 7. Comparison of reconstruction performance relative to the ADMM-net

(A–C) These panels represent the results reconstructed by DAGAN, ADMM-net, and ACID, respectively.

(D–F) Counterparts of another case.

reproduced even after these adversarial attacks. Consistently, the PSNR and SSIM results of ACID were not significantly compromised by the adversarial attacks.

Stabilization of AUTOMAP

AUTOMAP, an important milestone in medical imaging, was used as another classic example by Antun et al.[15] to demonstrate the instabilities of deep tomographic reconstruction. To further test the stability of ACID, cases A1 and A2 with structural changes and A3 and A4 cases with adversarial attacks were used, as shown in Figure 10 (details on cases A1–A4 are in supplemental information, part I). It is observed that AUTOMAP demonstrated good ability against structural changes but that it suffered from adversarial attacks.[15] ACID produced significantly better image quality than AUTOMAP. Beyond the visual inspection, ACID achieved better PSNR and SSIM values than AUTOMAP. (See supplemental information, part I for more details.)

DISCUSSION

As clearly reviewed in the theoretical part of our article series,[34] the kernel awareness[26] is important to avoid the so-called cardinal sin. When input vectors are very close to the null space of the associated measurement matrix, if the input is slightly perturbed, then a large variation may be induced in the reconstructed image. If an algorithm lacks the kernel awareness, then it will be intrinsically vulnerable, suffering from false-positive and false-negative errors; for mathematical rigor, please see Theorem 3.1 in Gottschling et al.[26] For this reason, the deep tomographic networks were successfully attacked in Antun et al.[15] However, sparsity-promoting algorithms were designed with the kernel awareness, leading to an accurate and stable recovery of underlying images, as also shown in Antun et al.[15] As demonstrated by our results here, the kernel awareness has been embedded in the ACID scheme through both the CS-based sparsity constraint and the iterative refinement mechanism. Hence, ACID demonstrates a robust performance against noise, under adversarial attacks, and when the amount of input data is increased relative to what was used for network training. A different empirical method was also designed by peers,[48] which produced results complementary to ours.[49]

It is important to understand how a CS-based image recovery algorithm implements the kernel awareness. The sparsity-constrained solution is iteratively obtained so that the search for the solution is within a low-dimensional manifold. That is, prior knowledge known as sparsity helps narrow down the solution space. Indeed, natural and medical images allow low-dimensional manifold representations.[50] It is critical to emphasize that a deep neural network is data driven, and the resultant data-driven prior is rather powerful to constrain the solution space greatly. While sparsity prior is just one or a few mathematical expressions, deep prior is in a deep network topology with a large amount of parameters extracted from big data. In this study, we incorporated the Ell-50 and DAGAN network into the ACID workflow. In fact, ACID as a general framework can integrate more advanced reconstruction neural networks,[51] such as PIC-GAN[52] and SARA-GAN.[53] These two kinds of priors (sparsity prior and deep priors) can be combined in our ACID workflow in various ways to gain the merits from both sides. Because the combination of deep prior and sparsity prior is more informative than sparsity prior alone, ACID or similar networks would outperform classic algorithms, including CS-inspired sparsity-promoting methods. Indeed, with a deep reconstruction capability, ACID outperforms the representative CS-based methods for image reconstruction, including dictionary learning reconstruction methods (see details in supplemental information, part II.B). Indeed, we only quantitatively evaluated the main reconstruction results in terms of PSNR, NRMSE, SSIM, and FSIM. Our current evaluative metrics directly correspond to what were used in the *Proceedings of the National Academy of Sciences* (*PNAS*) study.[15] However, it will be valuable and interesting to assess the clinical influence of reconstructed results using other advanced assessment methods (local perturbation responses[54] and Frechet inception distance[55]). In addition, the used deep tomographic networks are based on CNN architectures. Recently, the transformer as an advanced deep learning technique was used for image reconstruction. For example, Pan et al. developed a multi-domain integrative Swin transformer network (MIST-net) for sparse-view reconstruction.[56] Furthermore, the Swin transformer was used for MRI reconstruction.[57] How to stabilize transformer-based deep reconstruction networks is also important. We will pursue studies along this direction in the near future.[58]

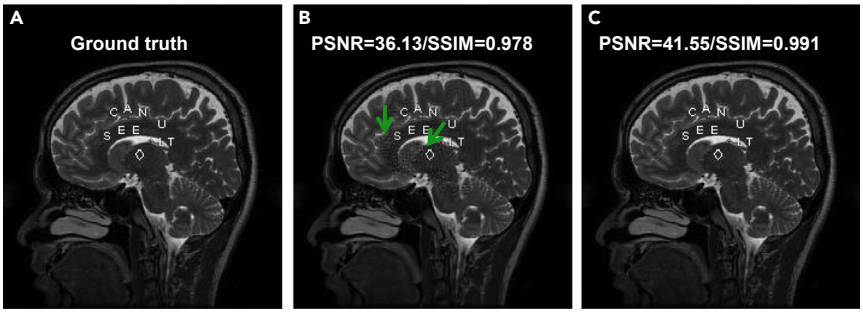


Figure 8. Stabilization of MoDL using the ACID strategy

(A–C) These panels represent a representative reference, corresponding results reconstructed by MoDL and ACID (with MoDL built in), respectively, where adversarial attacks were applied to the MoDL network, which was then successfully defended using the ACID scheme with MoDL built in.

This study represents our specific response to the challenge presented in the *PNAS* paper by Antun et al.[15] For that purpose and as the first step, the deep reconstruction networks and associated datasets we used are the same as what were used in the *PNAS* study, thereby making it clear and convincing for the readers to evaluate their relative performance. As a result, we also inserted the unrealistic features (e.g., bird, letters) used in the *PNAS* study. We emphasize that these experiments represent substantially easier inverse problems than real CT/MRI studies. How to evaluate and optimize the diagnostic performance of ACID-type algorithms in clinical tasks will be pursued in the future, which include real pathological features such as tumors.

In conclusion, our proposed ACID workflow has synergized deep network-based reconstruction, CS-inspired sparsity regularization, analytic forward mapping, and iterative data residual correction to systematically overcome the instabilities of the deep reconstruction networks selected in Antun et al.[15] and achieved better results than the CS algorithms used by them. It is emphasized that the ACID scheme is only an exemplary embodiment, and other hybrid reconstruction schemes of this type can be also investigated in a similar spirit.[59,60] We anticipate that this integrative data-driven approach will help promote the development and translation of deep tomographic image reconstruction networks into clinical applications.

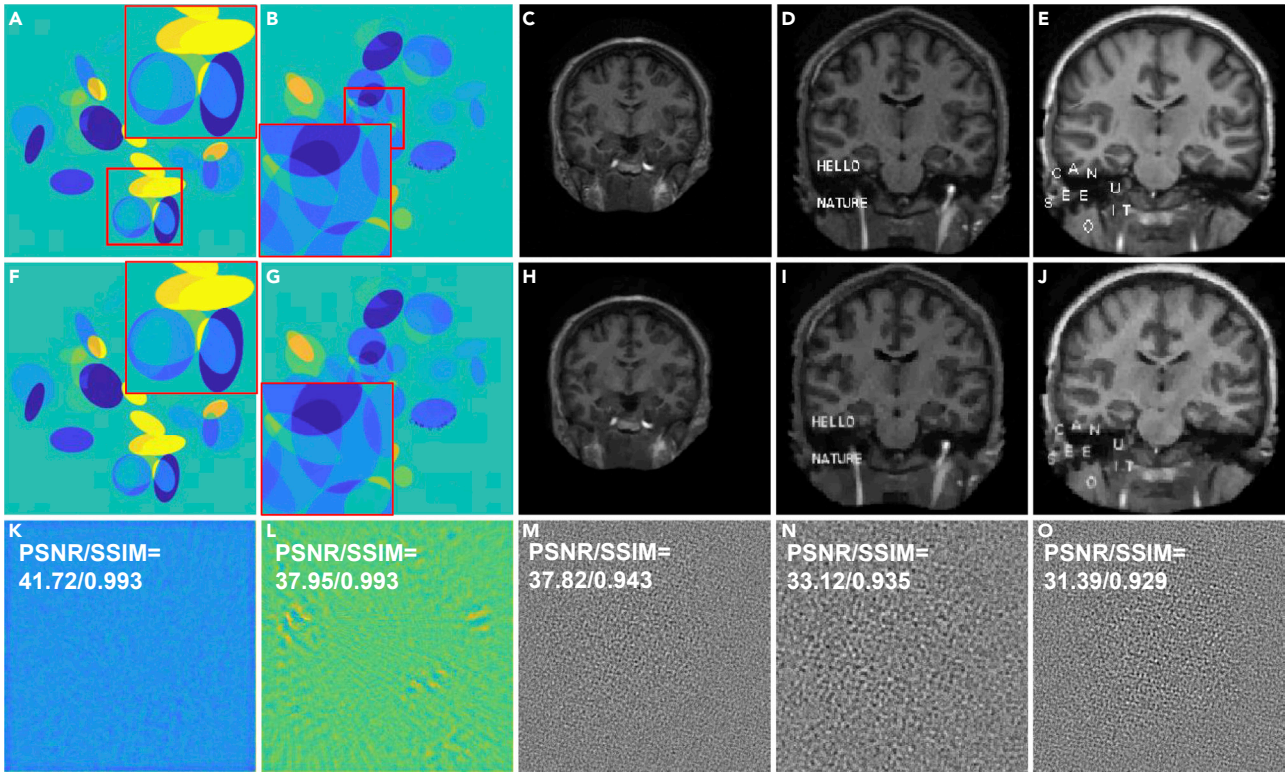


Figure 9. ACID being resilient against adversarial attacks

From left to right, the columns are ACID results in C6, C7, and M10–M12 cases, respectively. The first -third rows represent the ground truth plus tiny perturbation, reconstructed images, and corresponding perturbations.

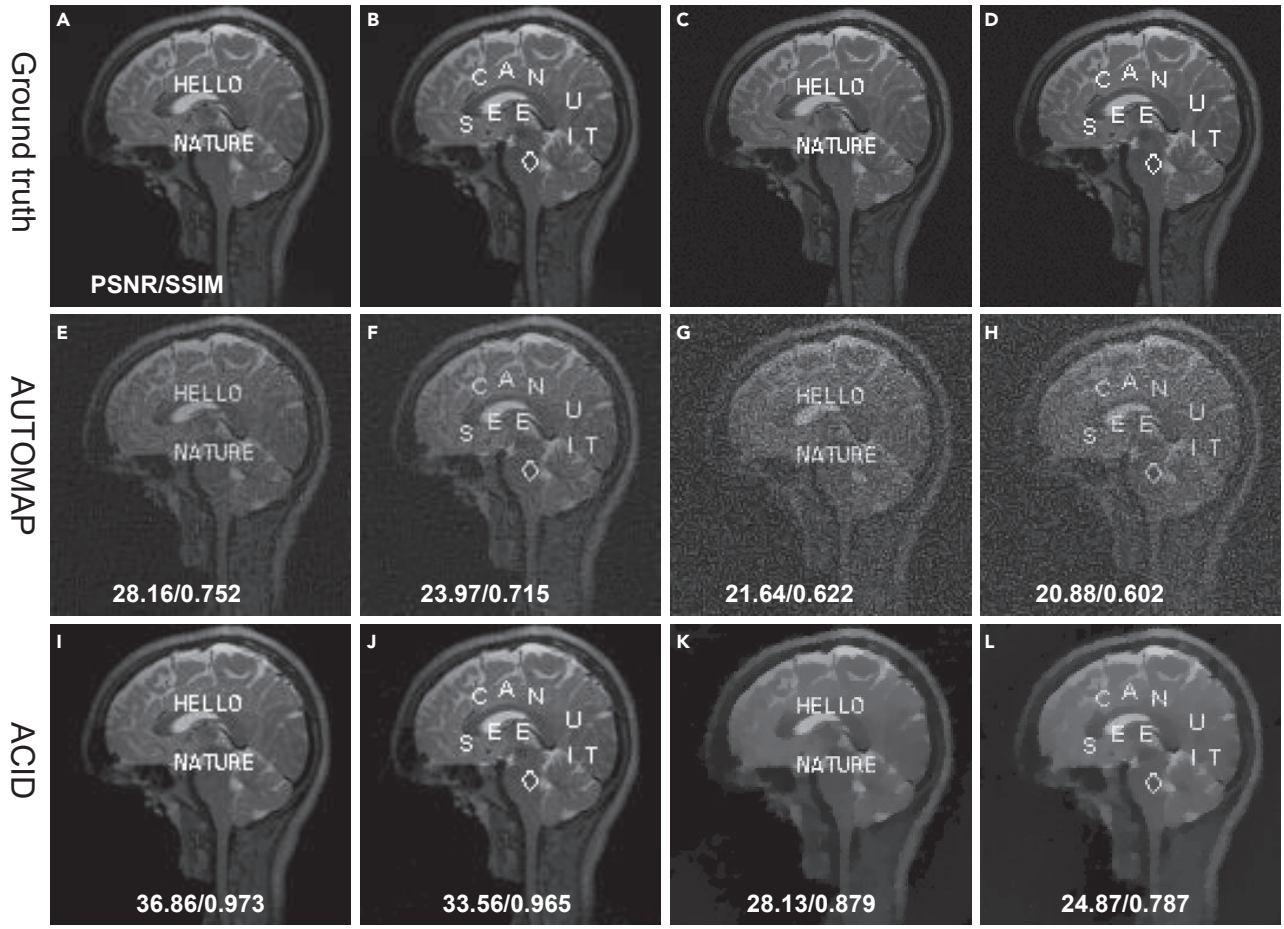


Figure 10. Stabilization of AUTOMAP using ACID

The first and second columns represent the reconstruction results from structural changes, where the first, second, and third rows represent the reference, AUTOMAP, and ACID (with AUTOMAP built in) results, respectively. Third and fourth columns are the counterparts under adversarial attack, where the first, second, and third rows denote the reference plus perturbation, AUTOMAP, and ACID (with AUTOMAP built in) results, respectively.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Hengyong Yu, PhD (e-mail: hengyong-yu@ieee.org).

Materials availability

The study did not generate new unique reagents.

Data and code availability

The codes, trained networks, test datasets, and reconstruction results are publicly available on Zenodo (https://zenodo.org/record/5497811).

Method details

Heuristic ACID scheme

In the imaging field, we often assume that the measurement $\boldsymbol{p}^{(0)} = \boldsymbol{A}\boldsymbol{f}^* + \boldsymbol{e}$, where $\boldsymbol{A} \in \mathbb{R}^{m\times N}$ is a measurement matrix (e.g., $\boldsymbol{A}$ is the Radon transform for CT[61] and the Fourier transform for MRI[62]), $\boldsymbol{p}^{(0)} \in \mathbb{R}^m$ is an original dataset, $\boldsymbol{f}^* \in \mathbb{R}^N$ is the ground truth image, $\boldsymbol{e} \in \mathbb{R}^m$ is data noise, and most relevant, $m < N$, meaning that the inverse problem is underdetermined. In the underdeterministic case, additional prior knowledge must be introduced to recover the original image uniquely and stably. Typically, we assume that $\boldsymbol{H} \in \mathbb{R}^{N\times N}$ is an invertible transform, $\boldsymbol{A}$ satisfies the restricted isometry property (RIP) of order s[63] (note that ACID works even without RIP, but in that case the solution may or may not be unique; see the theoretical part of our articles, part B[34]), and $\boldsymbol{H}\boldsymbol{f}^*$ is s-sparse. We further assume that the function $\Phi(\cdot)$ models a properly designed and well-trained neural network with the BREN property that continuously maps measurement data to an image. To solve the problem of reconstructing $\boldsymbol{f}$ from measurement $\boldsymbol{p}^{(0)}$, the ACID scheme is heuristically derived from the following iterative solution (see part B of our article series[34]):

$$\begin{cases} \boldsymbol{p}^{(k+1)} = \dfrac{\lambda\left(\mathbf{p}^{(0)} - \mathbf{A}\boldsymbol{f}^{(k)}\right)}{1+\lambda} \\ \boldsymbol{f}^{(k+1)} = \boldsymbol{H}^{-1} S_{\varepsilon}\left(\boldsymbol{H}\left(\boldsymbol{f}^{(k)} + \dfrac{1}{\lambda}\Phi\left(\boldsymbol{p}^{(k+1)}\right)\right)\right) \end{cases} \quad \text{(Equation 1)}$$

where k is the index for iteration, $k = 0, 1, 2, \ldots$, $\lambda > 0$ and $\varepsilon > 0$ are hyperparameters, $\boldsymbol{H}^{-1}$ is the inverse transform of $\boldsymbol{H}$, and $S_{\varepsilon}(\cdot)$ is the soft-thresholding filtering kernel function defined as

$$S_{\varepsilon}(x) = \begin{cases} 0, & |x| < \varepsilon \\ x - sgn(x)\varepsilon & \text{otherwise} \end{cases}. \quad \text{(Equation 2)}$$

In our experiments, $\boldsymbol{H}\boldsymbol{f}$ is specialized as a discrete gradient transform, and $\boldsymbol{H}^{-1}$ is interpreted as a pseudo-inverse[64] (see part B of our article series[34]). Under the same conditions described by Yu and Wang,[64] the ACID iteration would converge to a feasible solution subject to an uncertainty range proportional to the noise level (under the conditions and approximations discussed in part B of our article series[34]).

