Wasserstein GANs for MR Imaging: from Paired to Unpaired Training

Ke Lei, Morteza Mardani, John M. Pauly, and Shreyas S. Vasanawala

***Abstract*—Lack of ground-truth MR images impedes the common supervised training of neural networks for image reconstruction. To cope with this challenge, this paper leverages unpaired adversarial training for reconstruction networks, where the inputs are undersampled k-space and naively reconstructed images from one dataset, and the labels are high-quality images from another dataset. The reconstruction networks consist of a generator which suppresses the input image artifacts, and a discriminator using a pool of (unpaired) labels to adjust the reconstruction quality. The generator is an unrolled neural network – a cascade of convolutional and data consistency layers. The discriminator is also a multilayer CNN that plays the role of a critic scoring the quality of reconstructed images based on Wasserstein distance. Our extensive experiments with knee MRI datasets demonstrate that the proposed unpaired training enables diagnostic-quality reconstruction when there exists insufficient, or, no high-quality image labels. In addition, our adversarial training scheme can even achieve better image quality (as rated by expert radiologists) compared with the paired training schemes with pixel-wise loss.**

***Index Terms*—Wasserstein training, convolutional neural networks (CNN), fast reconstruction, diagnostic quality.**

I. Introduction

MAGNETIC resonance imaging (MRI) is commonly used clinically for its flexible contrast. The major shortcoming of MRI is its long scan time, especially for volumetric images. Undersampling is often necessary to reduce scan time and cope with motion, but reconstructing undersampled MRI is solving an undetermined system and conventional reconstruction methods such as compressed sensing (CS) are time intensive. Recently, data-driven methods based on neural networks (NN) are adopted to reconstruct MR images with rapid reconstruction speed. However, most of these models require supervised training on a large and specific set of labels, that are fully-sampled high-quality images.

Collecting such label is expensive or impossible in certain scenarios such as dynamic imaging. For instance, in dynamic contrast enhanced (DCE) imaging, the contrast is rapidly changing, or, for deformable moving organs in the chest, abdomen, or pelvis with respiratory motion, acquiring the ground truth image is a daunting task. On the other hand, basic 2D scans for static body parts, such as extremities and brain, are often fully-sampled with high quality to serve as labels. We aim to train a model for cases where there are no, or, only limited ground truth images. This is possible with unpaired training, where the labels can be different than the images being reconstructed (i.e. the inputs).

Ample research has been conducted during the last few years on deep learning for MRI reconstruction [1]–[7]. The majority of those works use paired training which demands a large amount of labels specifically for the task they are tackling. There are only a few attempts to cope with label scarcity, as in [8]–[14], using self-supervision and transfer learning. Unpaired training with adversarial objectives is an alternative that has been explored in computer vision for natural image translation tasks [15]. However, for medical imaging tasks it introduces the risk of hallucinating images that may adversely affect the subsequent diagnosis. The methods in [16]–[21], although adopting adversarial objectives, are still paired and rely heavily on some pixel-wise supervision, such as the ℓ_1 distance, for stablizing the training and reducing the hallucination risk. Adversarial methods used in these works were adopted from entropic generative adversarial networks (EGANs) [22] or least-squares GANs (LSGANs) [23]. Without the pixel-wise supervision, these methods return images with coherent artifacts.

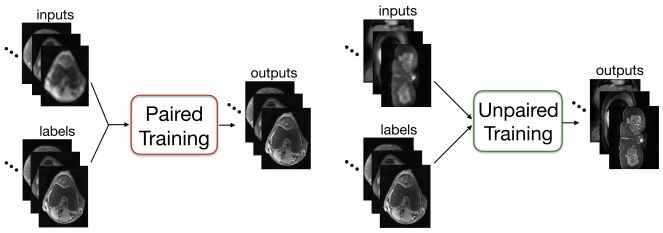

Fig. 1. High-level illustration of paired (left) vs. unpaired (right) training.

We introduce an unpaired training scheme for MRI reconstruction by leveraging adversarial training based on Wasserstein distance [24] and data consistency (DC). Our training scheme involves two networks, a generator (G) and a discriminator (D), which are trained simultaneously and interactively. The G performs the reconstruction by taking in the undersampled k-space and outputting diagnostic-quality images. It can take various kind of network architectures and types of data consistency. The D is a multilayer CNN that takes in the image reconstructed by the G and the image in the label set, and outputs a real number that reflects the distance between the two, derived from the Wasserstein distance [25]. Our model learns to approximate a desired distribution by this adversarial training process which does not require pairing between the

† Work in this paper was supported by the NIH R01EB009690 and NIH R01EB026136 award, and GE Precision Healthcare. The authors are with Stanford University, Departments of Electrical Engineering and Radiology. Emails: kllei, morteza, pauly, vasanawala@stanford.edu. Part of the results have been submitted to and presented at the 27th annual meeting of International Society of Magnetic Resonance in Medicine (ISMRM), Montreal, Canada, May 2019.

input and label.

Our proposed scheme is examined with different NN models under different scenarios of label availability. We perform extensive experiments with real-world knee and abdominal MRI datasets to show that: 1) unpaired adversarial training based on Wasserstein distance is superior to ℓ_1-supervised training and other adversarial training, 2) unpaired training can be applied when there is a small amount of labels available for the inputs or a separate set of available labels outside the input set, 3) adding Wasserstein distance based adversarial training objective to ℓ_1-supervision is the best choice when paired training is available.

Notation. The operators $\mathbb{E}[\cdot]$, $(\cdot)^{\mathsf{H}}$, $\odot$, $\mathcal{F}\{\cdot\}$, and $\mathcal{F}^{-1}\{\cdot\}$ denote the statistical expectation, matrix Hermitian, Hadamard product, 2D discrete Fourier transform (DFT), and inverse 2D discrete Fourier transform (IDFT), respectively. $\|\cdot\|_1$ and $\|\cdot\|_2$ refer to ℓ_1 and ℓ_2 norm, respectively.

II. Problem Statement and Preliminaries

A. Problem statement

MRI reconstruction, in a simplified standard setting, solves a linear inverse system $Y = \Phi(y) + u$, where Φ captures the forward model of an MRI examination and u captures the noise and uncertainties in the system, to find image $y \in \mathbb{C}^n$ from partial frequency domain samples $Y \in \mathbb{C}^m$ ($m < n$). Our goal is to learn an inverse mapping G so that for test data Y we can automatically recover its corresponding y as $G(Y)$. We approximate this mapping by a trained NN. Normally, training such a NN requires a set of inputs $I = \{Y_i\}_{i=1}^M$ and a set of labels $L = \{y_j\}_{j=1}^M$ because a traditional pixel-wise supervised training objective is defined on pairs of Y_i and y_j where $i = j$. We use *paired training* in this case.

In this paper, we consider two scenarios where a set of noisy inputs I is easily available but its corresponding label set L is not available. First, we have a 'partial' label set

$$L_p = \{y_j\}_{j=1}^N \ (N \ll M), \ L_p \subset L.$$

Second, we have a 'disjoint' label set

$$L_d = \{y_j\}_{j=M+1}^{M+1+q} \ (q \in \mathbb{Z}^+), \ L_d \cap L = \varnothing.$$

That is, for our training dataset, we either have a limited number of labels for the inputs or a different set of inexpensive labels, so pairs of Y and y cannot be used for the training. We use *unpaired training* in these two cases. In the sequel, we use adversarial training based on the Wasserstein distance for unpaired learning.

