Received: 25 August 2021 | Revised: 8 February 2022 | Accepted: 22 February 2022

DOI: 10.1002/mrm.29227

RESEARCH ARTICLE

Magnetic Resonance in Medicine

High fidelity deep learning-based MRI reconstruction with instance-wise discriminative feature matching loss

Ke Wang[1,4] | **Jonathan I. Tamir**[2] | **Alfredo De Goyeneche**[1] | **Uri Wollner**[3] | **Rafi Brada**[3] | **Stella X. Yu**[1,4] | **Michael Lustig**[1]

[1]Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, California, USA

[2]Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas, USA

[3]GE Global Research, Herzliya, Israel

[4]International Computer Science Institute, University of California at Berkeley, Berkeley, California, USA

Correspondence
Ke Wang, Electrical Engineering and Computer Sciences, University of California at Berkeley, CA 94720, USA.
Email: kewang@berkeley.edu;

Funding information
Foundation for the National Institutes of Health, Grant/Award Numbers: R01EB009690, R01EB026136, R01HL136965, U01EB029427; GE Healthcare

Purpose: To improve reconstruction fidelity of fine structures and textures in deep learning- (DL) based reconstructions.

Methods: A novel patch-based Unsupervised Feature Loss (UFLoss) is proposed and incorporated into the training of DL-based reconstruction frameworks in order to preserve perceptual similarity and high-order statistics. The UFLoss provides instance-level discrimination by mapping similar instances to similar low-dimensional feature vectors and is trained without any human annotation. By adding an additional loss function on the low-dimensional feature space during training, the reconstruction frameworks from under-sampled or corrupted data can reproduce more realistic images that are closer to the original with finer textures, sharper edges, and improved overall image quality. The performance of the proposed UFLoss is demonstrated on unrolled networks for accelerated two- (2D) and three-dimensional (3D) knee MRI reconstruction with retrospective under-sampling. Quantitative metrics including normalized root mean squared error (NRMSE), structural similarity index (SSIM), and our proposed UFLoss were used to evaluate the performance of the proposed method and compare it with others.

Results: In vivo experiments indicate that adding the UFLoss encourages sharper edges and more faithful contrasts compared to traditional and learning-based methods with pure ℓ_2 loss. More detailed textures can be seen in both 2D and 3D knee MR images. Quantitative results indicate that reconstruction with UFLoss can provide comparable NRMSE and a higher SSIM while achieving a much lower UFLoss value.

Conclusion: We present UFLoss, a patch-based unsupervised learned feature loss, which allows the training of DL-based reconstruction to obtain more detailed texture, finer features, and sharper edges with higher overall image quality under DL-based reconstruction frameworks. (Code available at: https://github.com/mikgroup/UFLoss)

KEYWORDS

compressed sensing, convolutional neural network, deep learning, feature loss, image reconstruction

1 | INTRODUCTION

MRI offers tremendous benefits to both science and medicine, but unfortunately, MRI data acquisition is inherently time-consuming. As a result, there is great interest in reconstructing diagnostic quality images from limited measurements to shorten scan times. Over the past decades, numerous computational approaches have been proposed to address this problem, including parallel imaging (PI)[1-3] and compressed sensing (CS).[4] PI leverages multiple receiver coils to acquire multiple-view images simultaneously for efficient image reconstruction. CS incorporates prior information about the system and signal to constrain the image reconstruction. Both PI and CS have successfully enabled a broad range of clinical applications, and all major MRI vendors have implemented products based on them.

Nonetheless, there remain several challenges with PI and CS. (1) The regularization functions used in CS are hand-crafted (e.g., sparse transformation) or rely on simple learned features (e.g., dictionary learning[5]), which are known to be suboptimal at modeling the underlying data distribution.[6] (2) CS reconstruction is sensitive to the tuning parameters. (3) The reconstruction time of CS is relatively long due to iterative optimization.

To overcome these limitations, end-to-end deep learning (DL)-based reconstruction methods[6-12] have been proposed to learn the regularization terms directly from a large training dataset. Two representative approaches include the Variational Network[6] and Model-based Deep Learning (MoDL).[11] Both methods consist of unrolling a conventional iterative CS reconstruction and replacing the regularization step with learnable activation functions or Convolutional Neural Networks (CNNs). End-to-end training is performed in a supervised learning manner. These unrolled learning-based methods have shown great potential at further accelerating reconstruction from under-sampled k-space measurements, well beyond the capabilities of combined PI and CS (PICS).