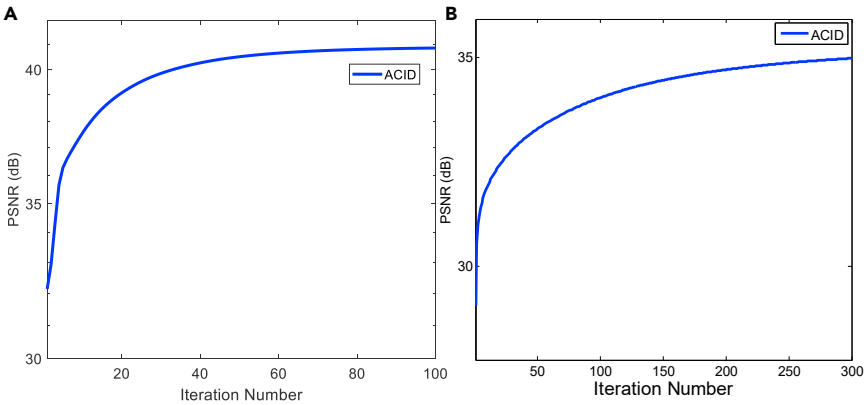


Figure 11. Convergence of the ACID iteration in terms of PSNR

(A and B) These panels show the convergence curves in the C1 and M2 cases, respectively.

Selected unstable networks stabilized in the ACID framework

The Ell-50 and DAGAN networks are two examples of unstable deep reconstruction networks chosen to validate the effectiveness of ACID, both of which were used in Antun et al.[15] and suffered from the three kinds of instabilities. In addition, the results from stabilizing AUTOMAP, a milestone deep tomographic network, were also included.

The projection data for Ell-50-based CT reconstruction were generated using the *radon* function in MATLAB R2017b, where 50 indicates the number of projections. For fair comparison, we only used the trained networks by Jin et al.,[36] which were the same as those used in Antun et al.[15] The test image for case C1 was provided by Antun et al.,[15] which can be downloaded from the related website.[15] Case C2 with the bird icon and text "A BIRD?" was provided by Gottschling et al.[26] and downloaded from the specified website.[15] The test images are of 512 × 512 pixels containing structural features without any perturbation. To generate adversarial attacks, the proposed method in Antun et al.[15] was used. Then, we obtained C3 and C4 images by adding perturbations to C1 and C2 images, respectively. Furthermore, Gaussian noise with zero mean and deviation 15 HU over the pixel value range was superimposed in case C1 to obtain case C5, and adversarial attacking was performed on the whole ACID workflow by perturbing C1 and C2 images to generate C6 and C7 images, respectively. More details on the datasets and implementation details are in the supplemental information.

To evaluate ACID in the MRI case, the DAGAN network was used,[38] which was proposed for single-coil MRI reconstruction. In this study, we set the subsampling rate to 10% and subsampled the resultant images with the 2D Gaussian sampling pattern. Also, the DAGAN network was re-trained, with the same hyperparameters and training datasets as those used by Yang et al.[38] The test images were a series of brain images, each of which consists of 256 × 256 pixels. Case M1 was randomly chosen from the test dataset,[38] then the phrase "HELLO NATURE" was placed in the image as structural changes. Case M2 was obtained in the same way as in Antun et al.,[15] where the sentence "CAN U SEE IT" and "◇" were added to the original image. Also, we applied the same attacking technique used in Antun et al.[15] to generate adversarial samples. These perturbations were added to M1 and M2 images to obtain M3 and M4 images, respectively.[15] Furthermore, the Gaussian noise with zero mean and deviation of 15 over the pixel value range [0, 255] was superimposed to cases M1 and M2 to obtain M5 and M6 images. M7 was randomly chosen from the DAGAN test dataset, which can be freely downloaded.[38] In addition, cases M8 and M9 were generated by putting a radial mask of a 20% subsampling rate on the M1 and M2 images, which were used to compare ACID with ADMM-net.[41] Cases M10 and M12 were generated by directly perturbing the entire ACID system. The comparative results are given in supplemental information, part I.B.

ACID was then compared with AUTOMAP for MRI reconstruction. The AUTOMAP network was tested on subsampled single-coil data. The trained AUTOMAP weights used in our experiments were provided by Zhu et al.[16] The AUTOMAP network took a vectorized subsampled measurement data as its input. First, the complex *k*-space data were computed using the discrete Fourier transform of an MRI image. Second, subsampled *k*-space data were generated with a subsampling mask. Lastly, the measurement data were reshaped into a vector and fed into the AUTOMAP network. In this study, the images of 128 × 128 pixels at a subsampling rate of 60% were used for testing. The original image was provided in Antun et al.,[15] "HELLO NATURE," "CAN U SEE IT," and "◇" were added to the test image to generate A1 and A2 with the structural changes. The perturbations were added to images A1 and A2 to obtain images A3 and A4.[15] For representative results, please see supplemental information, part I.C.

Image quality assessment

To quantitatively compare the results obtained with different reconstruction methods, the PSNR was used to measure the difference between a reconstructed image and the corresponding ground truth image. Also, the SSIM was used to assess the similarity between images. In addition, the NRMSE and FSIM[65] are also used to assess the main results. For qualitative analysis, the reconstructed results were visually inspected for structural changes (i.e., the inserted text, bird, and patterns) and artefacts induced by perturbation. In this context, we focused mainly on details such as edges and integrity such as overall appearance.

To highlight the merits and stability of the ACID scheme, the representative CS-based methods served as the baseline. For CT, the sparsity-regularized method combining X-lets (shearlets) and total variation (TV) was used,[66] which is consistent with the selection in Antun et al.[15] For MRI, the total generalized variation (TGV) method was chosen.[67] All of the parameters, including the number of iterations for these CS methods, were optimized for fair comparison, as further detailed in the supplemental information.

Numerical verification of convergence

To verify the convergence of the ACID iteration, we numerically investigated the convergence rate and computational cost. We used PSNR as the metric to reflect the convergence of ACID (Figure 11). It can be seen that the ACID iteration converged after approximately 30 iterations for CT, and became stable after 250 iterations for MRI. In this study, we set the number of iterations to 100 and 300 for CT and MRI, respectively. In addition, we empirically showed the convergence of ACID in terms of the Lipschitz constant (see part B of our article series[34] for details).

ACID parameterization

The ACID method mainly involves two parameters, λ and ε, as defined in Equation 1. These parameters were optimized based on our quantitative and qualitative analyses, as summarized in Table 2.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.patter.2022.100474.

ACKNOWLEDGMENTS

W.W. was partially supported by the Li Ka Shing Medical Foundation, Hong Kong. V.V. is partially supported by the Li Ka Shing Medical Foundation, Hong Kong. W.W., H.S., W.C., C.N., H.Y., and G.W. are partially supported by NIH grants U01EB017140, R01EB026646, R01CA233888, R01CA237267, and R01HL151561, USA.

AUTHOR CONTRIBUTIONS

G.W. initiated the project, and supervised the team in collaboration with H.Y. and V.V.; W.W., H.Y., and G.W. designed the ACID network; W.W. and D.H.

Table 2. Parameters optimized in the experiments

Variables	C1	C2	C3	C4	C5	C6	C7	M1
$\varepsilon(10^{-3})$	0.700	0.700	0.500	1.100	0.35	0.100	0.700	0.333
λ	0.76	0.76	14	2.4	60	3.0	0.76	0.1
Variables	M2	M3	M4	M5	M6	M7	M8	M9
$\varepsilon(10^{-3})$	0.333	0.500	0.500	0.667	0.667	0.333	0.500	0.500
λ	0.1	0.01	0.01	0.01	0.01	0.1	0.01	0.01
Variables	M10	M11	M12	A1	A2	A3	A4	
$\varepsilon(10^{-3})$	0.100	0.100	0.100	0.33	0.33	2.0	2.0	
λ	0.01	0.01	0.01	3.10	3.10	6.10	6.10	

conducted the experiments; H.Y., W.C., and G.W. established the mathematical model and performed the theoretical analysis; W.W., H.Y., and G.W. drafted the manuscript; W.W., D.H., and H.S. worked on user-friendly codes/data sharing; and all of the co-authors participated in discussions, contributed technical points, and revised the manuscript iteratively.

DECLARATION OF INTERESTS

G.W. is an advisory board member of *Patterns*. An invention disclosure was filed to the Office of Intellectual Property Optimization of Rensselaer Polytechnic Institute in August 2020, and the US Non-provisional Patent Application was filed in August 2021. The authors declare no competing interests.

Received: November 26, 2021
Revised: December 24, 2021
Accepted: March 1, 2022
Published: April 6, 2022

REFERENCES

1. (2020). Number of Magnetic Resonance Imaging (MRI) Units in Selected Countries as of 2019. Health, Pharma & Medtech, Medical Technology. https://www.statista.com/statistics/271470/mri-scanner-number-of-examinations-in-selected-countries/
2. (2018). Over 75 Million CT Scans Are Performed Each Year and Growing Despite Radiation Concerns. iData Research Intelligence Behind The Data. https://idataresearch.com/over-75-million-ct-scans-are-performed-each-year-and-growing-despite-radiation-concerns/
3. Fuchs, V.R., and Sox, H.C., Jr. (2001). Physicians' views of the relative importance of thirty medical innovations. Health Aff. *20*, 30–42. https://doi.org/10.1377/hlthaff.20.5.30.
4. Wang, G. (2016). A perspective on deep imaging. IEEE Access *4*, 8914–8924. https://doi.org/10.1109/ACCESS.2016.2624938.
5. Wang, G., Ye, J.C., Mueller, K., and Fessler, J.A. (2018). Image reconstruction is a new frontier of machine learning. IEEE Trans. Med. Imaging *37*, 1289–1296. https://doi.org/10.1109/TMI.2018.2833635.
6. Cong, W.X., Xi, Y., Fitzgerald, P., De Man, B., and Wang, G. (2020). Virtual monoenergetic CT imaging via deep learning. Patterns *1*, 100128. https://doi.org/10.1016/j.patter.2020.100128.
7. Khater, I.M., Nabi, I.R., and Hamarneh, G. (2020). A review of super-resolution single-molecule localization microscopy cluster analysis and quantification methods. Patterns *1*, 100038. https://doi.org/10.1016/j.patter.2020.100038.
8. Born, J., Beymer, D., Rajan, D., Coy, A., Mukherjee, V.V., Manica, M., Prasanna, P., Ballah, D., Guindy, M., Shaham, D., et al. (2021). On the role of artificial intelligence in medical imaging of covid-19. Patterns *2*, 100269. https://doi.org/10.1016/j.patter.2021.100269.
9. Wu, W.W., Hu, D., Niu, C., Yu, H.Y., Vardhanabhuti, V., and Wang, G. (2021). DRONE: dual-domain residual-based optimization network for sparse-view CT reconstruction. IEEE Trans. Med. Imaging *40*, 3002–3014. https://doi.org/10.1109/TMI.2021.3078067.
10. Wang, G., Zhang, Y., Ye, X.J., and Mou, X.Q. (2019). Machine Learning for Tomographic Imaging (IOP Publishing), p. 410.
11. Perlman, O., Zhu, B., Zaiss, M., Rosen, M.S., and Farrar, T.C. (2022). An end-to-end AI-based framework for automated discovery of CEST/MT MR fingerprinting acquisition protocols and quantitative deep reconstruction (AutoCEST). Magn. Reson. Med. 19. https://doi.org/10.1002/mrm.29173.
12. Li, H.Y., Zhao, H.T., Wei, M.L., Ruan, H.X., Shuang, Y., Cui, T.J., Del Hougne, P., and Li, L. (2020). Intelligent electromagnetic sensing with learnable data acquisition and processing. Patterns *1*, 100006. https://doi.org/10.1016/j.patter.2020.100006.
13. Wang, G., Ye, J.C., and De Man, B. (2020). Deep learning for tomographic image reconstruction. Nat. Mach Intell. *2*, 737–748. https://doi.org/10.1038/s42256-020-00273-z.
14. Chen, Y., Schönlieb, C.B., Liò, P., Leiner, T., Dragotti, P.L., Wang, G., Rueckert, D., Firmin, D., and Yang, G. (2022). AI-based reconstruction for fast MRI—a systematic review and meta-analysis. Proc. IEEE *110*, 224–245. https://doi.org/10.1109/JPROC.2022.3141367.
15. Antun, V., Renna, F., Poon, C., Adcock, B., and Hansen, A.C. (2020). On instabilities of deep learning in image reconstruction and the potential costs of AI. Proc. Natl. Acad. Sci. U S A *117*, 30088–30095. https://doi.org/10.1073/pnas.1907377117.
16. Zhu, B., Liu, J.Z., Cauley, S.F., Rosen, B.R., and Rosen, M.S. (2018). Image reconstruction by domain-transform manifold learning. Nature *555*, 487–492. https://doi.org/10.1038/nature25988.
17. Knoll, F., Murrell, T., Sriram, A., Yakubova, N., Zbontar, J., Rabbat, M., Defazio, A., Muckley, M.J., Sodickson, D.K., Zitnick, C.L., et al. (2020). Advancing machine learning for MR image reconstruction with an open competition: overview of the 2019 fastMRI challenge. Magn. Reson. Med. *84*, 3054–3070. https://doi.org/10.1002/mrm.28338.
18. Natterer, F. (2001). The Mathematics of Computerized Tomography (SIAM Publisher).
19. Kak, A.C., and Slaney, M. (1988). Principles of Computerized Tomographic Imaging (IEEE Press), p. 329.
20. Feng, L., Benkert, T., Block, K.T., Sodickson, D.K., Otazo, R., and Chandarana, H. (2017). Compressed sensing for body MRI. J. Magn. Reson. Imaging *45*, 966–987. https://doi.org/10.1002/jmri.25547.
21. Szczykutowicz, T.P., and Chen, G.H. (2010). Dual energy CT using slow kVp switching acquisition and prior image constrained compressed sensing. Phys. Med. Biol. *55*, 6411–6429. https://doi.org/10.1088/0031-9155/55/21/005.
22. Wu, W.W., Hu, D., An, K., Wang, S., and Luo, F. (2020). A high-quality photon-counting CT technique based on weight adaptive total-variation and image-spectral tensor factorization for small animals imaging. IEEE Trans. Instrumentation Meas. *70*, 14. https://doi.org/10.1109/TIM.2020.3026804.
23. Wu, W.W., Zhang, Y.B., Wang, Q., Liu, F.L., Chen, P., and Yu, H.Y. (2018). Low-dose spectral CT reconstruction using image gradient ℓ0–norm and

tensor dictionary. Appl. Math. Model. *63*, 538–557. https://doi.org/10.1016/j.apm.2018.07.006.