B. Wasserstein distance

Wasserstein distance is a measure of the distance between two probability distributions [25]. We particularly look at Wasserstein-1 distance in this paper. Here we first introduces Wasserstein-1 distance in its original definition which is intuitive but intractable to optimize on, then transform it into a tractable form which can be approximated by computationally efficient training objectives.

Wasserstein-1 distance is also known as the earth-mover's (EM) distance (see Fig. 3). This quantity intuitively reflects the minimum cost (i.e., mass times distance) to transport a pile of sand to another pile (with different location and shape). One advantage of this metric is that it is continuous and differentiable almost everywhere, unlike the Jensen-Shannon (JS) divergence deployed by the original EGANs [22], and χ^2 distance deployed by the LSGANs [23]. Formally, the Wasserstein-1 distance between probability mass P_r and P_g is defined as

$$W(P_r, P_g) = \inf_{J \in \mathcal{J}(P_r, P_g)} \int \|a - b\|_1 \, dJ(a, b) \qquad (1)$$

where $\mathcal{J}(P_r, P_g)$ is the set of all joint distributions for a and b whose marginals are P_r and P_g (both defined on a compact space $\mathcal{X}$), respectively.

The infimum in (1) is highly intractable, but using Kantorovich-Rubinstein duality [26] one can alternatively write it as

$$W(P_r, P_g) = \sup_{\|f\|_L \le 1} \mathbb{E}_{a \sim P_r}[f(a)] - \mathbb{E}_{b \sim P_g}[f(b)] \qquad (2)$$

where the supremum is over all 1-Lipschitz functions $f : \mathcal{X} \to \mathbb{R}$. f is 1-Lipschitz if $|f(a) - f(b)| \le \|a - b\|$. In practice, it is hard to enforce the Lipschitz continuity constraint directly. [27] introduces a viable alternative using gradient-norm regularization as would be discussed later. According to [27, Proposition 1], there is an 1-Lipschitz function f^* which maximizes $\mathbb{E}_{a \sim P_r}[f(a)] - \mathbb{E}_{b \sim P_g}[f(b)]$. This f^* has gradient norm equal to 1 almost everywhere under P_r and P_g. So we aim to search for a f whose gradient norm is close to 1 in order to minimize the Wasserstein distance.

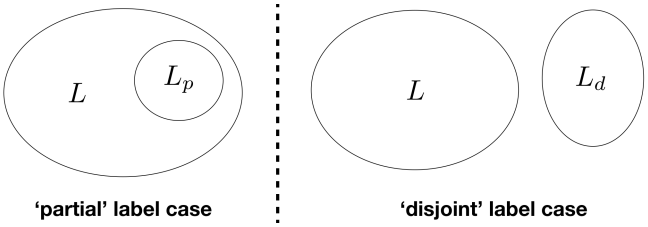

Fig. 2. Venn diagrams for two cases of label availability. L is the set of labels required for paired training. L_p and L_d are the set of labels used by unpaired training in two cases, respectively.

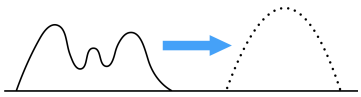

Fig. 3. Illustration of the sand pile transport explanation for earth mover's distance.

III. Adversarial Training of Neural Networks

Theoretically, adversarial training is derived to learn a desired probability distribution by minimizing some distance between the generated data (i.e. model outputs) distribution and the label data distribution. In this paper, we use the Wasserstain-1 distance introduced in the previous section. Practically, adversarial training refers to the scheme when two networks, the generator (G) and discriminator (D), are trained

simultaneously with feedback from each other's output. This scheme is used for our unpaired training approach as it does not required pixel-wise supervision. Adversarial training can also be combined with pixel-wise supervised training in the paired case.

A. Unpaired training

The ground-truth label is often not present for certain imaging scenarios. It is thus important to replace the pixel-wise supervision completely so that no pairing is needed. Then, one can leverage the available labels from other datasets that are more amenable to fully sampled acquisition. For instance, using 2D images as training labels for cine or 3D imaging inputs. In principle, adversarial training aims to approximate, in terms of some measures of distances, a probability distribution of interest: a distribution of images in the label set. Thus there is no need for each specific label to be the corresponding ground-truth of the input. Unpaired training has proved successful in image style transfer tasks (for instance, converting zebras to horses, and vice versa) such as [15] with adversarial training alone. These tasks in natural images do not necessarily require authentic output images.

Adversarial training without paired supervision for medical images, however, introduces a hallucination risk. The pixel authenticity is crucial and needs to be guaranteed. Fortunately, for the considered de-aliasing problem one has the k-space data and the forward model at hand to somewhat enforce the G outputs to adhere to the k-space data. This is ensured by the DC layers embedded into the G network. DC partially alleviates the hallucination risk, but overfitting is still a risk. This mainly emanates from the unstable training of GANs with stochastic gradient descent. EGANs and LSGANs training objectives are derived from JS and χ^2 divergences, respectively, which are not continuous (w.r.t. network parameters) for disjoint distributions. As a result, the gradients are not informative for training.

Wasserstein-1 distance is continuous under disjoint distributions. See [24] for concrete examples. Note that image distributions are in high-dimensional spaces and often disjoint. Therefore, we use Wasserstein GAN (WGAN) [24] objectives, derived from the Wasserstein-1 distance, for our unpaired training. Figure 4 illustrates the unpaired training procedure of our model. Intuitively, a D network serves as a critic which scores the images reconstructed by a G network by giving an estimate of the Wasserstein-1 distance between the G output and the label, and the G is optimized based on the feedback from D. Formally, the D network serves the role of f in equation (2) and we train it to approximate f^*. Since our goal is for the reconstructed images to be as good as the labels, let the labels $y \sim P_r$ and the output from G $G(x_{\text{zf}}) \sim P_g$. The G aims to minimize $W(P_r, P_g)$ with a given f. Then from equation (2) (under some assumptions [24]) we have the principle version of the adversarial training objective

$$\min_{G} \max_{\|D\|_L \leq 1} \mathbb{E}_{y\sim P_r}[D(y)] - \mathbb{E}_{G(Y)\sim P_g}[D(G(x_{\text{zf}}))] \qquad (3)$$

where the maximum is over all the 1-Lipschitz functions D.

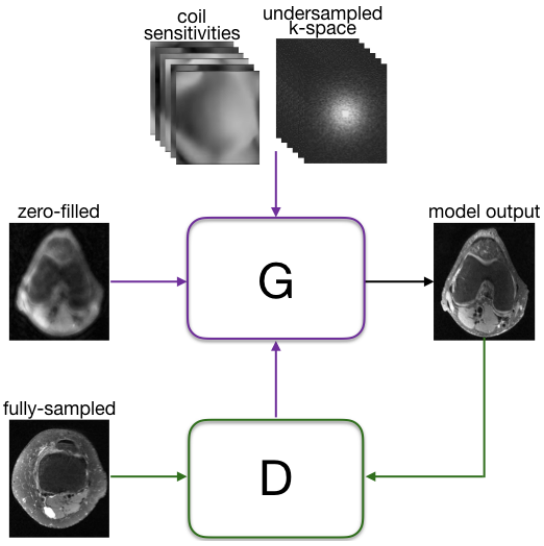

Fig. 4. Unpaired adversarial training.