It is well-known that the performance of DL-based methods is dependent on the loss function used for training. The most commonly used loss functions for training are pixel-wise ℓ_1, ℓ_2 and the patch-wise structural similarity index (SSIM)[13] losses.[6,7,11] However, these loss functions are usually hand-crafted or based on local statistics, which do not necessarily capture the perceptual information of fine structures, which results in images with degraded perceptual quality and blurring when compared to unaccelerated scans.[8,14]

To address these issues, Generative Adversarial Networks (GANs)[15-17] with adversarial losses have been proposed to exploit the implicit feature information by incorporating discriminators into the reconstruction pipeline.[8,14,18] Unfortunately, GANs are notoriously hard to train, easily fall into mode collapse, and are sensitive to hyperparameter selections. Additionally, the adversarial loss is a less-constrained instance-to-set loss function, where improper training parameters may result in unexpected instabilities during the training and artifacts in the reconstructions.[19,20]

Aside from the adversarial loss, recent works in computer vision have shown that CNN-based perceptual losses can be used to learn high-level image feature representations.[21,22] These perceptual loss functions are based on feature layers of classification networks (such as the VGG Net[23]). They are typically designed to work for natural images with a fixed channel number (RGB) and are usually trained in a supervised manner with human-annotated labels, for example, from ImageNet.[24] Therefore, simply using perceptual VGG losses may not be ideal for MRI reconstruction tasks. For MR datasets, the dimensionality of the data can vary from application to application (e.g., two- and three-dimensional [2D/3D] complex-valued data, 2D/3D dynamic data), while at the same time, human-annotated labels for MR images are much harder to obtain. More importantly, it is also unclear what kind of human annotations would be best for comparing the image quality for MR images.

In this work, we propose a novel unsupervised learned feature loss (Figure 1) to capture the perceptual and high-order statistical difference within MR images, which we call Unsupervised Feature Loss (UFLoss). The UFLoss is a large-patch-wise loss function that provides instance-level discrimination by mapping similar patches to similar low-dimensional feature vectors using a pretrained mapping network (which we refer to as UFLoss feature mapping network or UFLoss network).[25] The rationale of using features from large-patches (typically 40×40 pixels for a 300×300 pixels image) is that we want our UFLoss to capture mid-level structural and semantic features instead of using small patches (typically around 10×10 pixels), which only contain local edge information. On the other hand, we avoid using global features due to the fact that our training set (typically around 5000 slices) is usually not large enough to capture common and general features at a large-image scale.

Different from adversarial loss, UFLoss is a more-constrained instance-to-instance loss function, which leads to more stable training with clear and straightforward stop criterion. Meanwhile, unlike the VGG perceptual loss, pretraining the UFLoss network requires no supervision, and thus is able to capture high-level structural information specifically for MR images without any human annotations. Similar to the VGG perceptual loss, UFLoss can also be easily incorporated into the