24. Candes, E.J., and Tao, T. (2006). Near-optimal signal recovery from random projections: universal encoding strategies? IEEE Trans. Inf. Theor. *52*, 5406–5425. https://doi.org/10.1109/TIT.2006.885507.
25. Herman, G.T., and Davidi, R. (2008). Image reconstruction from a small number of projections. Inverse Probl. *24*, 45011–45028. https://doi.org/10.1088/0266-5611/24/4/045011.
26. Gottschling, N.M., Antun, V., Adcock, B., and Hansen, A.C. (2020). The troublesome kernel: why deep learning for inverse problems is typically unstable. Preprint at arXiv. 2001.01258.
27. Chen, G.H., Tang, J., and Leng, L. (2008). Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets. Med. Phys. *35*, 660–663. https://doi.org/10.1118/1.2836423.
28. Rudin, L.I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena *60*, 259–268. https://doi.org/10.1016/0167-2789(92)90242-F.
29. Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., and Ma, Y. (2012). Robust recovery of subspace structures by low-rank representation. IEEE Trans. Pattern Anal. Machine Intelligence *35*, 171–184. https://doi.org/10.1109/TPAMI.2012.88.
30. Tosic, I., and Frossard, P. (2011). Dictionary learning. IEEE Signal Process. Mag. *28*, 27–38. https://doi.org/10.1109/MSP.2010.939537.
31. Danilova, M., Dvurechensky, P., Gasnikov, A., Gorbunov, E., Guminov, S., Kamzolov, D., and Shibaev, I. (2020). Recent theoretical advances in non-convex optimization. Preprint at arxiv. 2012.06188.
32. Wang, X., Yan, J., Jin, B., and Li, W. (2019). Distributed and parallel ADMM for structured nonconvex optimization problem. IEEE Trans. Cybern *51*, 4540–4552. https://doi.org/10.1109/TCYB.2019.2950337.
33. Barber, R.F., and Sidky, E.Y. (2016). MOCCA: mirrored convex/concave optimization for nonconvex composite functions. J. Mach Learn Res. *17*, 1–51.
34. Wu, W.W., Hu, D., Cong, W.X., Shan, H.M., Wang, S.Y., Niu, C., Yan, P.K., Yu, H.Y., Vardhanabhuti, V., and Wang, G. (2022). Stabilizing deep tomographic reconstruction– Part B: convergence analysis and adversarial attacks. Patterns *3*, 100475.
35. Schwab, J., Antholzer, S., and Haltmeier, M. (2019). Deep null space learning for inverse problems: convergence analysis and rates. Inverse Probl. *35*. https://doi.org/10.1088/1361-6420/aaf14a.
36. Jin, K.H., McCann, M.T., Froustey, E., and Unser, M. (2017). Deep convolutional neural network for inverse problems in imaging. IEEE Trans. Image Process. *26*, 4509–4522. https://doi.org/10.1109/TIP.2017.2713099.
37. Yu, H.Y., and Wang, G. (2009). Compressed sensing based interior tomography. Phys. Med. Biol. *54*, 2791–2805. https://doi.org/10.1088/0031-9155/54/9/014.
38. Yang, G., Yu, S., Dong, H., Slabaugh, G., Dragotti, P.L., Ye, X., Liu, F., Arridge, S., Keegan, J., Guo, Y., et al. (2018). DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. IEEE Trans. Med. Imaging *37*, 1310–1321. https://doi.org/10.1109/TMI.2017.2785879.
39. Dai, H., Li, H., Tian, T., Huang, X., Wang, L., Zhu, J., and Song, L. (2018). Adversarial attack on graph structured data. In International conference on machine learning (PMLR), pp. 1115–1124.
40. Zheng, T., Chen, C., and Ren, K. (2019). Distributionally adversarial attack. Proc. AAAI Conf. Artif. Intelligence *33*, 2253–2260.
41. Yang, Y., Sun, J., Li, H., and Xu, Z. (2016). Deep ADMM-Net for compressive sensing MRI. In Advances in neural information processing systems, pp. 10–18.
42. Hammernik, K., Klatzer, T., Kobler, E., Recht, M.P., Sodickson, D.K., Pock, T., and Knoll, F. (2018). Learning a variational network for reconstruction of accelerated MRI data. Magn. Reson. Med. *79*, 3055–3071. https://doi.org/10.1002/mrm.26977.
43. Zhang, H.M., Liu, B.D., Yu, H.Y., and Dong, B. (2020). MetaInv-net: meta inversion network for sparse view CT image reconstruction. IEEE Trans. Med. Imaging *40*, 621–634. https://doi.org/10.1109/TMI.2020.3033541.
44. Chen, H., Zhang, Y., Chen, Y., Zhang, J., Zhang, W., Sun, H., Lv, Y., Liao, P., Zhou, J., and Wang, G. (2018). LEARN: learned experts' assessment-based reconstruction network for sparse-data CT. IEEE Trans. Med. Imaging *37*, 1333–1347. https://doi.org/10.1109/TMI.2018.2805692.
45. Aggarwal, H.K., Mani, M.P., and Jacob, M. (2019). MoDL: model-based deep learning architecture for inverse problems. IEEE Trans. Med. Imaging *38*, 394–405. https://doi.org/10.1109/TMI.2018.2865356.
46. Chen, G., Hong, X., Ding, Q., Zhang, Y., Chen, H., Fu, S., Zhao, Y., Zhang, X., Ji, H., Wang, G., et al. (2020). AirNet: fused analytical and iterative reconstruction with deep neural network regularization for sparse-data CT. Med. Phys. *47*, 2916–2930. https://doi.org/10.1002/mp.14170.
47. Xu, L., Lu, C., Xu, Y., and Jia, J. (2011). Image smoothing via L_0 gradient minimization. In Proceedings of the 2011 SIGGRAPH Asia Conference, ACM Transactions on Graphics, *30*, Bala Kavita, ed. *(ACM), pp. 1–12*.
48. Genzel, M., Macdonald, J., and Marz, M. (2022). Solving inverse problems with deep neural networks - robustness included. IEEE Trans. Pattern Anal. Machine Intelligence. https://doi.org/10.1109/TPAMI.2022.3148324.
49. Wu, W.W., Hu, D., Cong, W.X., Shan, H.M., Wang, S., Niu, C., Yan, P.K., Yu, Y.Y., Vardhanabhuti, V., and Wang, G. (2020). Stabilizing deep tomographic reconstruction. Preprint at arXiv. 2008.01846.
50. Cong, W.X., Wang, G., Yang, Q.S., Li, J., Hsieh, J., and Lai, R. (2019). CT image reconstruction on a low dimensional manifold. Inverse Probl. Imaging *13*, 449–460. https://doi.org/10.3934/ipi.2019022.
51. Tschuchnig, M.E., Oostingh, G.J., and Gadermayr, M. (2020). Generative adversarial networks in digital pathology: a survey on trends and future potential. Patterns *1*, 100089. https://doi.org/10.1016/j.patter.2020.100089.
52. Lv, J., Wang, C., and Yang, G. (2021). PIC-GAN: a parallel imaging coupled generative adversarial network for accelerated multi-channel MRI reconstruction. Diagnostics *11*, 61. https://doi.org/10.3390/diagnostics11010061.
53. Yuan, Z., Jiang, M., Wang, Y., Wei, B., Li, Y., Wang, P., Menpes-Smith, W., Niu, Z., and Yang, G. (2020). SARA-GAN: self-attention and relative average discriminator based generative adversarial networks for fast compressed sensing MRI reconstruction. Front. Neuroinformatics *14*, 611666. https://doi.org/10.3389/fninf.2020.611666.
54. Chan, C.C., and Haldar, J.P. (2021). Local perturbation responses and checkerboard tests: characterization tools for nonlinear MRI methods. Magn. Reson. Med. *86*, 1873–1887. https://doi.org/10.1002/mrm.28828.
55. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017). GANs trained by a two time-scale update rule converge to a local nash equilibrium. In NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, Ulrike Luxburg, ed. (Curran Associates Inc), pp. 6629–6640. https://doi.org/10.5555/3295222.3295408.
56. Pan, J.Y., Wu, W.W., Gao, Z.F., and Zhang, H.Y. (2021). Multi-domain integrative Swin transformer network for sparse-view tomographic reconstruction. Preprint at arXiv. 2111.14831.
57. Huang, J., Fang, Y., Wu, Y., Wu, H., Gao, Z., Li, Y., Del Ser, J., Xia, J., and Yang, G. (2022). Swin transformer for fast MRI. Preprint at arXiv. 2201.03230.
58. Seitzer, M., Yang, G., Schlemper, J., Oktay, O., Würfl, T., Christlein, V., Wong, T., Mohiaddin, R., Firmin, D., and Keegan, J. (2018). Adversarial and perceptual refinement for compressed sensing MRI reconstruction. In International conference on medical image computing and computer-assisted intervention (Springer).
59. Colbrook, M.J., Antun, V., and Hansen, A.C. (2009). On the existence of stable and accurate neural networks for image reconstruction. https://www.mn.uio.no/math/english/people/aca/vegarant/nn_stable.pdf.
60. Antun, A., Colbrook, M.J., and Hansen, A.C. (2021). Can stable and accurate neural networks be computed?–On the barriers of deep learning and Smale's 18th problem. Preprint at arXiv. 2101.08286.

61. Katsevich, A. (2002). Analysis of an exact inversion algorithm for spiral cone-beam CT. Phys. Med. Biol. *47*, 2583–2597. https://doi.org/10.1088/0031-9155/47/15/30.

62. Axel, L., Summers, R., Kressel, H., and Charles, C. (1986). Respiratory effects in two-dimensional Fourier transform MR imaging. Radiology *160*, 795–801. https://doi.org/10.1148/radiology.160.3.3737920.

63. Candes, E.J., and Tao, T. (2005). Decoding by linear programming. IEEE Trans. Inf. Theor. *51*, 4203–4215. https://doi.org/10.1109/TIT.2005.858979.

64. Yu, H.Y., and Wang, G. (2010). A soft-threshold filtering approach for reconstruction from a limited number of projections. Phys. Med. Biol. *55*, 3905–3916. https://doi.org/10.1088/0031-9155/55/13/022.

65. Zhang, L., Zhang, L., Mou, X.Q., and Zhang, D. (2011). FSIM: a feature similarity index for image quality assessment. IEEE Trans. Image Process. *20*, 2378–2386. https://doi.org/10.1109/TIP.2011.2109730.

66. Ma, J., and März, M. (2016). A multilevel based reweighting algorithm with joint regularizers for sparse recovery. Preprint at arXiv. 1604.06941.

67. Knoll, F., Clason, C., Bredies, K., Uecker, M., and Stollberger, R. (2012). Parallel imaging with nonlinear reconstruction using variational penalties. Magn. Reson. Med. *67*, 34–41. https://doi.org/10.1002/mrm.22964.

Patterns, Volume *3*

Supplemental information

Stabilizing deep tomographic reconstruction: Part A. Hybrid framework and experimental results

Weiwen Wu, Dianlin Hu, Wenxiang Cong, Hongming Shan, Shaoyu Wang, Chuang Niu, Pingkun Yan, Hengyong Yu, Varut Vardhanabhuti, and Ge Wang

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Table of Contents

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

I. Deep Networks & Datasets

I.A. Ell-50 Network

I.A.1. Narrative. Ell-50 is a special form of FBPConvNet, which is a classic neural network for CT imaging proposed in Ref. 1. The FBPConvNet with multiple-solution decomposition and residual learning[2] was proposed to remove sparse-data artifacts and preserve image features and structures. The reconstruction performance of the FBPConvNet was validated, outperforming the total variation-regularized iterative reconstruction using the realistic phantoms. Besides, it was very fast to reconstruct an image on GPUs. In this study, the training dataset mainly contains ellipses with different intensities, sizes and locations. The network is named Ell-50, indicating that the measurements were collected from 50 different views. This network was trained by the authors of Ref. 1, which can be freely downloaded (https://github.com/panakino/FBPConvNet).

I.A.2. Network Architecture. The Ell-50 network was trained to reconstruct $\boldsymbol{f}$ from measurements $\boldsymbol{p} = \boldsymbol{A}\boldsymbol{f}$, where $\boldsymbol{A}$ represents a subsampling system matrix, with which only 50 uniformly spaced radial lines are collected. Because the FBPConvNet is an image post-processing network, it is trained from filtered backprojection (FBP) reconstruction images rather than directly learning a mapping from $\boldsymbol{p}$ to $\boldsymbol{f}$. The network first employs FBP to convert $\boldsymbol{p}$ to $\widehat{\boldsymbol{f}} = \boldsymbol{A}^{+}\boldsymbol{p}$ where $\mathbf{A}^{+}$ represents the FBP and is considered as the first layer of the neural network.

The FBPConvNet is a useful model based on U-Net[3], which is considered as an encoder-decoder pair. The main features of U-Net based FBPConvNet are summarized as the following three features: multilevel decomposition, multichannel filtering, and skip connections (including concatenation operator and residual learning). The network input is an image with 512 ×512 pixels, where it is first down-sampled 4 times for encoding, and then the resultant low-dimensional image features are up-sampled to 512 ×512 pixels. Besides, the skip concatenation operator is employed in this network. The Ell-50 network consists of convolutional and deconvolutional layers, and each convolutional and deconvolutional layer is followed by batch normalization (BN) and ReLU layers. The sizes of filters and stride in the Ell-50 network were set to 3×3 and 1×1, respectively. Moreover, the Ell-50 network details are shown in Fig. S1.

I.A.3. Network Training. The few-view and full-view FBP images are treated as the input and ground-truth of the Ell-50, respectively. In this study, the network was implemented using the MatConvNet[4] toolbox with a slight modification to train and evaluate the performance. To prevent the divergence of the cost function, the MatConvNet[4] toolbox was slightly modified by clipping the computed gradients to a fixed range[5]. In this study, we only used the pre-trained network weights of Ref. 1 that were publicly available at GitHub (https://github.com/panakino/FBPConvNet). Such a configuration is consistent with the literature [6]. The loss function plays an important role in controlling the image quality, and the mean square error (MSE) between the network output and the ground truth is considered in Ell-50. Since the employed network was performed on a TITAN Black GPU graphic processor (NVIDIA Corporation), the

total training time took about 15 hours with 101 epochs[6]. Regarding the learning rate, it was decreased logarithmically from 0.01 to 0.001. Besides, the batch size, momentum, and clipping value were set to 1, 0.99 and 10^{-2}, respectively. For the Ell-50 network, it was implemented in MATLAB with the MatConvNet platform based on Window 10 system with one NVIDIA TITAN XP graphics processing units (GPUs) installed on a PC (16 CPUs @3.70GHz, 32.0GB RAM and 8.0GB VRAM).

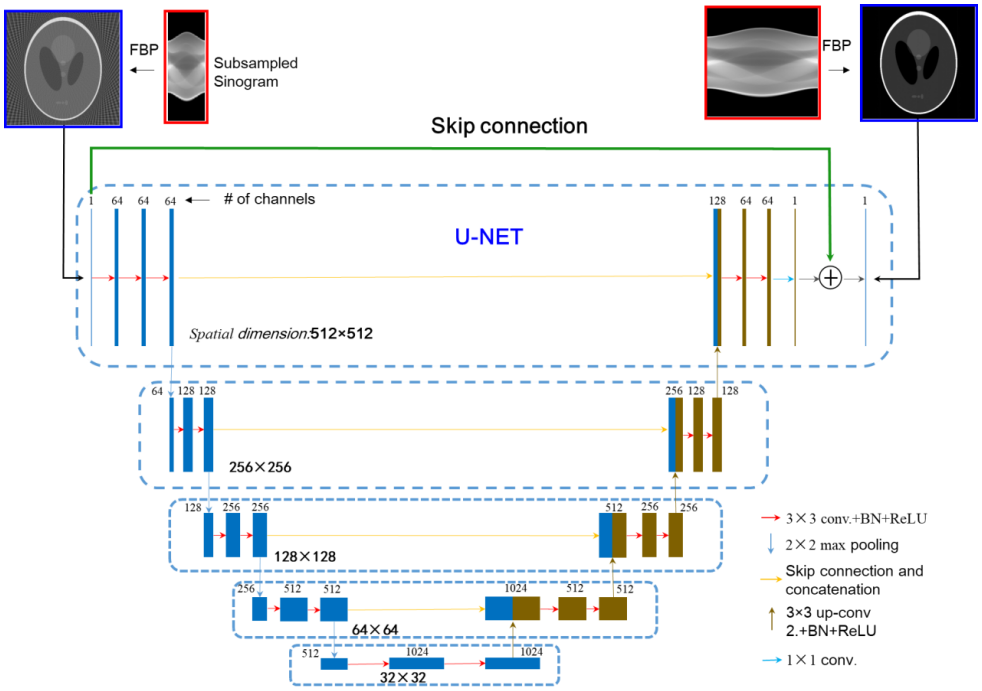


Fig. S1. Architecture for Ell-50.

I.A.4. Training Data. Regarding the training dataset, the number of training images is 475. The training images were reconstructed via FBP with sparse-view measurement. The dynamic range of the reconstructed images was controlled to the range of [−500, 500] HU. Since only ellipses with different sizes, locations and intensities were simulated, the projections were accurately and analytically computed[7]. The scanning geometry was set to produce parallel beams[8]. The number of full projections and the number of detector units were set to 1,000 and 729, respectively. Especially, the functions of *radon* and *iradon* in MATLAB were employed to realize the projection and backprojection operations. For sparse-data reconstruction, only 50 views were extracted from full projections, and then FBP reconstructed images were input to the selected networks in this paper. This case is a typical sparse-view reconstruction[9-11]. The ground truths are FBP images from full projections (i.e., 1000 views).

I.A.5. Test Data. To demonstrate the instability of neural network (i.e., Ell-50), the additional symbol “ ” and the text “CAN YOU SEE IT” were first embedded in the original image, which was provided by the authors of Ref. 6. These artificial features were to mimic the structure changes and further validate the instability of the neural network in this case (see Fig. S2). In this study, the image with the symbol “ ” and text “CAN YOU SEE IT” was also treated as case C1 to validate the instability of Ell-50 and the stability of our proposed ACID

method. Besides, a slightly complicated phantom with the inserted logo of a bird and text “A BIRD?” was provided by the authors of Ref. 12 and downloaded from Ref. 6, which is defined as case C2. The test image consists of 512×512 pixels, and it contains structural features without tiny perturbation. To generate adversarial attacks, the proposed method in [6] was employed to induce tiny perturbations. For case C3, an original image was randomly selected from the test datasets of https://github.com/panakino/FBPConvNet, which contains no perturbation. Here, the tiny perturbation is added to the original image with the same technique used in Ref. 6, and then we obtained the case C3. Regarding case C4, the same technique used in Ref. 6 was employed to generate the tiny perturbation and then embedded into case C1 to obtain C4. Furthermore, a Gaussian noise image with zero mean and standard deviation of 15 in HU over the pixel value range was superimposed to case 1 to obtain C5 image. To validate the ability of ACID against adversarial attacks, the adversarial samples (see Section III in this supplementary information for details) for the whole ACID were generated and added into the C3 and C1 images respectively to obtain the cases of C6 and C7. The searched adversarial attacks in the whole ACID flowchart are greater than those used in a single neural network (i.e., ell-50) in terms of L_2-norm.