As introduced in II.B, the Lipschitz constraint on the D is enforced by searching for the 1-Lipschitz function f^* which has the end-to-end gradient norm equal to unity. It is still not computationally practical to enforce this gradient norm everywhere. [27] introduces the gradient penalty (GP) term that penalizes the gradient norm of D w.r.t. random samples drawn from real and fake distributions from diverging from 1. Adopting this GP term and rearranging (3), we finally arrives at the following differentiable and fast-to-compute training objectives that approximately minimize the Wasserstein-1 distance defined in (1). The training objective for the D is

$$\text{(P1.D)} \quad \min_{\Theta_d} \mathbb{E}\big[D(G(x_{\text{zf}}); \Theta_d)\big] - \mathbb{E}\big[D(y; \Theta_d)\big] + \eta\mathbb{E}\big[(\|\nabla_{\hat{x}} D(\hat{x}; \Theta_d)\|_2 - 1)^2\big]$$

where Θ_d is the network parameters in D and η controls the strength of the GP. The random sample $\hat{x} := \alpha G(x_{\text{zf}}) + (1 - \alpha)y$ with $0 \leq \alpha \leq 1$.

The specific training objective for the G is derived directly from (3) as

$$\text{(P1.G)} \quad \min_{\Theta_g} -\mathbb{E}\big[D(G(x_{\text{zf}}; \Theta_g))\big]$$

where Θ_g is the network parameters in G and $x_{\text{zf}} = \Phi^{-1}(Y)$ is the zero-filled (ZF) image (inverse Fourier reconstruction from the ZF undersampled k-space measurements) input to the G. We refer to the above two equations as the unpaired WGAN objectives.

SGD algorithm. Θ_g, Θ_d are updated in an alternating fashion based on stochastic gradient descent (SGD) to optimization for (P1.D) and (P1.G) during training for each mini-batch of size b. First, the random samples $\{\hat{x}_i\}_{i=1}^b$ are drawn by uniformly sampling b different αs and linearly combining the corresponding G output and label in the current mini-batch. The mini-batch gradient of (P1.D) w.r.t. Θ_d is calculated given the labels $\{y\}_{i=1}^b$, the G outputs $\{x_i\}_{i=1}^b$, and random samples. Likewise, the G gradient (P1.G) is calculated, and the gradient steps are updated iteratively.

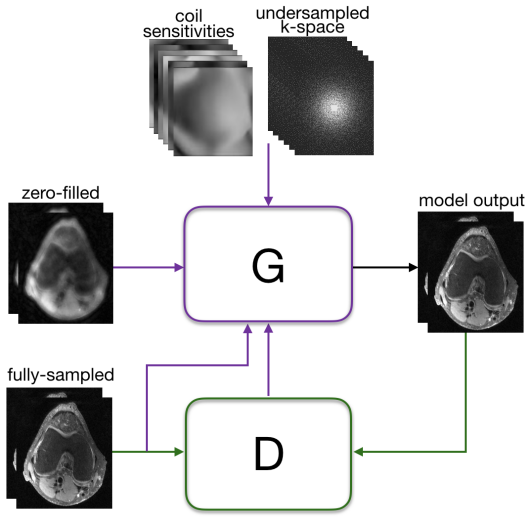

Fig. 5. Paired training with adversarial objectives.

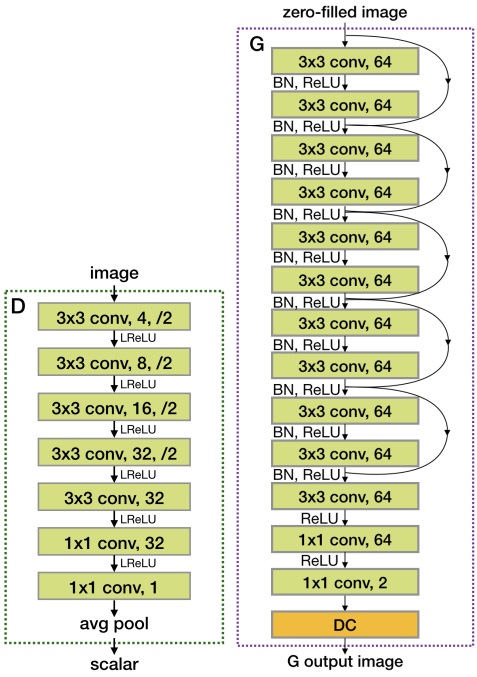

Fig. 6. The discriminator network (left) and plain generator network with 'hard' DC (right). BN and ReLU are applied after the summation with skip connections.

B. Paired training

Supervised learning of the inverse mapping is common in the MR imaging context using pixel-wise losses. These approaches achieve stable training but the resulting images are typically blurry especially at high undersampling rates. [18] shows that adding adversarial training to the pixel-wise supervised training improves the sharpness and perceptual quality of the reconstructed images. LSGAN [23] objectives are combined with pixel-wise ℓ_1 supervision in their work where the ℓ_1 supervision helps control the high-frequency noise and stabilize their training.

In our case, although the adversarial training alone is already relatively stable, adding more supervision when possible with a pixel-wise objective further improves the reconstruction quality. We find a pure ℓ_1 objective gives superior results than a pure ℓ_2 objective so ℓ_1 is used for the pixel-wise supervision. Now G aims to output images close to its ground-truth label in terms of ℓ_1 distance, and simultaneously gain a high score from D. The pixel-wise supervision is added to the G objective in (P1.G) which becomes

$$\min_{\Theta_g} -(1-\lambda)\mathbb{E}\big[D(G(x_{\text{zf}};\Theta_g))\big] + \lambda\mathbb{E}\big[\big\|y - G(x_{\text{zf}};\Theta_g)\big\|_1\big]. \tag{4}$$

We consider two models when paired training is possible. When $\lambda < 1$, the D is trained with the same objective defined by (P1.D), and we refer to the model as WGAN+ℓ_1 hybrid model. We find that starting with $\lambda = 1$ and linearly increasing it with training steps provides a more refined initial phase and leads to a higher-quality final output. When $\lambda = 1$, the training loss is the (paired) ℓ_1 loss, only the G is involved, and we refer to the model as ℓ_1-net. This is the traditional pixel-wise supervised paired training. Figure 5 illustrates the paired training procedure of our hybrid model.

C. Generator networks with data consistency

The G network takes the zero-filled input image, x_{zf}, which is simply an inverse DFT on the fully-sampled k-space masked (i.e. element-wise multiplied) by zeros and ones. Two channels are used to represent real and imaginary parts of a image separately. The G is supposed to output a high quality version of its input image as visually close as possible to the labels quality. Our training methods work with various types of G networks and data consistencies: from a standard ResNet [28] with a 'hard' DC (Fig. 6) (as that in [18]) to the state-of-the-art unrolled network with iterative 'soft' DC (Fig. 7).

Unrolled networks were introduced recently and show superior performance for image recovery and restoration tasks [29]–[33]. They are inspired by iterative inference algorithms [34]. The iterative process can be envisioned as a state-space model which at the k-th iteration takes an image estimate x_k, moves it towards the affine subspace of data consistent images, and then applies a proximal operator to obtain x_{k+1}. The state-space model is expressed as

$$v_{k+1} = g(x_k) \tag{5}$$

$$x_{k+1} = NN(v_{k+1}) \tag{6}$$

where g is a DC operation with a learnable step size μ that combines the ZF data with output of the previous iteration, x_k. Unfolding this recursion for a fixed number K of iterations, one ends up with a recurrent NN (Fig. 7), where x_K is the generator output.