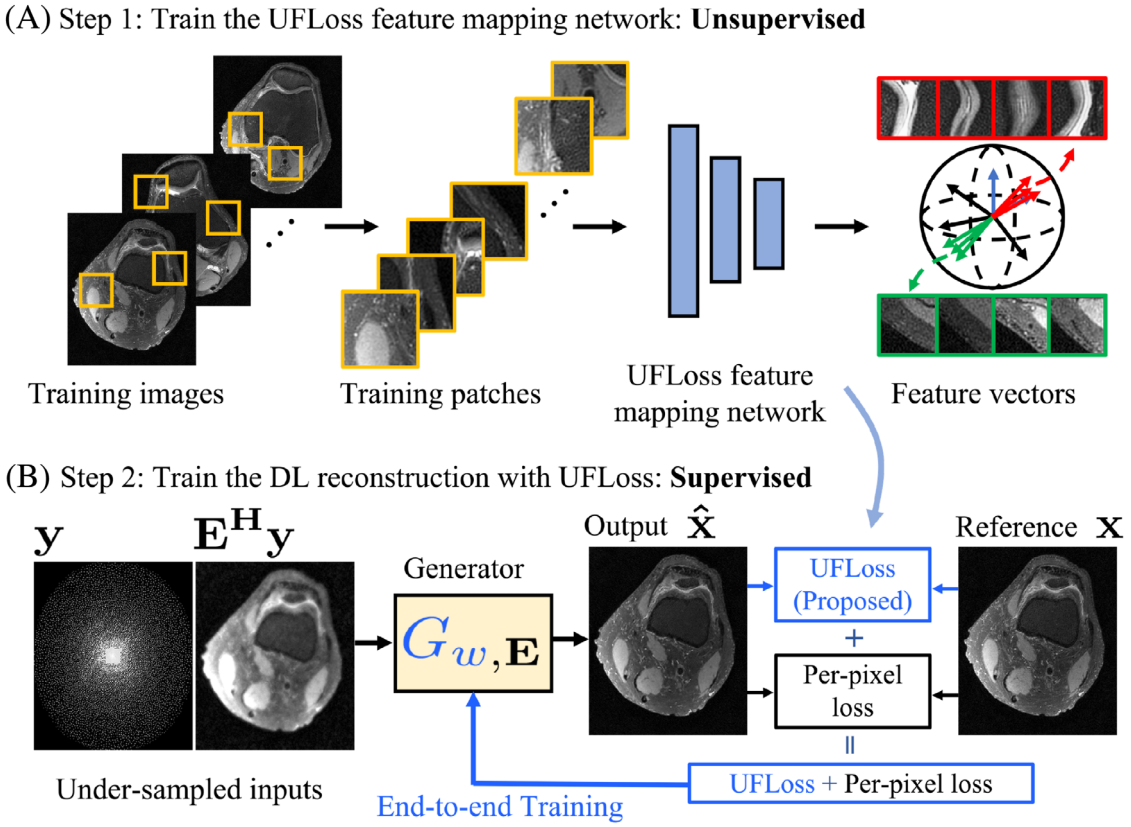


FIGURE 1 Overview of training the deep learning (DL)-based reconstruction with Unsupervised Feature Loss (UFLoss). We split the pipeline into two steps. (A) Step 1: We pretrain the UFLoss feature mapping network on fully sampled image patches without human annotations, where the aim of the training is to maximally separate out all the patches in the feature space. (B) Step 2: For the training of the DL-based reconstruction, $G_{w,\mathbf{E}}$ represents a reconstruction network with learnable parameters w, and given system encoding operator $\mathbf{E}$. The inputs of $G_{w,\mathbf{E}}$ are under-sampled k-space $\mathbf{y}$, and zero-filled reconstruction $\mathbf{E}^{\mathbf{H}}\mathbf{y}$. We feed-forward $\mathbf{E}^{\mathbf{H}}\mathbf{y}$ through $G_{w,\mathbf{E}}$ to obtain the output reconstruction results. We adopt the pretrained UFLoss network from (A) to compute the UFLoss in the feature space. Then, end-to-end training is performed with respect to the combination of UFLoss and per-pixel loss. Note that the training of DL-based reconstruction with UFLoss is still supervised

training of DL-based reconstruction networks without modifying the network architecture. Figure 1 shows the overall pipeline for using our UFLoss to train a DL-based reconstruction. We first pretrain the UFLoss network on fully sampled image patches without accompanying annotated labels (Figure 1A). This step maps patches to a lower-dimensional space while attempting to maximally separate them in the feature space. The outcome is that similar patches end up being close together in the feature space while dissimilar ones end up further apart. This pretrained feature mapping network is then adopted to compute the UFLoss during the training of the DL-based reconstruction (Figure 1B), which corresponds to the ℓ_2 distance in the feature space summed across all images patches. End-to-end training is performed with respect to a combination of UFLoss and per-pixel ℓ_1/ℓ_2 or SSIM losses.

To demonstrate the power of UFLoss, we focus on a representative unrolled DL-based reconstruction framework: MoDL.[11] We conduct experiments to show that UFLoss is a valid loss function sensitive to increasing low-level intensity deformation. Our results for patch retrieval and patch correlation in MR images demonstrate that *visually similar* patches are indeed close in the feature space.

In terms of computation costs, our UFLoss is added during training as an additional loss function without modifying the reconstruction network architecture. This imposes about 50% increase in training time and memory requirements during training. However, in inference time, the UFLoss has no impact at all on the reconstruction time as well as the memory requirements. Our experiments on 2D and 3D in vivo data show that the addition of the UFLoss encourages more realistic reconstructions with more subtle details and improved overall image quality compared to conventional and learning-based methods with other losses (pure ℓ_2 loss and ℓ_2+VGG perceptual loss).

2 | THEORY

2.1 | Unrolled reconstruction for under-sampled MRI

In conventional under-sampled MRI, the PICS inverse problem can be formulated as:[4]

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{E}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda Q(\mathbf{x}), \tag{1}$$

where $\mathbf{x}$ is the image to be reconstructed, and $\mathbf{y}$ is the measured data in k-space. $\mathbf{E}$ describes the system encoding matrix, which can be further expanded to: $\mathbf{E} = \mathbf{UFS}$, where $\mathbf{F}$ is the Fourier transform operator, $\mathbf{S}$ represents the multiple sensitivity maps, and $\mathbf{U}$ corresponds to the k-space sampling operator. For the Cartesian case, $\mathbf{U}$ is a diagonal matrix with 1s corresponding to collected k-space and 0s to unacquired k-space. For non-Cartesian, $\mathbf{U}$ is a k-space resampling operator from a Cartesian grid to the acquired non-Cartesian trajectory. The goal of this problem is to reconstruct the image which has the lowest error compared to the measured k-space data in the least-squares sense. However, when the sampling rate is below the Nyquist rate, Equation (1) becomes ill-posed. Therefore, a regularization term $Q(\mathbf{x})$ with a weighting parameter λ, which incorporates prior knowledge about the image, is added to constrain the optimization problem. For conventional CS MRI, $Q(\mathbf{x})$ is often chosen to promote sparsity in a certain transform domain such as wavelets or finite spatial differences.