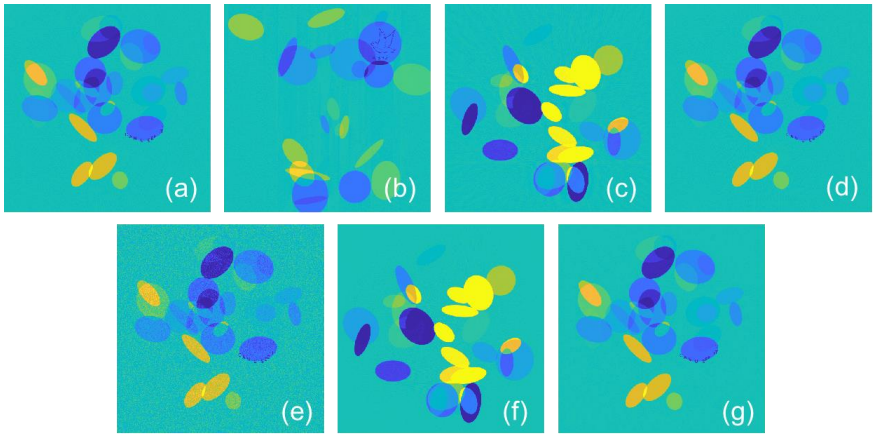


Fig. S2. Test images used to validate the effectiveness of ACID in stabilizing ell-50 for CT study. (a)-(g) correspond to the C1-C7 cases, respectively. The display window is [-150, 150] HU.

I.B. DAGAN Network

I.B.1. Narrative. The DAGAN network is to reduce aliasing artifacts with the U-Net[3] based generator[13]. To enhance the ability of the reconstruction method in preserving image texture and edges, DAGAN incorporates an innovative content loss and adversarial loss. Besides, it also introduces frequency-domain features to encourage coherence in image and frequency domains. Compared with the traditional CS-based and some other deep learning methods[14-16], the DAGAN method achieved superior performance in retaining image details. Besides, as one of the post-processing methods, the speed of DAGAN reconstruction is very fast. In this study, the DAGAN network was tested on a single-coil MRI with 10% and 20% subsampling rates. The trained weights are not available online, however, the authors of Ref. 13 provide the implementation details of DAGAN. With this help, we retrained the DAGAN with different subsampling rates and masks. The architecture, training parameters, and test data are summarized in the following subsections.

I.B.2. Network Architecture. The DAGAN[13] network was proposed for fast MRI reconstruction from subsampled measurement data. In the case of DAGAN, the measurement data is $\boldsymbol{p} = \boldsymbol{Af}$, where $\boldsymbol{A}$ is the subsampled discrete Fourier transform. The aim of DAGAN is to recover $\boldsymbol{f}$ from the degraded image $\hat{\boldsymbol{f}}$ that is reconstructed directly via inverse Fourier transform from the zero-filled measurement data.

To restore high-quality MR images from measurement, DAGAN adopted a conditional generative adversarial network (GAN)[17-19] model. It consists of two modules: generator and discriminator. The generator is to recover the image, and the discriminator is to distinguish the recovered image and the ground-truth. The goal is to make the discriminator fail, and hence improve the recovered image quality. The authors of Ref. [13] provided three variants of DAGAN, and we selected the full model version (Pixel-Frequency-Perceptual-GAN-Refinement) in our experiments. According to parameter settings in Ref. 13 and the codes provided by the authors of Ref. 6, we retrained the DAGAN.

The architecture of the generator is illustrated in Fig. S3. It adopted the basic U-Net type structure, which contains 8 convolutional layers and 8 deconvolutional layers. All of them are followed by batch normalization layers to accelerate training convergence and overcome overfitting. The leaky ReLU layers are adopted as the activation function with a slope equal to 0.2 when the input is less than 0. Additionally, skip connections are employed to concentrate on encoder and decoder features to gain reconstruction details and promote the information flow. The hyperbolic tangent function is applied as the activate function for the output of the last convolutional layer. Then a global skip connection, adding the input data and the output of the hyperbolic tangent function together, is then clipped by a ramp function to scale the output of the generator to the range [-1,1]. The global skip connection can accelerate the training convergence and improve the performance of the network. The DAGAN network architecture was shown in Fig. S3. For more detailed information on the DAGAN network, please refer to Ref. 13.

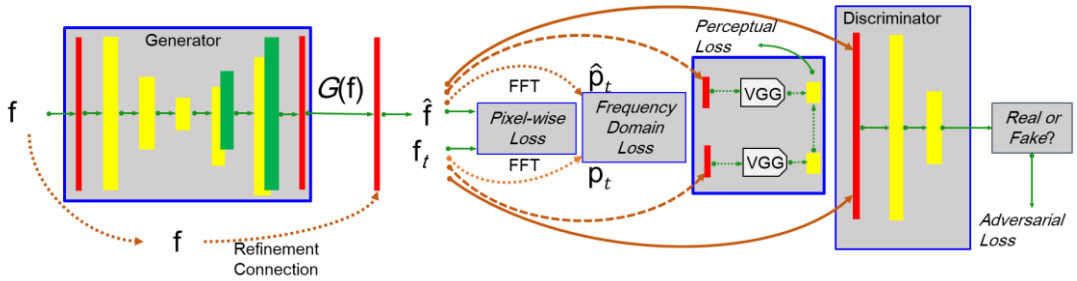


Fig. S3. The architecture of the DAGAN network.

I.B.3. Training Parameters. The loss function of DAGAN is formulated as follows:

$$L_{DAGAN} = \sigma_1 L_{Img} + \sigma_2 L_{frq} + \sigma_3 L_{VGG} + \sigma_4 L_D \;\; s.t. \sigma_1, \sigma_2, \sigma_3, \sigma_4 > 0\,, \quad \text{(S.1.1)}$$

where L_{Img} computes the Euclidean distance in the image domain between the generated image and ground truth, and L_{frq} accounts for the counterpart in the k-space. To constrain the similarity loss L_{VGG} in the feature space, the trained VGG-16 was used to optimize the L_2-distance between feature maps of the generated image and ground truth, which is the same as Ref. 20. In particular, the feature maps generated of the conv4 layer in VGG-16 were used to

calculate L_{VGG}. Last, L_D is the adversarial loss using a cross entropy to make the generated image more realistic. $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ are the hyper-parameters to balance different constraint terms. According to Ref. 13, they were set to 15, 0.1, 0.0025 and 1, respectively. The generator and the discriminator were optimized using the Adam algorithm[21] with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. Specifically, the learning rate was initially set to 0.0001, which was decreased half every 5 epochs, and the batch size is 25. To prevent overfitting, an early stopping strategy was adopted via measuring the loss L_{frq} on the validation set, and the stopping number was set to 10.

I.B.4. Training Data. The datasets for training the DAGAN network were provided by the MICCAI 2013 Grand Challenge and are publicly available in https://my.vanderbilt.edu/masi/workshops/. More details about the training datasets are given in https://github.com/tensorlayer/DAGAN. Specifically, to exclude the negative influence on the DAGAN network, all the images that have more *T*% background pixels were dropped. In our experiments, the threshold *T* was set to 90. After data preprocessing, there are 15,912 images for training and 4,977 images for testing. All the images are T1-weighted brain MR images. Again, the data augmentation methods were applied to eliminate overfitting, including image flipping, rotation, shifting, and so on[13].

In the experiments, the DAGAN is to recover images from 10% subsampling rate using a 2D Gaussian mask and the radial mask of a 20% subsampling rate, respectively. Two models of the DAGAN network were trained for these two subsampling masks. All the codes were implemented with TensorLayer and Tensorflow frameworks[13].

I.B.5. Testing Data. To test the robustness of DAGAN in terms of small structural changes, adversarial attacks and noise, the symbols "HELLO NATURE" and "CAN YOU SEE IT" were embedded in two different original images, which are denoted as Cases M1 and M2, respectively. Specifically, the image with the symbol "CAN YOU SEE IT" was provided by the authors of Ref. 6 (download in https://github.com/vegarant/Invfool). The original image with the symbol "HELLO NATURE" was produced (downloaded from https://github.com/tensorlayer/DAGAN). In cases M1 and M2, there are two test images used to demonstrate the instability of the DAGAN network with respect to small structural changes. Next, to explore the performance of the DAGAN network in terms of adversarial attacks and small structural changes, the tiny perturbations derived from Ref. 6 were added into cases M1 and M2 to generate cases M3 and M4. Last, to test the DAGAN network in terms of anti-noising, the noise was superimposed to cases M1 and M2 to obtain cases M5 and M6. In our ablation study of ACID, we randomly selected one original image as M7 from the DAGAN test dataset (https://github.com/tensorlayer/DAGAN). Furthermore, cases M8 and M9 were generated by applying the radial mask of a 20% subsampling rate on the M1 and M2 images, which were used to compare the performances between ACID and the classic Alternating Direction Method of Multipliers (ADMM)-net[22]. Regarding the stability of ACID, the tiny perturbations from ACID were added into M7, M1 and M2, and then the images with tiny perturbations were marked as M10, M11 and M12. The tiny perturbations from M11 and M12 are greater than the perturbations within M3 and M4 in terms of the L_2-norm. Except for M8 and M9, all the rest of the images were recovered from the k-space data collected at a 10% subsampling rate

using the Gaussian mask. All the images from the references of M1-M12 are shown in Fig. S4.

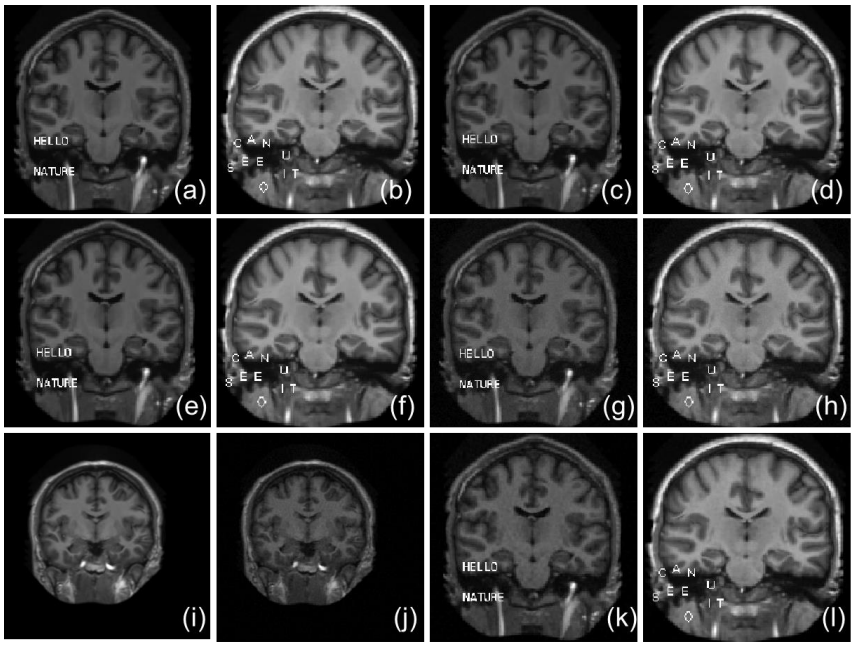


Fig. S4. Test images for showing the instability of neural networks. (a)-(l) correspond to the M1-M12 cases, respectively.

I.C. AUTOMAP Network

I.C.1. Narrative. Proposed as a framework for image reconstruction, the automated transform by manifold approximation (AUTOMAP) transfers sensor data to a high-quality image with a mapping function between the sensor and image domains[23]. The AUTOMAP demonstrated its advantages in various magnetic resonance imaging acquisition modes using the same architecture and hyperparameters. In this study, the AUTOMAP neural network was tested on the single-coil MRI with subsampled data. The trained ATUOMAP used in our experiments is provided by Ref. 6. The architecture, training details, and test data of AUTOMAP are in the following sub-sections.

I.C.2. Network Architecture. The AUTOMAP[23] presents a framework for image reconstruction by translating sensor-domain signals into the image domain directly via domain-transform manifold learning. For MRI reconstruction, four subsampling strategies were applied to access the performance of the AUTOMAP, which are Radon projection, spiral non-Cartesian Fourier, under-sampled Cartesian Fourier, and misaligned Fourier.

The AUTOMAP network takes a vectorized measurement data as input which is sub-sampled from the full-sampled k-space data. First, we can obtain the complex k-space data using the discrete Fourier transform on the MR images. Then, the subsampled k-space data are generated via a subsampling mask. Next, these measurement data are reshaped into vectors. Last, the vectorized measurement data are fed into the AUTOMAP network. In this paper, the images with the size of 128×128 and 60% subsampling rate are tested for MRI reconstruction. There are two fully connected layers in the AUTOMAP network, which have 25,000 and 128x128 nodes, respectively. The activate function of the first fully connected layer is the hyperbolic tangent function, and

the output of the second fully connected layer then subtracts the mean value of itself. Next, it is reshaped into a feature map with the same size as the reconstructed image. Furthermore, two convolutional layers are applied to extract essential features from their input data. Each of them contains 64 filters with a size of 5×5 and the stride of 1×1. The activation function of the first convolution layer is a hyperbolic tangent function and the other is rectified linear unit (ReLU). The last convolutional layer has one filter with the size of 7×7 and a stride of 1×1. The output of the network is the corresponding reconstruction image. The trained weights were provided by the authors of Ref. 6.

I.C.3. Training Parameters. The whole optimization problem of the AUTOMAP is defined as follow:

$$L_{AUTOMAP} = L_{rec} + \lambda_1 L_{fea} \ . \tag{S.1.2}$$

The loss function of AUTOMAP $L_{AUTOMAP}$ consists of two terms, i.e., L_{rec} and L_{fea}. L_{rec} is employed to evaluate the Euclidean distance between the predicted image provided by the AUTOMAP network and the ground-truth image. L_{fea} is ℓ_1-norm to constrain the feature maps produced by the activation function of the second convolutional layer. $\lambda_1 > 0$ is to balance the two terms. The total loss function is optimized by the RMSProp algorithm with momentum 0 and decay 0.9[23]. The learning rate is 0.00002 and the batch size is 100. The network was trained and stopped after 100 epochs.

I.C.4. Training Data. Selected in the MGH-USC HCP public dataset (http://www.humanconnectomeproject.org/data/), there are 50,000 images from 131 subjects in total. Specifically, the training images are 128×128 matrices, which were subsampled from the central part cropped from the original image. Meanwhile, all the training datasets were scaled to a given range. In the Fourier space, the subsampled measurement data were produced by a Poisson-disk mask of a 60% subsampling rate.

To improve the generalization ability of the AUTOMAP network, the data augmentation strategy was applied. 1.0% multiplicative noise was added to the input to promote manifold learning during the course of network training, and it is beneficial for the trained network learning robust representations from corrupted inputs. In fact, the specific additional noise distribution of the corruption process is not subject to the additive Gaussian noise during the process of evaluation. The corresponding training datasets with the size of 128×128 are cropped from original MR images by using four types of reflections. All the related codes were implemented in the TensorFlow framework[6].

I.C.5. Testing Data. To validate the instability of the AUTOMAP network, the symbol "♡" was first added to the original MR image, which was also provided by the author [6]. This simple symbol was used to simulate small structural changes in the patient and then test the instability of the AUTOMAP network reconstruction. All the test data were downloaded from Ref. 6. In addition, the "HELLO NATURE", "CAN U SEE IT" and "♢" were added to the original test image to generate A1 and A2 with the structural changes. The resultant tiny perturbations were added to A1 and A2 images to obtain A3 and A4 images (see Fig. S5).

I.C.6. Reconstruction Results. Here, to demonstrate the advantages of ACID, a typical reconstruction network, AUTOMAP, was selected as an example, and

the reconstruction results of Fig. S5 (a) are in Fig. S6. As shown in Fig. S6, ACID produced significantly better image quality than AUTOMAP. The PSNR was improved by ACID to 36.0 dB, well above 27.8 dB of AUTOMAP. Also, the SSIM of ACID reached 0.971, while the counterpart of AUTOMAP was 0.730. It further demonstrates that ACID achieves better image quality than AUTOMAP. The related reconstruction results of A1-A4 are in the main body of the paper.