The data consistency (DC) step ensures the k-space of the generated image is consistent with the actual input k-space data. 'Soft' DC used in the unrolled network is a gradient descent step [35]

$$g(x) = x + \mu\left[\sum_{i=1}^{c} \mathcal{F}^{-1}\big\{\Omega \odot \mathcal{F}\{x \odot s_i\}\big\} \odot s_i^H - x_{\text{zf}}\right] \tag{7}$$

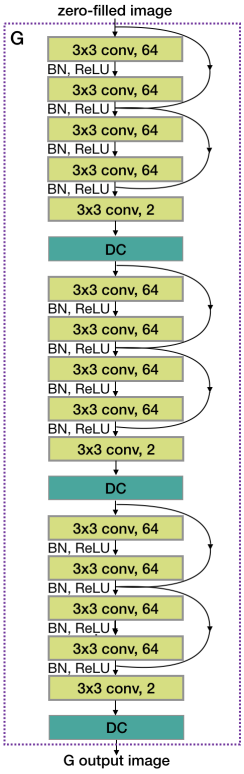

Fig. 7. The unrolled generator network with 'soft' DC.

where there are c coil maps $s_i \in \mathbb{C}^n$. Alternatively, a simpler 'hard' DC can be used at the end of a plain network with no need for learnable parameters:

$$g(x) = \sum_{i=1}^{c} \mathcal{F}^{-1}\{Y^i + (1-\Omega) \odot \mathcal{F}\{x \odot \mathbf{s}_i\}\} \odot \mathbf{s}_i^H \quad (8)$$

where $Y^i \in \mathbb{C}^n$ is the binary Ω masked k-space measurement from the ith coil.

D. Discriminator network

D takes two kinds of inputs: the G output and the labels. For paired training, both inputs are complex-valued and represented by two channels. For unpaired training, D can also take single-channel magnitude images. We explore this relaxation so that datasets which consist of only magnitude images and no k-space data can also be used as labels. The G output is always complex-valued and is converted to magnitude image before feeding to the D when the label is magnitude image.

D outputs a real-valued scalar. A 7-layer plain CNN is used for D as shown in Fig. 6, where the architectural details are provided . For the first four layers, the number of feature maps is doubled from 4 to 32, and a stride of 2 is used. Leaky ReLU nonlinearity (LReLU) [36] activation is used for all layers except the last one. The last layer averages out the seventh layer features to end up with a scalar score.

IV. Experiments

The effectiveness of the unpaired WGAN scheme is assessed for single- and multi-coil MR acquisition models with Cartesian sampling. Extensive experiments and evaluations are performed to compare unpaired and paired models trained with various objectives. We show that WGAN is most suitable for unpaired training, WGAN unpaired training is better than ℓ_1 supervised paired training, unpaired training can be used for partial and disjoint label cases (defined in II.A), and hybrid WGAN+ℓ_1 training gives the best reconstruction quality of all.

Knee MRI dataset. This dataset[1] [37] includes 19 subjects scanned with a 3T GE MR750 whole-body MR scanner. Each subjects knee was placed in an 8-channel HD knee coil. Fully sampled images are acquired with a 3D FSE CUBE sequence with proton density weighting including fat saturation. Other parameters include FOV = 160 mm (sagittal), TR = 1550 ms, TE = 25 ms, slice thickness 0.6 mm (sagittal). For each subject we have a complex-valued 3D volume of size $320 \times 320 \times 256$. The fully-sampled data used for reference images below takes over 41 minutes to collect for one subject. Axial slices of size 320×256 are the input for training and test. For the partial label case (IV.B), 17 subjects are used for training and 2 subjects for testing. For the disjoint label and paired case (IV.C, D), 13 subjects are used for training and 6 subjects for testing. The inputs are undersampled by a variable density poisson mask Ω with a fully-sampled center of size 20×20.

A. Network architecture and training

The plain G network is a deep ResNet [28] with 5 residual blocks (RBs) followed by 3 Conv layers. The D network consists of 7 Conv layers with LReLU nonlinearity; see Fig. 6. Also, as shown in Fig. 7, the unrolled G has $K = 3$ iterations, each with two RBs. Batch normalization (BN) [38] and ReLU are used after each layer except the last Conv layer for both plain and unrolled G. We set the gradient penalty coefficient $\eta = 10$. Adam optimizer is used with the momentum parameter $\beta = 0.9$, mini-batch size 4, and learning rate 10^{-4}. For paired GAN+ℓ_1 training, $\lambda = 0.99$ is used. Fully-sampled images are windowed to increase the brightness of the labels. The model is implemented in Tensorflow and the source code is available online at GitHub[2].

B. Unpaired training with partial labels

We start with a single-coil plain G model (as in the work GANCS [18]) for the partial labels scenario. Undersampled data are obtained by applying a n-fold undersampling mask to the 'k-space' of the fully-sampled image. Fully-sampled k-space in the single-coil case is obtained by a 2D DFT of the complex-valued image reconstructed from actual fully-sampled multi-coil measurement. The inputs to the single-coil model are 3-fold undersampled.

We first show that WGAN is indeed more suitable for our task than EGAN [22] and LSGAN [23]. Here we keep using all labels used in the paired training and only remove the ℓ_1 supervision. GANCS trained without ℓ_1 objective, that is, with merely LSGAN or EGAN objective, outputs images with heavy coherent artifacts (see Fig. 8).

[1]Available at: mridata.org/list?project=Stanford%20Fullysampled%203D%20FSE%20Knees

[2]https://github.com/lisakelei/Unpaired-GANCS/

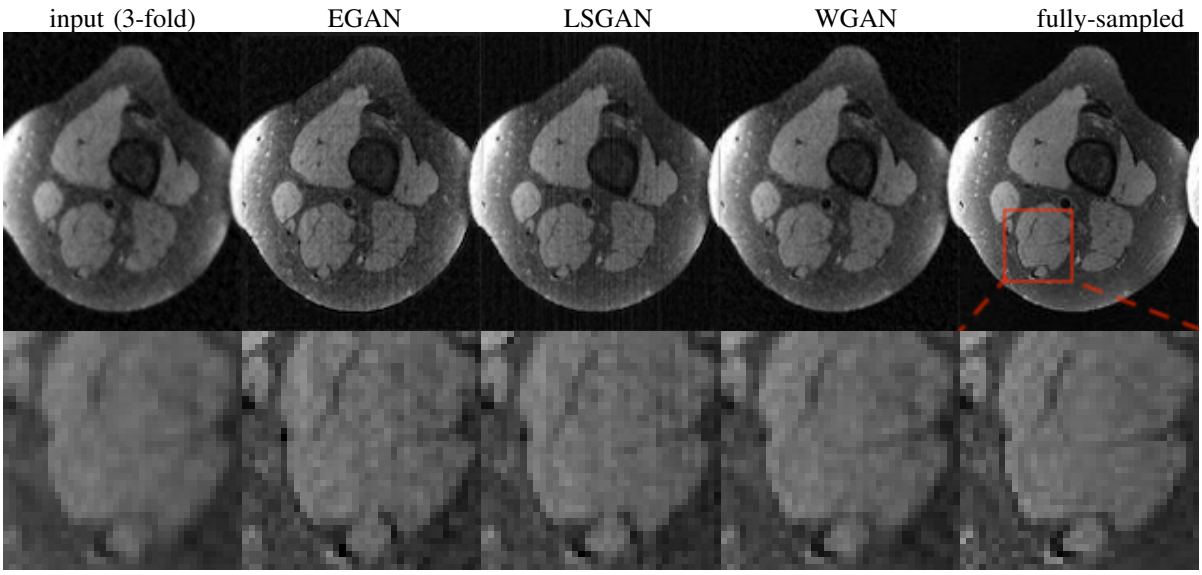

Fig. 8. A representative test sample from the single-coil plain G model trained with 6 subject labels and various unpaired GAN objectives: EGAN, LSGAN, WGAN-GP. The bottom row shows the region in red box zoomed-in.

TABLE I
QUANTITATIVE EVALUATIONS OF THE MULTI-COIL PLAIN MODEL TRAINED WITH UNPAIRED WGAN OR PAIRED ℓ_1 LOSS, WITH 5 TO 9-FOLD UNDERSAMPLED INPUTS, AND 3 TO 17 SUBJECT LABELS FOR TRAINING. THE CASES NOT INDICATED BY 'ℓ_1' ARE FOR UNPAIRED TRAINING.

Experiments	5 fold			7 fold		9 fold	
	ℓ_1 17 subjects	6 subjects	3 subjects	6 subjects	3 subjects	6 subjects	3 subjects
SNR	18.76	18.81	18.48	17.66	17.23	16.72	16.34
SSIM	0.747	0.873	0.869	0.842	0.835	0.819	0.813

We then test WGAN based unpaired training with different numbers of partially available labels. The single-coil model is trained using inputs of undersampled volumes from $M = 17$ subjects and labels of fully-sampled volumes from $N = 3$ and 6 subjects. Two sample images are shown in Fig. 9. The reconstructed images are diagnostically valuable and the result from $N = 3$ is not significantly worse than that from $N = 6$. Switching to a 8-coil model, we examine the partial label unpaired training with same M and N's but different undersampling ratios of 5, 7, and 9. Table I lists the average SNR and SSIM over slices in two test volumes. The SNR of each output image is obtained by averaging pixel-wise SNRs where the signal is the pixel value of the fully-sampled image and the noise is the absolute difference between pixel values of the output and fully-sampled images. Note that minimizing the paired ℓ_1 loss is the same as maximizing SNR so ℓ_1-net tends to get a higher SNR regardless of its visual quality. Quality of the output images varies with the undersampling ratio (i.e. quality of the input), but outputs from $N = 3$ are not significantly worse than that from $N = 6$. The above experiments with the single- and multi-coil models show that we can push the number of labels to as little as 1/5 of that used in the paired case.

All subsequent experiments are done with a multi-coil model with k-space data from 8 coils. The coil sensitivities are extracted by the ESPIRiT algorithm [39].

Fig. 10 shows the outputs from unpaired WGAN loss compared with that from paired ℓ_1 training. Compared to a model trained with pixel-wise losses, our model trained with pure WGAN loss not only allows for using fewer labels but also generates images with more realistic texture. Pixel-wise paired training (with double the labels of the unpaired training) while refining the edges better, oversmooths images.

Now we have shown that the proposed unpaired training is adequate in the case of partial label and we can achieve $N < \frac{1}{5}M$.

C. Unpaired training with disjoint datasets

We now switch to the unrolled G which gives more accurate images (with around 2dB better SNR) compared to plain G and explore a more relaxed setting for the labels where there is no overlap between the input and label sets. Among the 13 subjects in the training set, undersampled raw data from 7 subjects are inputs, and fully-sampled magnitude images from 6 other subjects are labels. This setting reflects the case when we want to train a model for a dataset without any label using high-quality labels from some other datasets.

We train the model with pure WGAN-GP [27] objective on 10-fold undersampled inputs. The inference sample and quantitative score from this model along with some other models are shown in Fig.11 and 12 and Table II. The quantitative scores are averaged over 1920 test slices, and only the center 272 x 216 region out of a 320 x 256 image is used.

The conventional CS method can be used in this disjoint dataset set case thus included in the comparison. We use the CS-Wavelet implementation by the BART [40] toolbox. The

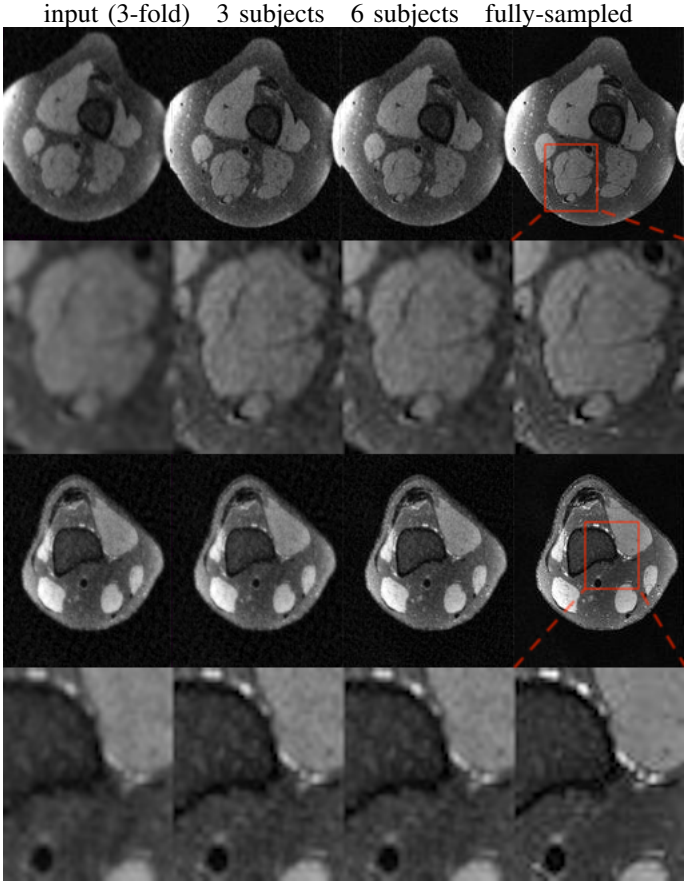

Fig. 9. Two representative test samples from the single-coil model. From left to right: 3-fold undersampled input, output from our model trained with 3 subject and 6 subject labels, and fully-sampled reference.

regularization parameters 0.05 is tuned to optimize for the perceptual quality of a small evaluation set. Sample reconstructed slices are also shown in Fig. 11 and Fig. 12.

We test our model on a DCE abdominal dataset where there is no fully-sampled data in reality. Here the input set and label set are disjoint because they are abdominal and knee datasets. Fig. 13 shows a test sample from our unrolled unpaired model comparing with the CS-Wavelet reconstruction.

Overall, the comparisons among these schemes and configurations indicate unrolled ResNets with WGAN training as the viable alternative to CS.

D. Paired training

In this section, we consider a supervised scenario with input and label pairs from 6 subjects. The network is trained with unrolled G and two different objectives.

We train the hybrid WGAN+ℓ_1 model with the first 500 batches with pure ℓ_1 objective, then linearly decrease λ to 0.99 within the first 1000 mini-batches. This is useful to stabilize training and improve final performance. We also train a model with only ℓ_1 objective (ℓ_1-net). Two test slices from these two models are shown in Fig.11 and 12, and the quantitative scores are shown in Table II. The conclusion is that when paired training is possible, adding WGAN objective to the classic ℓ_1-minimization leads to results that are visually sharper with higher SNR.

E. Radiologist evaluation

We notice the standard quantitative metrics (i.e. SNR, PSNR, MSE, SSIM, etc.) including those reported above do not reflect the visual quality of the reconstructed images well. To assess the diagnostic image qualities from different reconstruction methods, we perform an experiment based on the consensus of two radiologists. We asked them to rank the reconstructed volumes given by four reconstruction methods together with the fully-sampled volume according to five aspects: sharpness, level of coherent artifacts, visibility of anterior cruciate ligament (ACL), medial meniscus (MM) and medial collateral ligament (MCL). ACL, MM, and MCL are three structures in the knee that are commonly assessed.