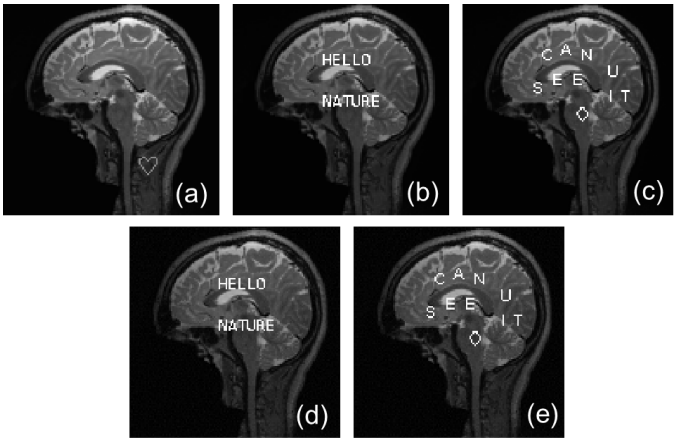


Fig. S5. Test images with a structural change and tiny perturbations for evaluation of the AUTOMAP stability. (a) was the test image provided in Ref. 6, and (b)-(e) represent the test images of A1-A4.

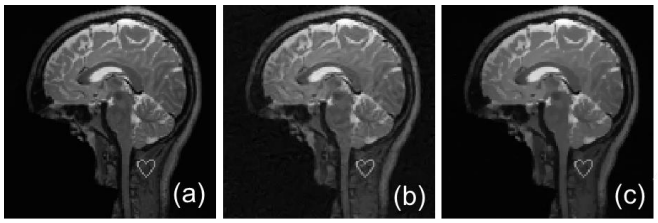


Fig. S6. ACID deep reconstruction with the embedded AUTOMAP network. (a) represents the original brain phantom, (b) and (c) represent the reconstructed results by AUTOMAP and ACID respectively.

I.D. ADMM-net

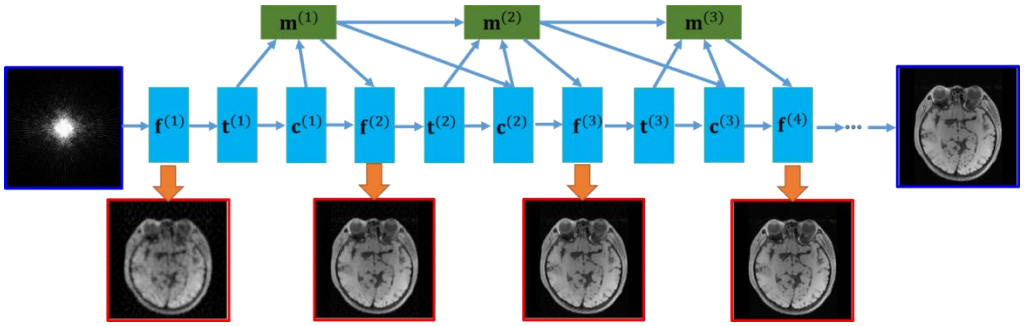


Fig. S7. The flowchart of the ADMM-net. $\boldsymbol{f}^{(k)}$, $\boldsymbol{c}^{(k)}$, $\boldsymbol{t}^{(k)}$ and $\boldsymbol{m}^{(k)}$ represent the construction layer, convolution layer, nonlinear transform layer, and multiplier update layer in the k-th stage.

I.D.1 Narrative. Inspired by the traditional Alternating Direction Method of Multipliers (ADMM) iterative optimization algorithm for CS-based MRI[24], the ADMM-net defined over a data flow graph was first proposed in Ref. 22. Regarding the training procedure, the network parameters (e.g. image shrinkage functions, transforms) are trained into an end-to-end architecture using the L-BFGS algorithm[25]. Regarding the testing step, it needs a similar

computational overhead with the ADMM. However, there is only one parameter to be chosen initially in the ADMM-net since others are automatically learned during the training step. The superior experiments on MRI image reconstruction demonstrate the advantages over fast MRI imaging and higher image quality. In this study, the ADMM-net is tested on the single-coil MRI with 20% subsampling and the trained weights are provided by the authors of Ref. 22. The architecture, training parameters and test data of ADMM-net are summarized in the following sections. The workflow of ADMM-net is given in Fig. S7.

I.D.2. Network Architecture. ADMM-net[22] is a classical unrolled iterative optimization algorithm for MRI reconstruction. Different from the traditional compressed sensing (CS) based methods[26] and data-driven based methods, ADMM-net can be trained end-to-end by incorporating a physic-guided model, and it achieves excellent performance in MR imaging with much less computational cost. The ADMM-net is derived from the ADMM algorithm via solving the sub-problem with deep learning networks. The CS-MRI model can be described as:

$$\underset{\boldsymbol{f}}{\operatorname{argmin}} \frac{1}{2}\left\|\boldsymbol{A}\boldsymbol{f}-\boldsymbol{p}^{(0)}\right\|_F^2+\sum_{l=1}^{L}\lambda_l g(D_l(\boldsymbol{f})) \tag{S.1.3}$$

where $\boldsymbol{f}\in C^N$ is the MR image to be reconstructed, $\boldsymbol{p}^{(0)}\in C^H$ denotes the under-sampled measurement data, $\boldsymbol{A}$ is the Fourier translation based system matrix with an under-sampled mask, D_l represents the transform operation, g is the regularization function, and λ_l is the regularization parameter. By introducing $t_l=D_l(\boldsymbol{f})$, $l=1,\dots,L$, (S.1.3) is converted into the following constraint optimization problem:

$$\underset{\boldsymbol{f},\{\mathrm{t}_l\}_{l=1}^{L}}{\operatorname{argmin}} \frac{1}{2}\left\|\boldsymbol{A}\boldsymbol{f}-\boldsymbol{p}^{(0)}\right\|_F^2+\sum_{l=1}^{L}\lambda_l g(\mathrm{t}_l),\quad \mathrm{t}_l=D_l(\boldsymbol{f}), l=1,\dots L. \tag{S.1.4}$$

(S.1.4) is a constraint programming procedure and it can be further converted into the following unconstraint problem

$$\underset{\mathbf{f},\{\mathrm{t}_l\}_{l=1}^{L},\{\alpha_l\}_{l=1}^{L}}{\operatorname{argmin}} \frac{1}{2}\left\|\boldsymbol{A}\boldsymbol{f}-\boldsymbol{p}^{(0)}\right\|_F^2+\sum_{l=1}^{L}\lambda_l g(\mathrm{t}_l)-\sum_{l=1}^{L}\langle \mathrm{t}_l-D_l(\boldsymbol{f}),\alpha_l\rangle+\frac{1}{2}\sum_{l=1}^{L}\gamma_l\|\mathrm{t}_l-D_l(\boldsymbol{f})\|_F^2, \tag{S.1.5}$$

where α_l $(l=1,\dots,L)$ are Lagrange multipliers and γ_l $(l=1,\dots,L)$ are the corresponding penalty parameters. (S.1.5) can be solved using the ADMM algorithm [27] as the following three sub-problems:

$$\boldsymbol{f}^{(k+1)}=\underset{\boldsymbol{f}}{\operatorname{argmin}} \frac{1}{2}\left\|\mathbf{A}\mathbf{f}-\boldsymbol{p}^{(0)}\right\|_F^2-\sum_{l=1}^{L}\langle \mathrm{t}_l^{(k)}-D_l(\mathbf{f}),\alpha_l^{(k)}\rangle+\frac{1}{2}\sum_{l=1}^{L}\gamma_l\left\|\mathrm{t}_l^{(k)}-D_l(\boldsymbol{f})\right\|_F^2, \tag{S.1.6}$$

$$\mathrm{t}_l{}^{(\mathrm{k}+1)}=\underset{\{\mathrm{t}_l\}_{l=1}^{L}}{\operatorname{argmin}} \lambda_l g(\mathrm{t}_l)-\langle \mathrm{t}_l-D_l\left(\boldsymbol{f}^{(\mathrm{k}+1)}\right),\alpha_l\rangle+\frac{1}{2}\gamma_l\left\|\mathrm{t}_l-D_l\left(\boldsymbol{f}^{(\mathrm{k}+1)}\right)\right\|_F^2,\ l=1,\dots,L, \tag{S.1.7}$$

$$\alpha_l{}^{(\mathrm{k}+1)}=\alpha_l{}^{(\mathrm{k})}+\mathrm{t}_l{}^{(\mathrm{k}+1)}-D_l\left(\boldsymbol{f}^{(k+1)}\right), l=1,\dots,L. \tag{S.1.8}$$

Finally, these three sub-problems can be updated iteratively using deep neural blocks. Regarding the ADMM-net, the above optimization with one separate variable update can be generalized as four type layers: reconstruction layer ($\boldsymbol{f}^{(k+1)}$), convolutional layer ($\{D_l\left(\boldsymbol{f}^{(k+1)}\right)\}_{l=1}^{L}$), non-linear layer ($\{t_l{}^{(\mathrm{k}+1)}\}_{l=1}^{L}$), and multiplier update layer ($\{\alpha_l{}^{(\mathrm{k}+1)}\}_{l=1}^{L}$). More details related to the construction and organization of the layers in ADMM-net can be referred to Ref. 22. The

ADMM-net takes the sub-sampled k-space data as the input and finally generates the reconstructed image through the iterative process.

I.D.3. Training Parameters. The ADMM-net adopts the normalized mean square error (NMSE) as the loss function to optimize the neural network. The image reconstructed from fully-sampled k-space data was used as the reference image and the corresponding under-sampling data in the k-space was used as the input. The loss function is defined as

$$L_{NMSE} = \frac{1}{N_1}\sum_{n_1=1}^{N_1}\frac{\sqrt{\|\hat{\boldsymbol{f}}_{n_1}(\Theta)-\boldsymbol{f}_{n_1}\|_F^2}}{\sqrt{\|\boldsymbol{f}_{n_1}\|_F^2}}, \quad \text{(S.1.9)}$$

where $\hat{\boldsymbol{f}}_{n_1}$ and $\boldsymbol{f}_{n_1}$ are the generated image from ADMM-net and the reference image (as the label), respectively. N_1 is the number of training samples. Θ denotes parameters needed to be optimized in ADMM-net. The L-BFGS algorithm was used to minimize the loss function L_{NMSE}.

I.D.4. Training Data. The ADMM-net is trained with brain and chest MR image datasets (https://my.vanderbilt.edu/masi/workshops/). For each dataset, 100 images were randomly selected for training and 50 images for testing. In our experiments, all the under-sampled k-space data were generated with the radial mask of a 20% subsampling rate, as shown in Fig. S8. All the codes are in MATLAB with Intel core i7-4790k CPU, and the training and testing datasets were downloaded from https://github.com/yangyan92/Deep-ADMM-Net.

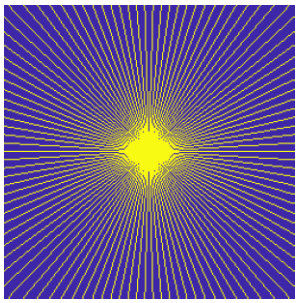


Fig. S8. Radial sampling mask of a 20% subsampling rate.

I.D.5. Testing Data. To validate the performance of ADMM-net about small structural changes, the same images as those in Fig. 2 in the main text of the paper with the symbols “CAN YOU SEE IT” and “HELLO NATURE” were employed with the radial mask of a 20% subsampling rate.

II. CS Based Reconstruction Methods

II.A. CS-inspired Reconstruction

II.A.1. Narrative. To demonstrate the advantages of our ACID in terms of stability against the benchmark compressed sensing (CS)-based methods[28-31], the related experiments are performed, and the reconstruction results are provided using the established methods[32-34]. Since the total variation minimization (individual or combination) is popular in the image reconstruction field with consideration of sparsity prior, it is respectively chosen for CT and MRI image reconstruction in this study. The specific details are given as follows.

II.A.2. X-ray CT Reconstruction. The re-weighting technique[35] combining both shearlets[36] and TV[37] was proposed to validate the stability in Ref. 6. In this study,

it served as a state-of-the-art CS-based comparison method for CT. The details can be found in Refs. 37 and 6. Here we only provide a brief summary as follows.

The mathematical model for this method is formulated as

$$\underset{\boldsymbol{f}}{argmin}\frac{\omega}{2}\|\boldsymbol{Af}-\boldsymbol{p}\|_F^2+\sum_{j=1}^{J}\vartheta_j\left\|W_j\psi_j\boldsymbol{f}\right\|_1+TGV_{\varrho}(\boldsymbol{f}), \tag{S.2.1}$$

where ϑ_j represents the *j*-th balance factor, W_j is a diagonal matrix, and ψ_j represents the *j*-th subband from the corresponding shearlet transformation. The $\mathrm{TGV}_{\varrho}(\mathbf{f})$ stands for the total generalized variation with the parameter ϱ. $\mathrm{TGV}_{\varrho}(\mathbf{f})$ is combined with the components from the first and second orders of the total variation of the reconstructed image. Furthermore, the parameters ϱ is introduced to balance these two terms. $\omega>0$ is to balance data fidelity and regularization of sparsity prior.

To solve the optimization problem (S.2.1), $\boldsymbol{d}=\psi'\boldsymbol{f}$ is introduced to represent the matrix format of $\sum_{j=1}^{J}\psi_j\boldsymbol{f}$ and (S.2.1) is split into three sub-problems:

$$\left\{\boldsymbol{f}^{(k+1)},\boldsymbol{d}^{(k+1)}\right\}=\underset{\boldsymbol{f},\boldsymbol{d}}{argmin}\frac{\omega}{2}\left\|\boldsymbol{Af}-\boldsymbol{y}^{(k)}\right\|_F^2+\|W\boldsymbol{d}\|_1+\frac{\omega_1}{2}\left\|\boldsymbol{d}-\psi'\boldsymbol{f}-\boldsymbol{b}^{(k)}\right\|_F^2+TGV_{\varrho}(\boldsymbol{f}), \tag{S.2.2}$$

$$\boldsymbol{b}^{(k+1)}=\boldsymbol{b}^{(k)}+\psi'\boldsymbol{f}^{(k+1)}-\boldsymbol{d}^{(k+1)}, \tag{S.2.3}$$

$$\boldsymbol{y}^{(k+1)}=\boldsymbol{y}^{(k)}+\boldsymbol{p}-\boldsymbol{Af}^{(k+1)}. \tag{S.2.4}$$

where *W* is the matrix format of $\vartheta_j W_j$. In (S.2.2)-(S.2.4), the four variables are updated iteratively. First, the minimization problem in (S.2.2) is optimized utilizing the multiple non-linear block Gauss-Seidel iterations[38]. Compared with the original re-weighting strategy[35], the weights in *W* are not only updated after convergence to the solution of (S.2.2), but also are put into the following split process. This unique weight updating strategy is further described in Ref. 6. In this work, the same strategy and configuration[6] were used (including the parameters, the number of iterations, etc.). Note that the number of iterations and the regularization parameters can be further optimized.

II.A.3. MRI Reconstruction. By extending the iteratively regularized Gauss–Newton method (IRGN) with variational penalties[39,40], the total generalized variation (TGV) based IRGN (IRGN-TGV) was proposed[41], and better reconstruction quality was achieved by combining estimation of image and coil sensitivities with TGV regularization. Indeed, the IRGN-TGV had superior noise suppression because of the TGV regularization. In addition, the IRGN-TGV approach can remove sampling artifacts arising from pseudorandom and radial sampling patterns. In this study, it was employed as a state-of-the-art to perform CS-based MRI experiments. Here we also give a brief summary of this method.

Mathematically, MRI is a typical inverse problem with the sampling operator $\boldsymbol{A}$ and the correspondingly k-space data $\boldsymbol{p}$ from the receivers. Besides, the spin density is given as $\boldsymbol{h}$, and $\boldsymbol{c}$ represents the unknown set of coil sensitivities. For the current iteration index k with the given $\boldsymbol{f}^{(k)}:=\left(\boldsymbol{h}^{(k)},\boldsymbol{c}^{(k)}\right)$, the solution $\Delta\boldsymbol{f}:=(\Delta\boldsymbol{h},\Delta\boldsymbol{c})$ is sought to minimize the following objective function:

$$\underset{\Delta\boldsymbol{f}}{argmin}\frac{1}{2}\left\|\boldsymbol{A}'\left(\boldsymbol{f}^{(k)}\right)\Delta\boldsymbol{f}+\boldsymbol{A}\left(\boldsymbol{f}^{(k)}\right)-\boldsymbol{y}\right\|_F^2+\frac{\alpha_k}{2}\left\|W_1\left(\boldsymbol{c}^{(k)}+\Delta\boldsymbol{c}\right)\right\|_1+\rho_k R\left(\boldsymbol{h}^{(k)}+\Delta\boldsymbol{h}\right). \tag{S.2.5}$$

Given $\alpha_k>0$, $\rho_k>0$, we have $\boldsymbol{f}^{(k+1)}:=\boldsymbol{f}^{(k)}+\Delta\boldsymbol{f}$, $\alpha_{k+1}=q_a\alpha_k$ and $\rho_{k+1}=$

$q_b\rho_k$ and $0 < q_a, q_b < 1$. $\boldsymbol{A}'\left(\boldsymbol{f}^{(k)}\right)$ represents the derivative of $\boldsymbol{A}\left(\boldsymbol{f}^{(k)}\right)$ with respect to $\boldsymbol{f}^{(k)}$. The term $W_1\left(\boldsymbol{c}^{(k)} + \Delta\boldsymbol{c}\right)$ represents the penalty on the Fourier coefficients, and R is a regularization term. In the original IRGN method, the conventional L_2 was considered. Since the TV regularization can introduce stair-casing artifacts and reduce the image quality if the penalty parameter is too large, the authors of Ref. 41 considered the second-order TGV (total generalization variation, TGV), which is a generalized TV. Compared with the conventional TV, the TGV avoids stair-casing in regions of smooth signal changes and improves the image quality[30,42]. Therefore, the authors of Ref. 41 employed TGV in IRGN and then generated IRGN-TGV for MRI. More details are in Ref. 30 and the corresponding code can be downloaded from https://www.tugraz.at/fileadmin/user_upload/Institute/IMT/files/misc/irgntv.zip. The parameters can be further tuned, depending on experimental designs[30].