Image volumes from 6 different subjects, 5 versions each, (30 volumes in total) are used for this test. Radiologists were blinded to the reconstruction schemes. Horos [41] software interface is used to visualize the images. For each subject, the five volumes (for different reconstructions) are ranked from best to worst with ties possible. We then convert the rankings to scores; the best score is 5, and the worst one is 1. When there is a tie, we take an average of the scores; for example, if second and third best scores are equally good, both would receive the score $\frac{4+3}{2} = 3.5$. The scores for all NN based methods are presented in Fig. 14.

F. Inference time

Table III shows the reconstruction time per 2D slice on a NVIDIA TITAN Xp GPU [42], averaged across two test volumes. The timing starts after the initial data reading and ends before the final data writing. The programs run in the same terminal. The CS reconstruction in BART [40] is done one volume at a time. Our generators reconstruct slice by slice but we time each volume as for the CS. The per slice time is obtained by dividing the total time by the number of slices. The 3-iteration unrolled network is only slightly slower than the plain network, while both NN based models are significantly (16 times) faster than the conventional CS-Wavelet.

V. Conclusions

This paper advocates an unpaired deep learning scheme for MRI reconstruction when high-quality training labels are scarce. Leveraging Wasserstein GANs with gradient penalty, a generator network based on plain or unrolled ResNets maps linear image estimates to mimic the image label distribution. The discriminator network then plays the role of a critic that estimates the distance of generator output images from the label images. The unpaired training objectives alleviates the need for pairing among the undersampled input and the high-quality labels. Our work far extends the scope of prior work [18] for imaging scenarios with scarce training labels and more realistic multi-coil models. Extensive experiments on knee and abdominal MRI datasets – deploying various network architectures under different data configurations and training

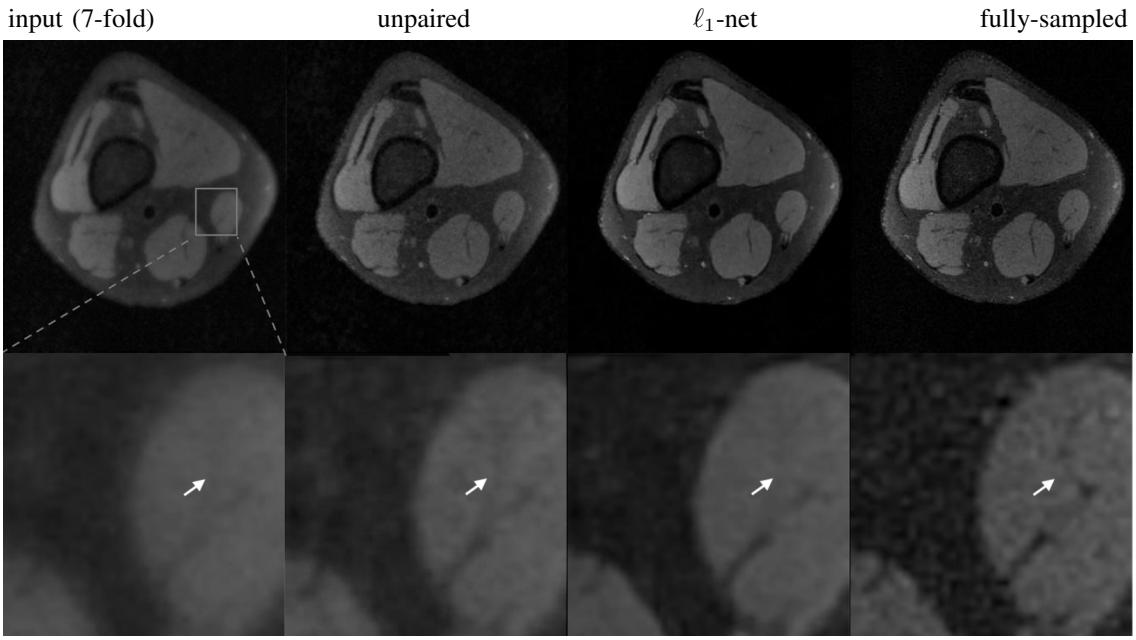

Fig. 10. A representative test sample from the multi-coil plain model trained with various objectives. From left to right: 7-fold undersampled input, output from our unpaired model trained with 6 subject labels, output from the same G trained with pure ℓ_1 loss and 17 subject labels, fully-sampled reference.

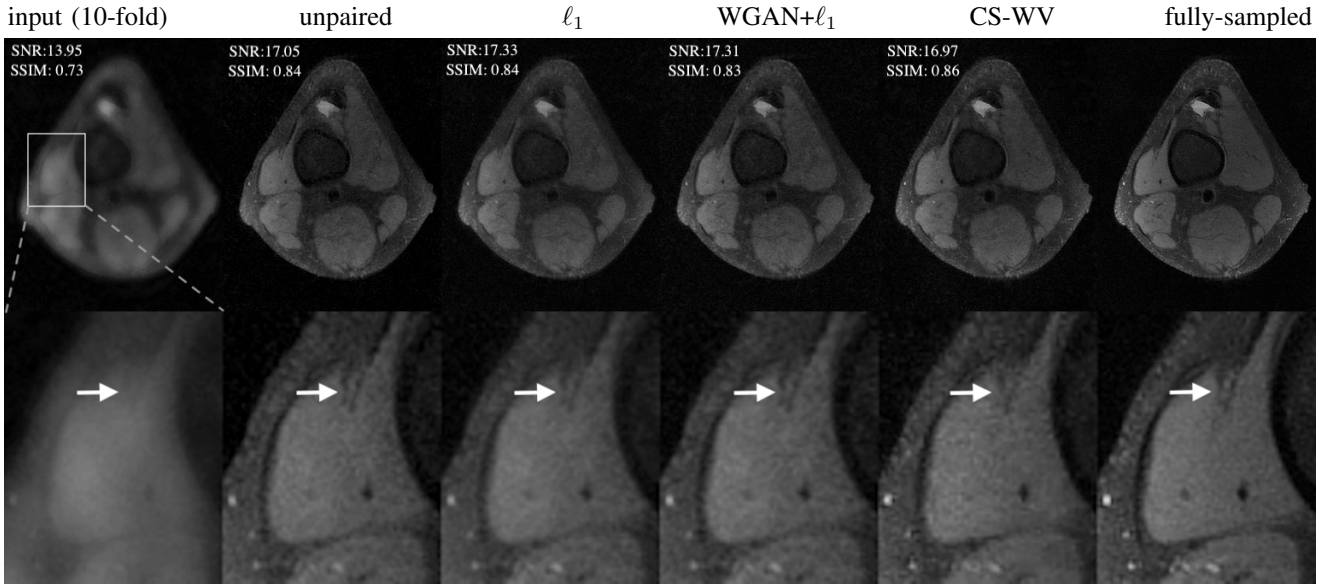

Fig. 11. Test samples from the multi-coil unrolled model trained with various objectives compared with CS. From left to right: 10-fold input, outputs from unpaired model trained with WGAN, paired model trained with ℓ_1, paired model trained with WGAN and ℓ_1, CS-Wavelet, fully-sampled reference.

TABLE II
QUANTITATIVE EVALUATIONS OF THE MULTI-COIL UNROLLED MODEL TRAINED WITH DIFFERENT OBJECTIVES COMPARED WITH CS-WV.

	input (10-fold)	unpaired WGAN	ℓ_1-net	WGAN+ℓ_1	CS-Wavelet
SNR	12.57	16.52	16.84	17.05	16.96
SSIM	0.692	0.822	0.828	0.824	0.847

TABLE III
INFERENCE TIME OF PLAIN G, UNROLLED G, AND CS.

Method	plain G	unrolled G	30-iter CS
Second per slice	0.022	0.025	0.4

schemes – corroborate the efficacy of Wasserstain distance based adversarial, and most importantly, unpaired, training with DC to give faithful reconstruction of MRIs and is a viable alternative to slow conventional methods.

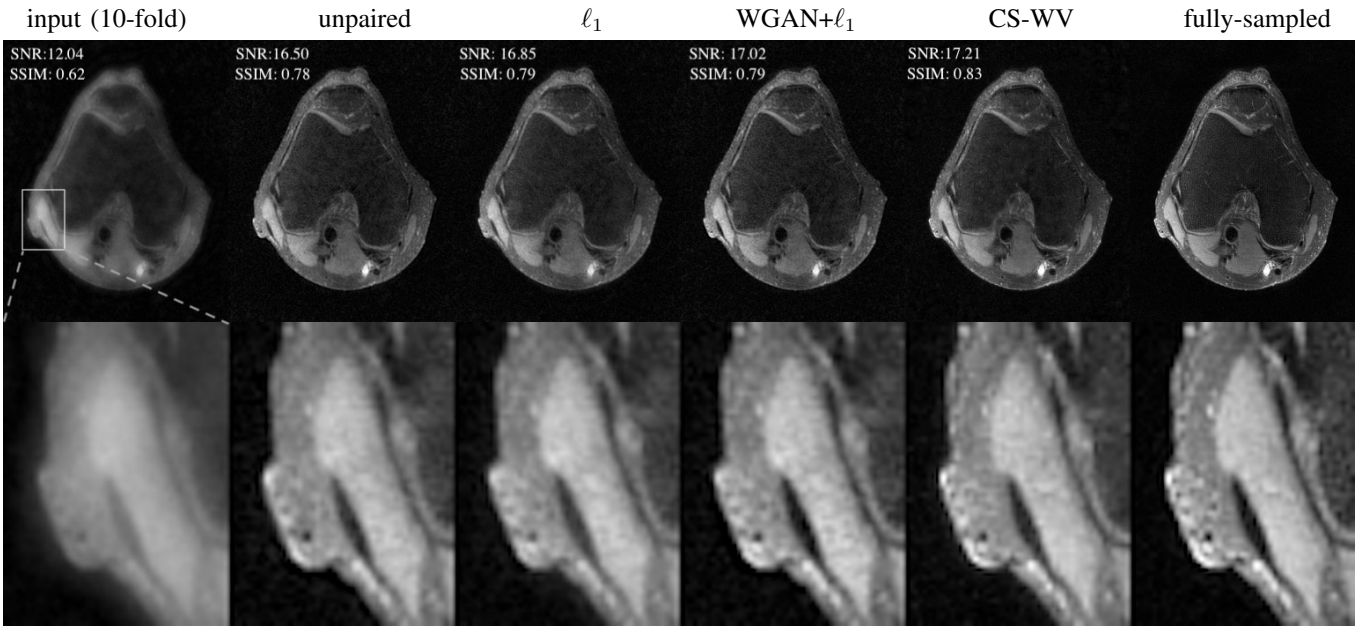

Fig. 12. Test samples from the multi-coil unrolled model trained with various objectives compared with CS. From left to right: 10-fold input, outputs from unpaired model trained with WGAN, paired model trained with ℓ_1, paired model trained with WGAN and ℓ_1, CS-Wavelet, fully-sampled reference.

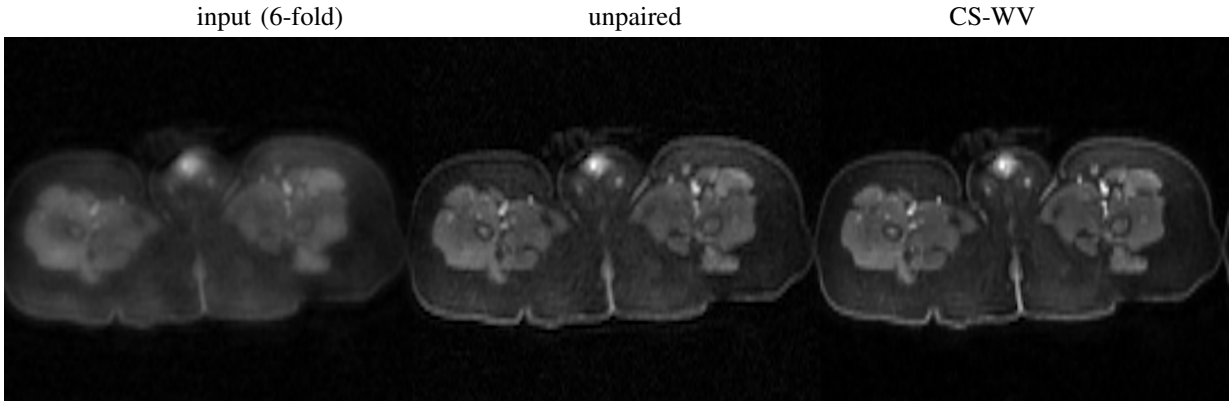

Fig. 13. A representative test sample of the unpaired multi-coil unrolled model and CS-Wavelet reconstruction on DCE abdominal input.

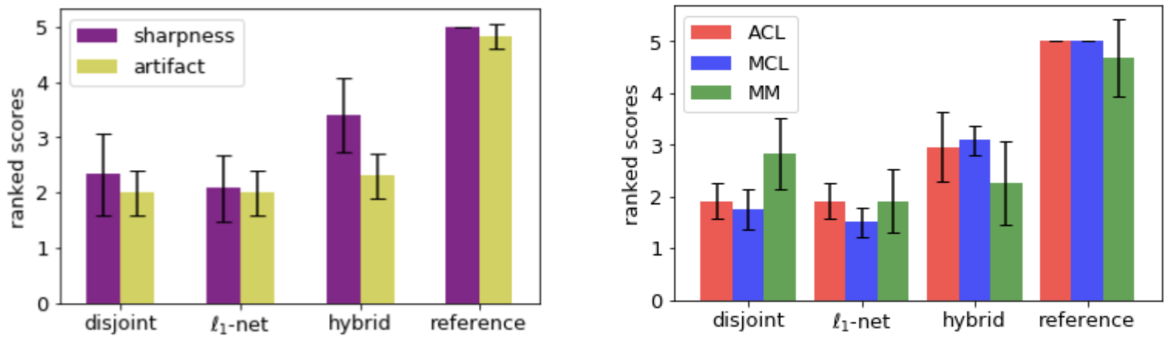

Fig. 14. Ranked score from radiologist review for outputs from disjoint unrolled model, paired unrolled ℓ_1 model, paired unrolled WGAN+ℓ_1 (hybrid) model, and fully-sampled reference. Aspects of rating: sharpness of the image, level of coherent artifacts, and visibility of three knee structures: ACL, MCL, MM.

In particular, the proposed unpaired training works with 1/5 of the labels needed in the paired case and when the training input and label are from disjoint datasets. When pairing is possible, training an unrolled network with WGAN+ℓ_1 objective is the optimal choice and in some cases better than CS-Wavelet reconstruction. All of our NN based models are

16 times faster than CS-Wavelet reconstruction.