II.B. Dictionary Learning-Based Reconstruction

II.B.1. Narrative. As a successful example, dictionary learning-based methods were developed for tomographic reconstruction, including MRI[43-45], Optical Coherence Tomography[46-49] and CT[33,50-52]. Dictionary learning based reconstruction methods explored the intrinsic properties using the trained dictionary with initial reconstruction results. The reconstruction process is usually divided into two steps: dictionary learning and image reconstruction. Without loss of generality, we compare the dictionary learning-based reconstruction method with our proposed ACID for CT and MRI.

II.B.2. Dictionary Learning Model. A number of image patches $\boldsymbol{f}_{i_d} \in \boldsymbol{\mathcal{R}}^{s\times s}$, $i_d = 1, \dots, I_d$, are extracted from the training datasets $\boldsymbol{f}$, and s represents the size of image patches. The set of $\boldsymbol{f}_{i_d}$, $i_d = 1, \dots, I_d$ is employed to train the global dictionary $\boldsymbol{D}_{ic} \in \boldsymbol{\mathcal{R}}^{S\times T_d}$, where $S = s \times s$ and T_d is the number of atoms. The aim of dictionary learning is to search representation coefficients with sparse-level space constrained by $\boldsymbol{q} \in \boldsymbol{\mathcal{R}}^{T_d\times I_d}$ based on the dictionary $\boldsymbol{D}_{ic}$. It can be explained by solving the following optimization expression:

$$\{\boldsymbol{D}_{ic}^*, \boldsymbol{q}^*\} = argmin\ \frac{1}{2}\sum_{i_d=1}^{I_d}\left\| \boldsymbol{f}_{i_d} - \boldsymbol{D}_{ic}\boldsymbol{q}_{i_d}\right\|_2^2 \ \text{ s.t. } \left\|\boldsymbol{q}_{i_d}\right\|_0 \le L_{dl}, \qquad \text{(S.2.6)}$$

where L_{dl} is the sparsity level of dictionary learning, $\|\cdot\|_0$ represents the quasi-l_0 norm, $\boldsymbol{q}_{i_d} \in \boldsymbol{\mathcal{R}}^{T_d\times 1}$ represents sparse representation coefficients for the i_d–th image patch. (S.2.6) is a constrained problem, and it is equivalent to the following unconstrained problem under a certain condition:

$$\{\boldsymbol{D}_{ic}^*, \boldsymbol{q}^*\} = argmin\left(\sum_{i_d=1}^{I_d}\left(\frac{1}{2}\left\|\boldsymbol{f}_{i_d} - \boldsymbol{D}_{ic}\boldsymbol{q}_{i_d}\right\|_2^2 + h_{i_d}\left\|\boldsymbol{q}_{i_d}\right\|_0\right)\right), \qquad \text{(S.2.7)}$$

where h_{i_d} represents a Lagrange multiplier, which needs to be optimized. Furthermore, (S.2.7) can be solved by an alternating minimization scheme. First, we need to update $\boldsymbol{q}_{i_d}$ with a fixed dictionary $\boldsymbol{D}_{ic}$,

$$\boldsymbol{q}^* = argmin\sum_{i_d=1}^{I_d}\left(\frac{1}{2}\left\|\boldsymbol{f}_{i_d} - \boldsymbol{D}_{ic}\boldsymbol{q}_{i_d}\right\|_2^2 + h_{i_d}\left\|\boldsymbol{q}_{i_d}\right\|_0\right). \qquad \text{(S.2.8)}$$

(S.2.8) can be solved using the matching pursuit (MP)[53] or orthogonal matching pursuit (OMP) algorithm[54].Then, we can update the dictionary with a fixed sparse representation coefficients $\boldsymbol{q}$. Many methods can be employed to train the dictionary $\boldsymbol{D}_{ic}$, such as K-SVD[55], discriminate K-SVD[56], coupled dictionary training[57], online learning technique[58] and online robust learning[59]. In this study, the K-SVD was employed.

II.B.3. Dictionary Learning-Based CT Reconstruction. The conventional dictionary learning was first employed to MR reconstruction from under-sampled k-space data[43]. Then, the dictionary learning was utilized to low-dose CT imaging in our previous work[33], few-view CT reconstruction[28] and material decomposition[60]. In this study, we only consider the dictionary learning-based sparse data CT reconstruction. The mathematical model of dictionary learning-based CT reconstruction can be written as follows:

$$\underset{\boldsymbol{f},\boldsymbol{q}}{\operatorname{argmin}} \frac{1}{2}\left\|\boldsymbol{p}^{(\mathbf{0})} - \boldsymbol{A}\boldsymbol{f}\right\|_2^2 + \varsigma \sum_{i_d=1}^{I_d}\left(\frac{1}{2}\left\|\wp_{i_d}\boldsymbol{f} - \boldsymbol{D}_{ic}\boldsymbol{q}_{i_d}\right\|_2^2 + h_{i_d}\left\|\boldsymbol{q}_{i_d}\right\|_0\right), \quad \text{(S.2.9)}$$

where $\varsigma > 0$ represents the regularization penalty parameter. $\wp_{i_d}$ is an operator to extract i_d–th image patch from $\boldsymbol{f}$. Regarding the optimization of (S.2.9), there are many strategies to reach such a goal. Here, the split-Bregman method is used to obtain its solution. First, we introduce a new variable $\boldsymbol{b}$ to replace $\boldsymbol{f}$ and (S.2.9) can be converted into the following constraint programming problem

$$\underset{\boldsymbol{f},\boldsymbol{q}}{\operatorname{argmin}} \frac{1}{2}\left\|\boldsymbol{p}^{(\mathbf{0})} - \boldsymbol{A}\boldsymbol{f}\right\|_2^2 + \varsigma \sum_{i_d=1}^{I_d}\left(\frac{1}{2}\left\|\wp_{i_d}\boldsymbol{b} - \boldsymbol{D}_{ic}\boldsymbol{q}_{i_d}\right\|_2^2 + h_{i_d}\left\|\boldsymbol{q}_{i_d}\right\|_0\right), s.t., \boldsymbol{f} = \boldsymbol{b}. \quad \text{(S.2.10)}$$

To optimize (S.2.10), it can be further converted into

$$\underset{\boldsymbol{f},\boldsymbol{b},\boldsymbol{\chi},\boldsymbol{q}}{\operatorname{argmin}} \frac{1}{2}\left\|\boldsymbol{p}^{(\mathbf{0})} - \boldsymbol{A}\boldsymbol{f}\right\|_2^2 + \varsigma \sum_{i_d=1}^{I_d}\left(\frac{1}{2}\left\|\wp_{i_d}\boldsymbol{b} - \boldsymbol{D}_{ic}\boldsymbol{q}_{i_d}\right\|_2^2 + h_{i_d}\left\|\boldsymbol{q}_{i_d}\right\|_0\right) + \frac{\varsigma_1}{2}\left\|\boldsymbol{f} - \boldsymbol{b} - \boldsymbol{\chi}\right\|_2^2, \quad \text{(S.2.11)}$$

where $\varsigma_1 > 0$ represents the coupling factor, and $\boldsymbol{\chi}$ is the error feedback. In (S.2.11), there are four variables $\boldsymbol{f}, \boldsymbol{b}, \boldsymbol{q}$ and $\boldsymbol{\chi}$. It can be split into the following three sub-problems:

$$\underset{\boldsymbol{f}}{\operatorname{argmin}} \frac{1}{2}\left\|\boldsymbol{p}^{(\mathbf{0})} - \boldsymbol{A}\boldsymbol{f}\right\|_2^2 + \frac{\varsigma_1}{2}\left\|\boldsymbol{f} - \boldsymbol{b}^{(k)} - \boldsymbol{\chi}^{(k)}\right\|_2^2, \quad \text{(S.2.12)}$$

$$\underset{\boldsymbol{b},\boldsymbol{q}}{\operatorname{argmin}} \sum_{i_d=1}^{I_d}\left(\frac{1}{2}\left\|\wp_{i_d}\boldsymbol{b} - \boldsymbol{D}_{ic}\boldsymbol{q}_{i_d}\right\|_2^2 + h_{i_d}\left\|\boldsymbol{q}_{i_d}\right\|_0\right) + \frac{\varsigma_1}{2}\left\|\boldsymbol{f}^{(k+1)} - \boldsymbol{b} - \boldsymbol{\chi}^{(k)}\right\|_2^2, \quad \text{(S.2.13)}$$

$$\boldsymbol{\chi}^{(k+1)} = \boldsymbol{\chi}^{(k)} - \tau_d\left(\boldsymbol{f}^{(k+1)} - \boldsymbol{b}^{(k+1)}\right), \quad \text{(S.2.14)}$$

where $\tau_d > 0$ represents the step length and it was set to 1 in this study. Regarding (S.2.12), it is solved by using the separable surrogate method[61]

$$\boldsymbol{f}_{j_1 j_2}^{(k+1)} = \boldsymbol{f}_{j_1 j_2}^{(k)} - \frac{\left[\boldsymbol{A}^T\left(\boldsymbol{A}\boldsymbol{f}^{(k)} - \boldsymbol{p}^{(\mathbf{0})}\right)\right]_{j_1 j_2} + \varsigma_1\left[\boldsymbol{f}^{(k)} - \boldsymbol{b}^{(k)} - \boldsymbol{\chi}^{(k)}\right]_{j_1 j_2}}{\left[\boldsymbol{A}^T\boldsymbol{A} + \varsigma_1\right]_{j_1 j_2}}, \quad \text{(S.2.15)}$$

where$[.]_{j_1 j_2}$ represents the $(j_1, j_2)^{th}$ pixel in the matrix. In practice, (S.2.15) was performed using two steps:

$$\boldsymbol{f}_{j_1 j_2}^{\left(k+\frac{1}{2}\right)} = \boldsymbol{f}_{j_1 j_2}^{(k)} - \frac{\left[\boldsymbol{A}^T\left(\boldsymbol{A}\boldsymbol{f}^{(k)} - \boldsymbol{p}^{(\mathbf{0})}\right)\right]_{j_1 j_2}}{\left[\boldsymbol{A}^T\boldsymbol{A} + \varsigma_1\right]_{j_1 j_2}}, \quad \text{(S.2.16)}$$

and

$$\boldsymbol{f}_{j_1 j_2}^{(k+1)} = \boldsymbol{f}_{j_1 j_2}^{\left(k+\frac{1}{2}\right)} - \frac{\varsigma_1\left[\boldsymbol{f}^{(k)} - \boldsymbol{b}^{(k)} - \boldsymbol{\chi}^{(k)}\right]_{j_1 j_2}}{\left[\boldsymbol{A}^T\boldsymbol{A} + \varsigma_1\right]_{j_1 j_2}}. \quad \text{(S.2.17)}$$

In fact, the number of iterations for $\boldsymbol{f}_{j_1 j_2}^{\left(k+\frac{1}{2}\right)}$ in (S.2.16) needs to be set to a good number (it was set to 10 in this study), and then $\boldsymbol{f}_{j_1 j_2}^{(k+1)}$ is updated. Since ς_1 is specific to scanning geometry, it is normalized into a new parameter γ_1 so that $\varsigma_1 = \gamma_1\|\boldsymbol{A}^T\boldsymbol{A}\|$; that is, we only need to select a geometrically-invariant γ_1. Regarding the optimization of (S.2.13), it is a typical dictionary learning-based

signal recovery problem, and there are a large number of algorithms to solve this problem[55,57]. To control the image recovery via dictionary learning, the parameters of sparsity level L_{dl} and the precision level ζ should be chosen; for more details, see our previous studies[28,33,62].

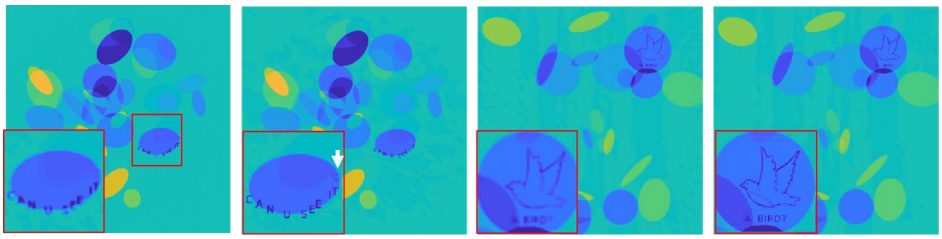
II.B.4. Dictionary Learning-Based MRI Reconstruction. The conventional dictionary learning methods are common for MR reconstruction[43,63-65]. In this study, the dictionary learning-based MRI (DLMRI)[43] was employed to highlight the advantages of the ACID with built-in DAGAN and TV. Regarding the reconstruction process of DLMRI, it is similar to the process for CT reconstruction. It is also divided into two steps: dictionary learning and image updating. Regarding the dictionary learning step, both MRI and CT are the same except that training images are different. Again, the dictionary used in CT reconstruction was trained from FBP results or updated results within the iteration process. In contrast, the dictionary utilized in MRI was trained from the inverse Fourier transform results. Regarding the image updating step, it is not necessary to update the image based on the fast Fourier transform. More details can be found in Ref. 43.

II.B.5. Experimental Results. To validate the outperformance of ACID in comparison with the dictionary learning-based CT reconstruction method (DLCT), we repeated the experiments design for the cases C1 and C2. Here, we adopt the FBP method to reconstruct images. Then, the FBP results were employed to train the dictionaries. In this study, only 1.0x10^4 image patches were extracted from FBP images to train the dictionary by the K-SVD algorithm. The size of extracted image patches was set to 6×6. The dictionary $\boldsymbol{D}_{ic}$ is overcomplete, and it can benefit the sparsity level enforcement. The number of atoms was set to 512. The sparsity level L_{dl} in the dictionary training can be set empirically, and it was chosen as 6. The number of iterations for the training dictionary was set to 100.

Note that the total variation is still treated as the compressed sensing-based sparsity for the built-in component in the ACID. Here, the parameters of γ_1, L_{dl} and ζ in DLCT were set to 0.22, 8 and 0.06, respectively. The number of outer iterations was set to 200. The implementation environment for training and reconstruction is the same as Ell-50. Specifically, the computational costs of dictionary training and reconstruction consume 139 and 561 seconds. However, the whole ACID with the built-in Ell-50 consumes about 70.5 seconds. In other words, the ACID is faster than the DLCT method.

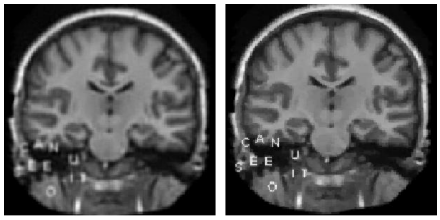
The reconstruction results from DLCT and ACID with C1 and C2 are in Fig. S9. It is observed that DLCT provides higher image quality than that obtained by the CS method. However, it is still worse than those obtained by the ACID. Besides, the proposed ACID method obtains better image edges and avoids blurred artifacts compared with the DLCT method. Especially, the insert texts in DLCT results (i.e., “CAN U SEE IT” and “A BIRD?”) are very blurry, and they failed to be discriminated against. These texts are clearly observed in ACID results. Regarding small features (*i.e.*, the symbol “ ”), they are totally missing in the DLCT result. However, they were still recovered by the ACID. In terms of quantitative assessment, our proposed ACID obtained the best results remarkably. More details for codes and test data are at https://zenodo.org/record/5497811.

To show the advantages of ACID with the built-in DAGAN, the reconstruction results in the M2 case from ACID and DLMRI are given in Fig. S10. The DLMRI obtained higher image quality than that obtained by the conventional CS method in the main text. However, it is still worse than those obtained by the ACID. Besides, the proposed ACID method obtained better image edges and avoided blurry artifacts compared with the DLMRI method. Especially, the inserted texts in DLCT results (i.e., "CAN U SEE IT") are very blurry, and they could not be discriminated. These texts are clearly observed in the ACID results. The small symbol was totally blurred in DLMRI result, which was still recovered by ACID. In terms of quantitative assessment, our proposed ACID obtained better results than those achieved by DLMRI method. The MATLAB code of the MRIDL method can be downloaded from http://www.ifp.illinois.edu/~yoram/DLMRI-Lab/DLMRI.html. The reconstruction parameters within DLMRI were optimized. Regarding the computational cost, under the same computing environment, DLMRI took 606.3 seconds, which is higher than that of CS reconstruction methods in the main text (i.e., 127.8 seconds). Our proposed ACID only took 9.2 seconds.



(a) 32.87/0.979 (b) 40.86/0.995 (c) 32.41/0.984 (d) 38.78/0.993

Fig. S9. Comparison study on the DLCT and ACID methods. (a) and (c), (b) and (d) are reconstructed results from DLCT and ACID, respectively. The numbers represent the quantitative results in terms of PSNR and SSIM, and the display window is [-150, 150] HU.



(a) 30.9/0.963 (b) 34.9/0.977

Fig. S10. Comparison study on the DLMRI and ACID methods. (a) and (b) are the reconstructed results from DLCT and ACID respectively. The numbers represent the quantitative results in terms of PSNR and SSIM.

III. ACID Implementation & More Results

III.A. ACID Implementation

For an implementation of the whole ACID system, it is considered as an iterative framework and listed in Algorithm 1. In the whole ACID framework, we need to have input data $\boldsymbol{p}^{(0)}$, a neural network Φ and a system matrix $\boldsymbol{A}$. Then, we should specify the stopping condition; i.e., the maximum number of iterations K. Finally, the parameters λ and ε should be given to control the iterative process and the regularization strength, all of which can be empirically picked up. When k=1, we need to compute $\Phi(\boldsymbol{p}^{(0)})$ and then normalize $\Phi(\boldsymbol{p}^{(0)})$. The goal of the normalization operator is to facilitate the adjustment of the regularization parameters for different applications. Then, we obtain the updated $\boldsymbol{b}^{(1)}$ using

the second formula in (1) in the main body of this paper. Next, $\boldsymbol{f}^{(1)}$ is updated by de-normalizing $\boldsymbol{b}^{(1)}$. When 1<k<K+1, we need to compute the residual data by $\boldsymbol{p}^{(k+1)} = \frac{\lambda(\boldsymbol{p}^{(0)} - \boldsymbol{A}\boldsymbol{f}^{(k)})}{1+\lambda}$. Since the residual data are not in the dynamic range of the original data, the residual data should be normalized into the original range to make sure the efficiency of the neural network (Line #9 in Algorithm 1). After the neural network predicts a residual image, the de-normalization operator should be applied on the prediction to ensure the consistency of the reconstruction results. Then, $\boldsymbol{f}^{(k)} + \frac{1}{\lambda}\Phi(\boldsymbol{p}^{(k+1)})$ is normalized and fed to the compressed sensing-based regularization module to encourage image sparsity. Finally, we obtain the updated image $\boldsymbol{f}^{(k+1)}$ after the de-normalization. More details on our codes and other materials are available at https://zenodo.org/record/5497811.

Algorithm 1. Pseudocode of the ACID workflow.

Input: Data $\boldsymbol{p}^{(0)}$, neural network Φ, system matrix $\boldsymbol{A}$, maximum number of iterations K, auxiliary parameters λ, ε, and k=1;

1. If k<K+1 do
2. if k=1 do
3. Computing $\Phi(\boldsymbol{p}^{(0)})$;
4. Normalizing $\Phi(\boldsymbol{p}^{(0)})$;
5. Updating $\boldsymbol{b}^{(1)}$ where the normalized $\Phi(\boldsymbol{p}^{(0)})$ is treated as the input;
6. Updating $\boldsymbol{f}^{(1)}$ by de-normalizing $\boldsymbol{b}^{(1)}$;
7. else do
8. Computing residual data using $\boldsymbol{p}^{(k+1)} = \frac{\lambda(\boldsymbol{p}^{(0)} - \boldsymbol{A}\boldsymbol{f}^{(k)})}{1+\lambda}$;
9. Normalizing the residual data $\boldsymbol{p}^{(k+1)}$ into the input range of neural network to obtain $\bar{\boldsymbol{p}}^{(k+1)}$;
10. Inputting $\bar{\boldsymbol{p}}^{(k+1)}$ into the neural network Φ and obtaining $\Phi(\bar{\boldsymbol{p}}^{(k+1)})$;
11. De-normalizing $\Phi(\bar{\boldsymbol{p}}^{(k+1)})$ to obtain $\Phi(\boldsymbol{p}^{(k+1)})$;
12. Normalizing $\boldsymbol{f}^{(k)} + \frac{1}{\lambda}\Phi(\boldsymbol{p}^{(k+1)})$
13. Updating $\boldsymbol{b}^{(k+1)}$;
14. Updating $\boldsymbol{f}^{(k+1)}$ by de-normalizing $\boldsymbol{b}^{(k+1)}$;
15. end
16. end
17. return $\boldsymbol{f}^{(K)}$

Output: Reconstructed image $\boldsymbol{f}^{(K)}$

III.B. Difference Images for Figures 2-4

Here we provide the difference images for figures 2-4 in the main text. Fig. S11 includes the difference images of Figure 2 in the main text. It can be seen that the results provided by ACID is closer to the ground truth. In addition, the differences of the text symbols in the competing approaches are more obvious than that obtained by our ACID. It further demonstrates our ACID can effective stabilize the deep tomographic network against structure changes.

Fig. S12 includes the difference images of Figure 3 in the main text. It can be seen that the results provided by ACID is also closer to the ground truth. In addition, the differences of inserted text symbols in the competing algorithms are more obvious than that obtained by our ACID. Fig. S12 further demonstrates our ACID can effectively stabilize the deep tomographic network against adversarial attacks. The difference images of ACID against noise are also provided in Fig. S13. It can be observed that the difference images with our ACID are the smallest among all the algorithms.

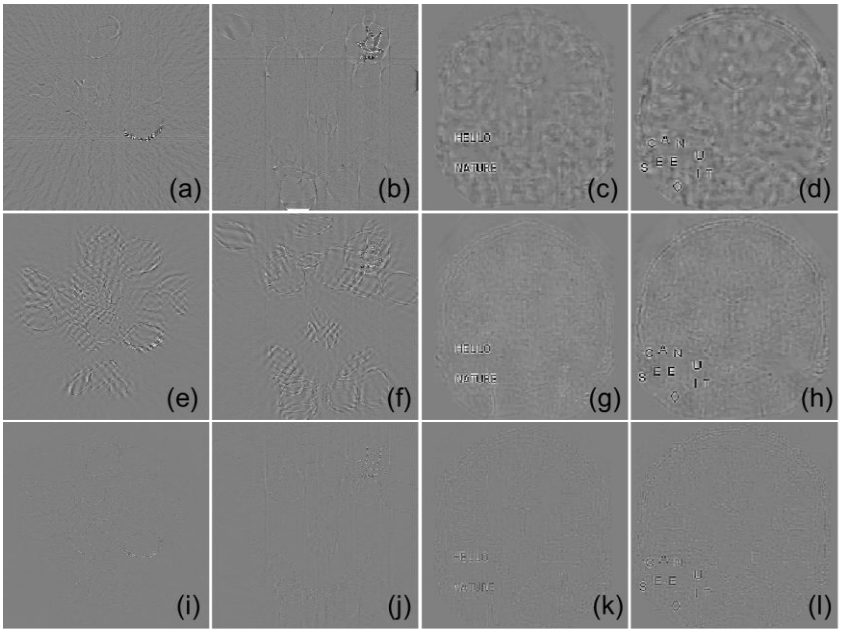


Fig. S11. **Difference images of ACID with small structural changes in the CT and MRI cases.** The 1st- 3rd rows are the difference images of Ell-50, CS-inspired and ACID results respectively with respect to the ground truth. The 1st-4th columns correspond to CT cases C1 and C2, and MRI cases M1 and M2, respectively. The sub-sampling rate of MRI is 10%. The display windows for CT and MRI are [-70 70] HU and [-0.5 0.5], respectively.

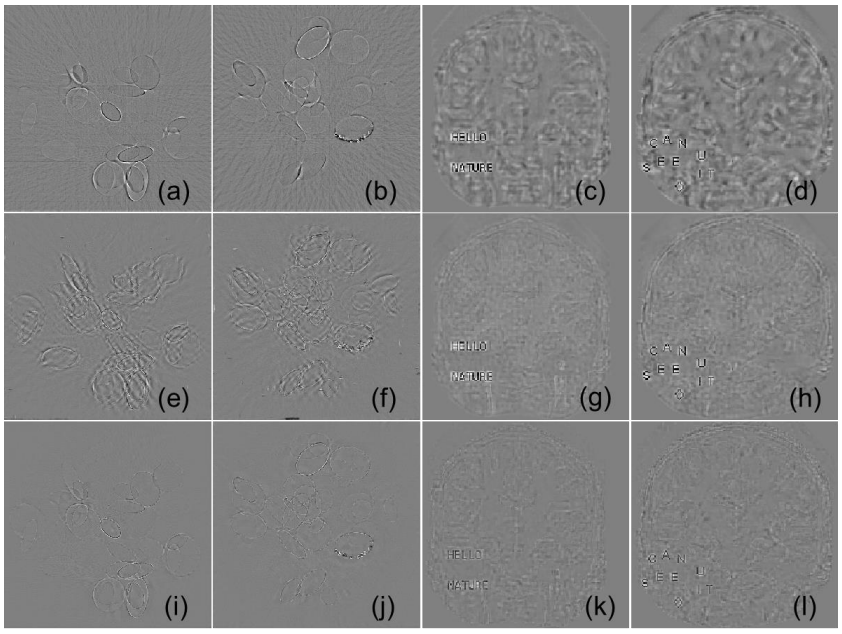


Fig. S12. **Difference images of ACID with adversarial attack in the CT and MRI cases.** The 1st -3rd rows are the difference images of Ell-50, CS-inspired and ACID results respectively with respect to the ground truth. The 1st-4th columns correspond to CT cases C3 and C4, and MRI cases M3 and M4, respectively. Each CT dataset contains 50 projections, and the sub-sampling rate of MRI is 10%. The display windows for CT and MRI are [-70 70] HU and [-0.5 0.5], respectively.

III.C. ACID Performance on Real CT Dataset

To validate the proposed ACID on real datasets, here we retrained the

FBPConvNet on mayo clinical datasets and the reconstruction results are

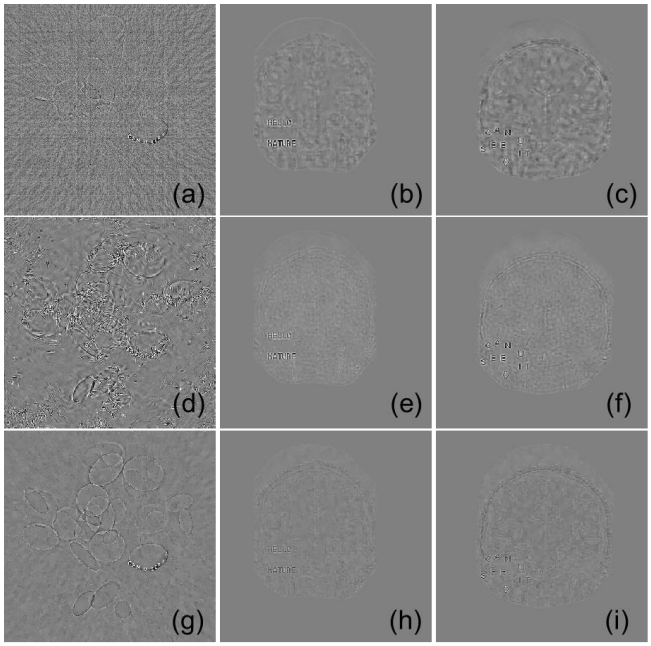


Fig. S13. **Difference images in the C5, M5 and M6 cases against noise.** In the first column, (a), (d) and (g) present the difference images between ground truth and that of EII-50, CS and ACID results on C5. The second and third columns are the counterparts of M5 and M6 showing the difference images between the ground truth and that of DAGAN, CS and ACID results, respectively.

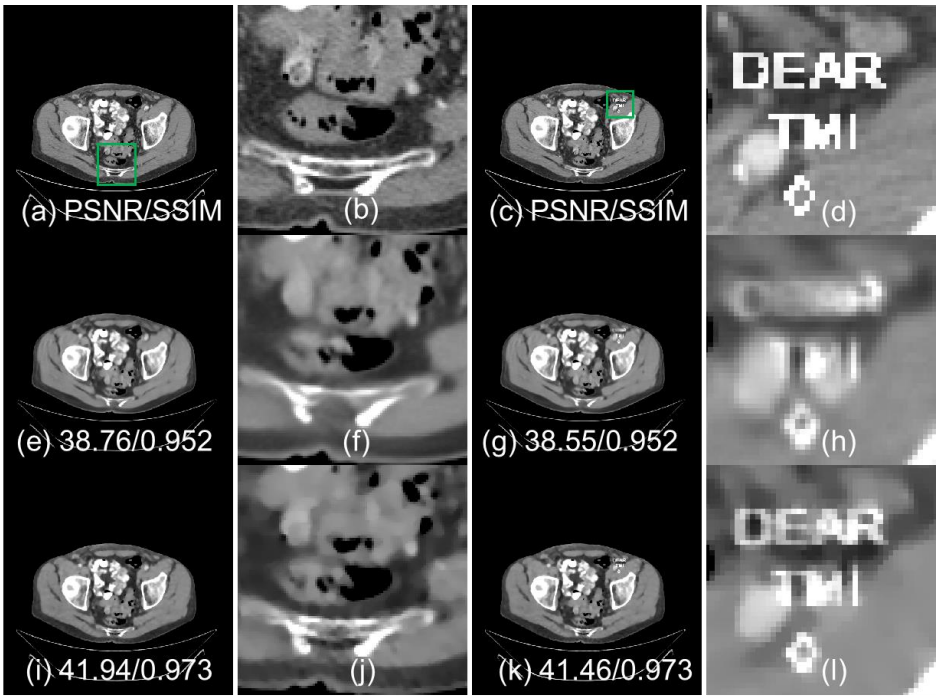


Fig. S14. **Reconstruction performance of ACID on Mayo clinical datasets.** In the first column, (a), (e) and (i) are the ground truth, FBPConvNet and ACID results. The second column are the magnified ROI in the first column. The third and fourth columns are the counterparts of the first and second columns with structural changes. The numbers indicate the PSNR and SSIM values, and the display window is [-160, 240] HU.

further given in Fig. S14. There are two cases in Fig. S14, where one has no

structural changes and the other has structural changes. For the case without structural changes, our ACID built with FBPConvNet can provide higher reconstructed image quality than FBPConvNet itself. For the case with structural changes, our ACID obviously provides clearer insert symbols than FBPConvNet. Furthermore, ACID has higher quantitative PSNR and SSIM results than the FBPConvNet.

III.D. ACID Against Distributional Robustness

As for robustness, distributional robustness is very important for image reconstruction. To further demonstrate the advantages of the generalization ability from our ACID, here we first use the test dataset of DAGAN to test AUTOMAP. Then, we also use the test datasets from AUTOMAP to test DAGAN. The results of Fig. S15 demonstrate the DAGAN has a relative weakness distributional robustness since its results contain several structural artifacts. However, these artifacts can be removed by our ACID framework with built-in DAGAN network. In addition, our ACID also provides higher PSNR as well as SSIM. The results of Fig. S16 demonstrate the AUTOMAP has good distributional robustness, but it also contains structure and other artifacts. However, these artifacts can be removed by our ACID framework with built-in AUTOMAP network. In addition, our ACID also provides higher PSNR as well as SSIM.

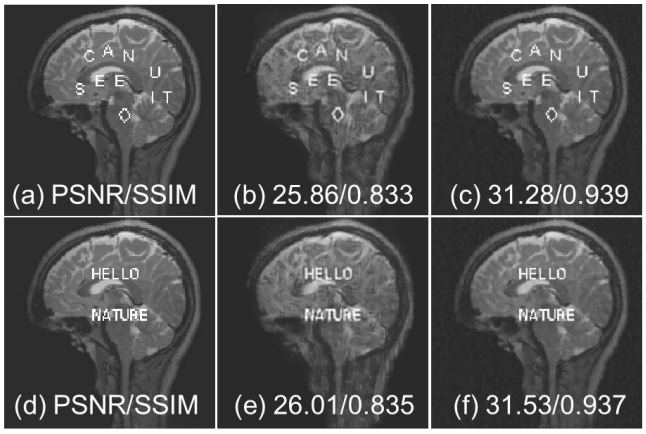


Fig. S15. Distributional robustness of DAGAN against AUTOMAP test datasets. (a)-(c) are the ground truth, DAGAN result and ACID result with built-in DAGAN network. (d)-(f) are the counterparts of (a)-(c) for another case. The numbers indicate PSNR and SSIM values.