REFERENCES

[1] B. Zhu, J. Z. Liu, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain transform manifold learning," *Nature*, vol. 555, 03 2018.

[2] F. Knoll *et al.*, "Deep Learning Methods for Parallel Magnetic Resonance Image Reconstruction," *arXiv e-prints*, p. arXiv:1904.01112, Apr 2019.

[3] F. Chen *et al.*, "Data-driven self-calibration and reconstruction for non-cartesian wave-encoded single-shot fast spin echo using deep learning," *Journal of Magnetic Resonance Imaging*, 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.26871

[4] C. M. Hyun, H. P. Kim, S. M. Lee, S. Lee, and J. K. Seo, "Deep learning for undersampled MRI reconstruction," *Physics in Medicine & Biology*, vol. 63, no. 13, p. 135007, June 2018.

[5] J. Y. Cheng, F. Chen, C. Sandino, M. Mardani, J. M. Pauly, and S. S. Vasanawala, "Compressed Sensing: From Research to Clinical Practice with Data-Driven Learning," *arXiv e-prints*, p. arXiv:1903.07824, Mar 2019.

[6] C. Qin, J. Schlemper, J. Caballero, A. N. Price, J. V. Hajnal, and D. Rueckert, "Convolutional Recurrent Neural Networks for Dynamic MR Image Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 38, pp. 280–290, 2017.

[7] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, "A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 491–503, Feb 2018.

[8] K. C. Tezcan, C. F. Baumgartner, and E. Konukoglu, "MR image reconstruction using the learned data distribution as prior," *CoRR*, vol. abs/1711.11386, 2017. [Online]. Available: http://arxiv.org/abs/1711.11386

[9] J. Tamir, S. Yu, and M. Lustig, "Unsupervised deep basis pursuit: Learning reconstruction without ground-truth data," in *Proceedings of the 27th Annual Meeting of ISMRM*, 2019.

[10] F. Ong and M. Lustig, "k-space Aware Convolutional Sparse Coding: Learning from Undersampled k-space Datasets for Reconstruction," in *Proceedings of the 26th Annual Meeting of ISMRM*, 2018.

[11] F. Chen, J. Y. Cheng, J. M. Pauly, and S. S. Vasanawala, "Semi-Supervised Learning for Reconstructing Under-Sampled MR Scans," in *Proceedings of the 27th Annual Meeting of ISMRM*, 2019.

[12] K. H. Jin, M. Unser, and K. M. Yi, "Self-supervised deep active accelerated MRI," *CoRR*, vol. abs/1901.04547, 2019. [Online]. Available: http://arxiv.org/abs/1901.04547

[13] S. U. H. Dar and T. Çukur, "A transfer-learning approach for accelerated MRI using deep neural networks," *CoRR*, vol. abs/1710.02615, 2017. [Online]. Available: http://arxiv.org/abs/1710.02615

[14] J. Lehtinen *et al.*, "Noise2noise: Learning image restoration without clean data," in *ICML*, 2018, pp. 2971–2980.

[15] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.

[16] I. Sánchez and V. Vilaplana, "Brain MRI super-resolution using 3D generative adversarial networks," *CoRR*, vol. abs/1812.11440, 2018. [Online]. Available: http://arxiv.org/abs/1812.11440

[17] G. Yang *et al.*, "Dagan: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, pp. 1310–1321, 2018.

[18] M. Mardani *et al.*, "Deep generative adversarial neural networks for compressive sensing (GANCS) MRI," *IEEE Transactions on Medical Imaging*, vol. 38, no. 1, pp. 167–179, July 2018.

[19] T. M. Quan, T. Nguyen-Duc, and W.-K. Jeong, "Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss," *IEEE Transactions on medical imaging*, vol. 37, no. 6, pp. 1488–1497, 2018.

[20] Z. Li, T. Zhang, and D. Zhang, "SEGAN: structure-enhanced generative adversarial network for compressed sensing MRI reconstruction," *CoRR*, vol. abs/1902.06455, 2019. [Online]. Available: http://arxiv.org/abs/1902.06455

[21] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 105–114.

[22] I. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[23] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Computer Vision (ICCV), 2017 IEEE International Conference on*. IEEE, 2017, pp. 2813–2821.

[24] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proceedings of the 34th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 70, 06–11 Aug 2017, pp. 214–223. [Online]. Available: http://proceedings.mlr.press/v70/arjovsky17a.html

[25] C. Villani, *The Wasserstein distances*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 93–111. [Online]. Available: https://doi.org/10.1007/978-3-540-71050-9_6

[26] ——, *Cyclical monotonicity and Kantorovich duality*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. [Online]. Available: https://doi.org/10.1007/978-3-540-71050-9_6

[27] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein GANs," in *Advances in Neural Information Processing Systems*, 2017, pp. 5769–5779.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 770–778.

[29] J. Y. Cheng, F. Chen, M. T. Alley, J. M. Pauly, and S. S. Vasanawala, "Highly scalable image reconstruction using deep neural networks with bandpass filtering," *CoRR*, vol. abs/1805.03300, 2018. [Online]. Available: http://arxiv.org/abs/1805.03300

[30] M. Mardani *et al.*, "Neural proximal gradient descent for compressive imaging," in *Proceedings of the 32Nd International Conference on Neural Information Processing Systems*, ser. NIPS'18, 2018, pp. 9596–9606.

[31] H. K. Aggarwal, M. P. Mani, and M. Jacob, "Multi-shot sensitivity-encoded diffusion MRI using model-based deep learning (MODL-MUSSELS)," *CoRR*, vol. abs/1812.08115, 2018. [Online]. Available: http://arxiv.org/abs/1812.08115

[32] Y. Yang, J. Sun, H. Li, and Z. Xu, "ADMM-CSNet: A Deep Learning Approach for Image Compressive Sensing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, pp. 521–538, 2018.

[33] Y. Chun and J. A. Fessler, "Deep BCD-Net Using Identical Encoding-Decoding CNN Structures for Iterative Image Recovery," in *2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, 2018, pp. 1–5.

[34] M. Lustig and J. Pauly, "SPIRiT: Iterative Self-consistent Parallel Imaging Reconstruction From Arbitrary k-Space," *Magnetic resonance in medicine: official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, vol. 64, pp. 457–71, 08 2010.

[35] S. Diamond, V. Sitzmann, S. P. Boyd, G. Wetzstein, and F. Heide, "Dirty pixels: Optimizing image classification architectures for raw sensor data," *CoRR*, vol. abs/1701.06487, 2017. [Online]. Available: http://arxiv.org/abs/1701.06487

[36] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, vol. 30, no. 1, 2013, p. 3.

[37] K. Epperson *et al.*, "Creation of Fully Sampled MR Data Repository for Compressed Sensing of the Knee," in *SMRT Conference*, 2013.

[38] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ser. ICML'15. JMLR.org, 2015, pp. 448–456.

[39] M. Uecker *et al.*, "ESPIRiTan eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA," *Magnetic Resonance in Medicine*, vol. 71, no. 3, pp. 990–1001, 2014.

[40] ——, "Berkeley advanced reconstruction toolbox," in *Proceedings of the 23rd Annual Meeting of ISMRM*, 2015.

[41] Nimble Co LLC d/b/a Purview, "Horos." [Online]. Available: https://horosproject.org

[42] NVIDIA. (2019) TITAN Xp Graphics Card with Pascal Architecture. [Online]. Available: https://www.nvidia.com/en-us/titan/titan-xp/