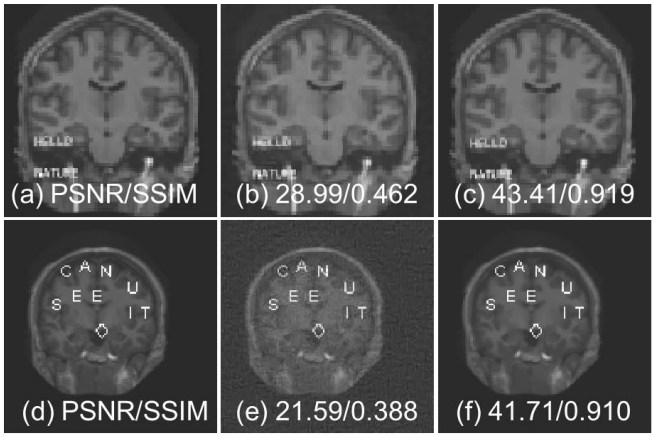


Fig. S16. Distributional robustness of AUTOMAP against DAGAN test datasets. (a)-(c) are ground truth, AUTOMAP result and ACID result with built-in AUTOMAP network. (d)-(f) are

counterparts of (a)-(c) for another case. The numbers indicate PSNR and SSIM values.

References

1 Jin, K. H., McCann, M. T., Froustey, E., and Unser, M. (2017). Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing,* 26, 4509-4522. 10.1109/TIP.2017.2713099.

2 He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, p.*770-778. 10.1109/CVPR.2016.90.

3 Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. in *International Conference on Medical Image Computing and Computer-assisted Intervention, LNCS, Vol. 9531, p.*234-241.

4 Vedaldi, A., and Lenc, K. (2015). Matconvnet: Convolutional neural networks for matlab. in *Proceedings of the 23rd ACM international conference on Multimedia,* p.689-692. 10.1145/2733373.2807412.

5 Pascanu, R., Mikolov, T., and Bengio, Y. (2013). On the difficulty of training recurrent neural networks. in *International conference on machine learning*, PMLR, 28(3), 1310-1318.

6 Antun, V., Renna, F., Poon, C., Adcock, B., and Hansen, A. C. (2020). On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proceedings of the National Academy of Sciences,* 117, 30088-30095. 10.1073/pnas.1907377117.

7 Yu, H.Y., Zhao, S.Y., and Wang, G. (2005). A differentiable Shepp–Logan phantom and its applications in exact cone-beam CT. *Physics in Medicine & Biology,* 50, 5583-5595. 10.1088/0031-9155/50/23/012.

8 Averbuch, A., Sedelnikov, I., and Shkolnisky, Y. (2011). CT reconstruction from parallel and fan-beam projections by a 2-D discrete Radon transform. *IEEE Transactions on Image Processing,* 21, 733-741. 10.1109/TIP.2011.2164416.

9 Zhang, Z., Liang, X., Dong, X., Xie, Y., and Cao, G. (2018). A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution. *IEEE Transactions on Medical Imaging,* 37, 1407-1417. 10.1109/TMI.2018.2823338.

10 Zang, G., Aly, M., Idoughi, R., Wonka, P., and Heidrich, W. (2018). Super-resolution and sparse view CT reconstruction. in *Proceedings of the European Conference on Computer Vision (ECCV),* p.137-153.

11 Li, H.Y., Zhao, T.H., Wei, M.L., Ruan, H.X., Shuang, Y., Cui, T.J., Del Hougne, P., and Li, L.L (2020). Intelligent electromagnetic sensing with learnable data acquisition and processing. *Patterns* 1, Article ID: 100006. 10.1016/j.patter.2020.100006.

12 Gottschling, N. M., Antun, V., Adcock, B., and Hansen, A. C. (2020). The troublesome kernel: why deep learning for inverse problems is typically unstable. *arXiv preprint, arXiv:2001.01258*.

13 Yang, G., Yu, S.M., Dong, H., Slabaugh, G., Dragottyi, P.L., Ye, X., Liu, F., Arridge, S., Keegan, J., Guo, Y., *et al.* (2017). DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Transactions on Medical Imaging,* 37, 1310-1321. 10.1109/TMI.2017.2785879.

14 Hwang, H., Rehman, H.Z.U., and Lee, S. (2019). 3D U-Net for skull stripping in brain MRI. *Applied Sciences,* 9, Article ID: 569, 515 pages. 10.3390/app9030569.

15 Dolz, J., Desrosiers, C., and Ayed, I.B. (2018). IVD-Net: Intervertebral disc localization and segmentation in MRI with a multi-modal UNet. in *International Workshop and Challenge on Computational Methods and Clinical Applications for Spine Imaging,* p.130-143, Springer.

16 Kermi, A., Mahmoudi, I., and Khadir, M.T. (2018). Deep convolutional neural networks using U-Net for automatic brain tumor segmentation in multimodal MRI volumes. in *International MICCAI Brainlesion Workshop,* p.37-48, Springer.

17 Zhao, J., Mathieu, M., and LeCun, Y. (2016). Energy-based generative adversarial network. *arXiv preprint, arXiv:1609.03126*.

18 You, C.Y., Li, G., Zhang, Y., Zhang, X., Shan, H.M., Li, M.Z., Ju, S.H., Zhao, Z., Zhang, Z.Y., Cong, W.X., *et al.* (2019). CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE). *IEEE Transactions on Medical Imaging,* 39, 188-203. 10.1109/TMI.2019.2922960.

19 Wang, G., Ye, J. C., Mueller, K., and Fessler, J.A. (2018). Image reconstruction is a new frontier of machine learning. *IEEE Transactions on Medical Imaging,* 37, 1289-1296. 10.1109/TMI.2018.2833635.

20 Yang, Q., Yan, P.K., Zhang, Y.B., Yu, H.Y., Shi, Y.Y., Mou, X.Q., Kalra, M.K., Zhang, Y., Sun., L., and Wang, G. (2018). Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Transactions on Medical Imaging,* 37, 1348-1357. 10.1109/TMI.2018.2827462.

21 Kingma, D.P., and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint, arXiv:1412.6980*.

22 Yang, Y., Sun, J., Li, H., and Xu, Z (2016). Deep ADMM-Net for compressive sensing MRI. in *Advances in neural information processing systems,* p.10-18.

23 Zhu, B., Liu, J.Z., Cauley, S.F., Rosen, B.R., and Rosen, M.S. (2018). Image reconstruction by domain-transform manifold learning. *Nature,* 555, 487-492. 10.1038/nature25988.

24 Liu, J., Lefebvre, A., and Nadar, M. (2013). Alternating direction of multipliers method for parallel MRI reconstruction, Google Patents, Patent # 8,879,811.

25 Lustig, M., Donoho, D.L., Santos, J.M. and Pauly, J.M. (2008). Compressed sensing MRI. *IEEE Signal Processing Magazine,* 25, 72-82. 10.1109/MSP.2007.914728.

26 Haldar, J. P., Hernando, D., and Liang, Z.P. (2011). Compressed-sensing MRI with random encoding. *IEEE Transactions on Medical Imaging,* 30, 893-903.

10.1109/TMI.2010.2085084.

27 Shi, W., Ling, Q., Yuan, K., Wu, G., and Yin, W. (2014). On the linear convergence of the ADMM in decentralized consensus optimization. *IEEE Transactions on Signal Processing,* 62, 1750-1761. 10.1109/TSP.2014.2304432.

28 Xu, M., Hu, D.L., Luo, F.L., Liu, F.L., Wang, S.Y., and Wu, W.W. (2021). Limited angle X ray CT Reconstruction using Image Gradient ℓ0 norm with Dictionary Learning. *IEEE Transactions on Radiation and Plasma Medical Sciences,* 5, 78-87. 10.1109/TRPMS.2020.2991887.

29 Ma, S., Yin, W., Zhang, Y., and Chakraborty, A. (2008). An efficient algorithm for compressed MR imaging using total variation and wavelets. in *2008 IEEE Conference on Computer Vision and Pattern Recognition,* p.1-8.

30 Knoll, F., Bredies, K., Pock, T. , and Stollberger, R. (2011). Second order total generalized variation (TGV) for MRI. *Magnetic Resonance in Medicine,* 65, 480-491. 10.1002/mrm.22595.

31 Knoll, F., Holler, M., Koesters, T., Otazo, R., Bredies, K., and Sodickson, D.K. (2017). Joint MR-PET reconstruction using a multi-channel image regularizer. *IEEE transactions on medical imaging,* 36, 1-16. 10.1109/TMI.2016.2564989.

32 Yu, H.Y., Wang, G., Hsieh, J., Entrikin, D.W., Ellis, S., Liu, B.D., and Carr, J.J. (2011). Compressive Sensing–Based Interior Tomography: Preliminary Clinical Application. *Journal of Aomputer Assisted Tomography,* 35, 762-764. 10.1097/RCT.0b013e318231c578.

33 Xu, Q., Yu. H.Y., Mou, X.Q., Zhang, L., Hsieh, J. and Wang, G. (2012). Low-dose X-ray CT reconstruction via dictionary learning. *IEEE Transactions on Medical Imaging,* 31, 1682-1697. 10.1109/TMI.2012.2195669.

34 Wu, W.W., Zhang, Y.B., Wang, Q., Liu, F.L., Chen, P.J., and Yu, H.Y. (2018). Low-dose spectral CT reconstruction using image gradient ℓ0–norm and tensor dictionary. *Applied Mathematical Modelling,* 63, 538-557. 10.1016/j.apm.2018.07.006.

35 Candes, E.J., Wakin, M.B., and Boyd, S.P. (2008). Enhancing sparsity by reweighted ℓ1 minimization. *Journal of Fourier Analysis and Applications,* 14, 877-905. 10.1007/s00041-008-9045-x.

36 Ma, J., März, M., Funk, S., Schulz-Menger, J., Kutyniok, G., Schaeffter, T., and Kolbitsch, C. (2018). Shearlet-based compressed sensing for fast 3D cardiac MR imaging using iterative reweighting. *Physics in Medicine & Biology,* 63, Article ID: 235004. 10.1088/1361-6560/aaea04.

37 Ma, J., and März, M. (2016). A multilevel based reweighting algorithm with joint regularizers for sparse recovery. *arXiv preprint, arXiv:1604.06941*.

38 Yoon, S., and Jameson, A. (1988). Lower-upper symmetric-Gauss-Seidel method for the Euler and Navier-Stokes equations. *AIAA Journal,* 26, 1025-1026 (1988). 10.2514/3.10007.

39 Block, K.T., Uecker, M., and Frahm, J. (2007). Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. *Magnetic Resonance in Medicine,* 57, 1086-1098. 10.1002/mrm.21236.

40 Ehrhardt, M.J., and Betcke, M.M. (2016). Multicontrast MRI reconstruction with structure-guided total variation. *SIAM Journal on Imaging Sciences,* 9, 1084-1106. 10.1137/15M1047325.

41 Knoll, F., Clason, C., Bredies, K., Uecker, M., and Stollberger, R. (2012). Parallel imaging with nonlinear reconstruction using variational penalties. *Magnetic Resonance in Medicine,* 67, 34-41. 10.1002/mrm.22964.

42 Bredies, K., Kunisch, K., and Pock, T. (2010). Total generalized variation. *SIAM Journal on Imaging Sciences,* 3, 492-526. 10.1137/090769521.

43 Ravishankar, S., and Bresler, Y. (2011). MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Transactions on Medical Imaging,* 30, 1028-1041. 10.1109/TMI.2010.2090538.

44 Caballero, J., Price, A.N., Rueckert, D., and Hajnal, J.V. (2014). Dictionary learning and time sparsity for dynamic MR data reconstruction. *IEEE Transactions on Medical Imaging,* 33, 979-994. 10.1109/TMI.2014.2301271.

45 Zhan, Z.F., Cai, J.F., Guo, D., Liu, Y., Chen, Z., and Qu, X. (2015). Fast multiclass dictionaries learning with geometrical directions in MRI reconstruction. *IEEE Transactions on Biomedical Engineering,* 63, 1850-1861. 10.1109/TBME.2015.2503756.

46 Fang, L., Li, S., McNabb, R.P., Nie, Q., Kuo, A.N., Toth, C.A., Izatt, J.A., and Farsiu, S. (2013). Fast acquisition and reconstruction of optical coherence tomography images via sparse representation. *IEEE Transactions on Medical Imaging,* 32, 2034-2049. 10.1109/TMI.2013.2271904.

47 Esmaeili, M., Dehnavi, A. M., Rabbani, H., and Hajizadeh, F. (2017). Speckle noise reduction in optical coherence tomography using two-dimensional curvelet-based dictionary learning. *Journal of Medical Signals and Sensors,* 7, 86-91. 10.1364/BOE.377021.

48 Albarrak, A., Coenen, F., and Zheng, Y. (2017). Volumetric image classification using homogeneous decomposition and dictionary learning: a study using retinal optical coherence tomography for detecting age-related macular degeneration. *Computerized Medical Imaging and Graphics,* 55, 113-123. 10.1016/j.compmedimag.2016.07.007.

49 Fang, L., Li, S., Cunefare, D., and Farsiu, S. (2016). Segmentation based sparse reconstruction of optical coherence tomography images. *IEEE Transactions on Medical Imaging,* 36, 407-421. 10.1109/TMI.2016.2611503.

50 Chen, Y., Shi, L., Feng, Q.J., Yang, J., Shu, H.Z., Luo, L.M., Coatrieux, J.L., and Chen, W.F. (2014). Artifact suppressed dictionary learning for low-dose CT image processing. *IEEE Transactions on Medical Imaging,* 33, 2271-2292. 10.1109/TMI.2014.2336860.

51 Fang, R., Chen, T., and Sanelli, P. C. (2013). Towards robust deconvolution of low-dose perfusion CT: Sparse perfusion deconvolution using online dictionary learning. *Medical image analysis,* 17, 417-428. 10.1016/j.media.2013.02.005.

52 Tan, S.Q., Zhang, Y.B., Wang, G., Mou, X.Q., Cao, G.H., Wu. Z.F., and YU, H.Y. (2015). Tensor-based dictionary learning for dynamic tomographic reconstruction. *Physics in Medicine & Biology,* 60, 2803-2818. 10.1088/0031-

9155/60/7/2803.

53 Mallat, S.G., and Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing,* 41, 3397-3415. 10.1109/78.258082.

54 Chen, S., Billings, S A., and Luo, W. (1989). Orthogonal least squares methods and their application to non-linear system identification. *International Journal of Control,* 50, 1873-1896. 10.1080/00207178908953472.

55 Aharon, M., Elad, M., and Bruckstein, A. (2006). K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing,* 54, 4311-4322. 10.1109/TSP.2006.881199.

56 Zhang, Q., and Li, B. (2010). Discriminative K-SVD for dictionary learning in face recognition. in *2010 IEEE computer society conference on computer vision and pattern recognition,* p.2691-2698. 10.1109/CVPR.2010.5539989.

57 Yang, J., Wang, Z., Lin, Z., Cohen, S., and Huang, T. (2012). Coupled dictionary training for image super-resolution. *IEEE Transactions on Image Processing,* 21, 3467-3478. 10.1109/TIP.2012.2192127.

58 Mairal, J., Bach, F., Ponce, J. and Sapiro, G. (2010). Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research,* 11, 19-60. 10.1145/1756006.1756008.

59 Lu, C., Shi, J., and Jia, J. (2013). Online robust dictionary learning. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* p.415-422. 10.1109/CVPR.2013.60.

60 Wu, W.W., Yu, H.J., Chen, P.J., Luo, F.L., Wang, Q., Zhu, Y.N., Zhang, Y.B., Feng, J., and Yu, H.Y. (2020). Dictionary learning based image-domain material decomposition for spectral CT. *Physics in Medicine & Biology,* 65, Article ID: 245006. 10.1088/1361-6560/aba7ce.

61 Elbakri, I. A. and Fessler, J.A. (2002). Statistical image reconstruction for polyenergetic X-ray computed tomography. *IEEE Transactions on Medical Imaging,* 21, 89-99. 10.1109/42.993128.

62 Zhang, Y.B., Mou, X.Q., Wang, G. and Yu, H.Y. (2017). Tensor-based dictionary learning for spectral CT reconstruction. *IEEE Transactions on Medical Imaging* 36(1), 142-154. 10.1109/TMI.2016.2600249.

63 Merlet, S., Caruyer, E., and Deriche, R. (2012). Parametric dictionary learning for modeling EAP and ODF in diffusion MRI. in *International Conference on Medical Image Computing and Computer-Assisted Intervention,* p.10-17, Springer. 10.1007/978-3-642-33454-2_2.

64 Lei, Y., Shu, H.K., Tian, S.B., Jeong, J.J., Liu, T., Shim, H., Mao, H., Wang, T.H., Jani, A.B., Curran, W.J., et al. (2018). Magnetic resonance imaging-based pseudo computed tomography using anatomic signature and joint dictionary learning. *Journal of Medical Imaging,* 5, Article ID: 034001. 10.1117/1.JMI.5.3.034001.

65 Lingala, S.G., and Jacob, M. (2013). Blind compressive sensing dynamic MRI. *IEEE Transactions on Medical Imaging,* 32, 1132-1145. 10.1109/TMI.2013.2255133.