A number of first-order iterative methods have been developed for efficiently solving the minimization problem in Equation (1) for the case where $Q(\mathbf{x})$ is convex.[26,27] To further develop fast and high-fidelity reconstructions, recent methods have attempted to directly learn the proximal function Q and the corresponding parameters from a large set of fully sampled training data in an unrolled fashion.[6-12]

A widely used unrolled reconstruction framework is MoDL,[11] where the reconstruction is formulated as:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{E}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{x} - D_w(\mathbf{x})\|_2^2. \tag{2}$$

In this formulation, D_w is a learned CNN denoiser/artifact removal network and w are the learned weighting parameters. The CNN-based prior $\|\mathbf{x} - D_w(\mathbf{x})\|_2^2$ results in high values when $\mathbf{x}$ is corrupted by noise and aliasing. Similar to the alternating direction method of multipliers,[27] we can solve the optimization problem in the following half-quadratic splitting steps:

$$\mathbf{z}^k = D_w(\mathbf{x}^k). \tag{3}$$

$$\begin{aligned}\mathbf{x}^{(k+1)} &= \arg\min_{\mathbf{x}} \|\mathbf{E}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\left\|\mathbf{x} - \mathbf{z}^k\right\|_2^2 \\ &= (\mathbf{E}^{\mathbf{H}}\mathbf{E} + \lambda\mathbf{I})^{-1}(\mathbf{E}^{\mathbf{H}}\mathbf{y} + \lambda\mathbf{z}^k).\end{aligned} \tag{4}$$

Equation (4) can be solved using the Conjugate Gradient (CG) Method while Equation (3) is viewed as a CNN-based forward-pass step. MoDL is formulated as an unrolled network, where in each iteration, a CG layer is followed by a CNN-based proximal step. The unrolled reconstruction can be denoted as $\hat{\mathbf{x}} = G_w(\mathbf{y}, \mathbf{E})$, where $\mathbf{y}$, $\mathbf{E}$, and $w$correspond to the under-sampled k-space measurements, the encoding matrix, and the learnable weights of the reconstruction network, respectively. Training the unrolled model becomes supervised learning with a predefined loss function:

$$\min_{w} \sum_i \mathcal{L}(G_w(\mathbf{y}_i, \mathbf{E}_i), \mathbf{x}_i), \tag{5}$$

where $\mathbf{x}_i$ is the ith fully sampled ground truth image, and $\mathbf{y}_i$ is the retrospectively under-sampled k-space computed by applying the encoding matrix $\mathbf{E}_i$ to generate $\mathbf{y}_i = \mathbf{E}_i\mathbf{x}_i$. The loss function $\mathcal{L}(\cdot)$ can be combinations of ℓ_1, ℓ_2, SSIM, and other losses. Once trained, a new under-sampled scan denoted by $\mathbf{y}$ with the encoding operator $\mathbf{E}$ is reconstructed as:

$$\hat{\mathbf{x}} = G_w(\mathbf{y}, \mathbf{E}). \tag{6}$$

2.2 | UFLoss feature mapping network

As shown in Figure 2A, a patch-wise mapping network (UFLoss feature mapping network) is trained to map patches from image-space to a low-dimensional unit-norm feature space, aiming to capture high-level structural differences. The UFLoss network can then be used for training a DL-based reconstruction. In contrast to conventional supervised computer vision tasks, the UFLoss network is trained from fully sampled image patches in an unsupervised fashion. In other words, the training does not use any human annotation, which has been challenging to obtain in large-scale MRI datasets. The training is motivated by contrastive learning,[28] where a feature mapping function f_θ is learned such that each patch is maximally separated from other patches in a lower-dimensional hypersphere feature space.

Mathematically, we formulate our unsupervised feature mapping using the softmax criterion. Suppose we have N patches $\{\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_N\}$ cropped from the fully sampled images from the training set, with their corresponding unit-norm features $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_N\}$ with $\mathbf{v}_i = f_\theta(\mathbf{p}_i) \in \mathbb{R}^d$. For a certain patch $\mathbf{p}$ with feature $\mathbf{v} = f_\theta(\mathbf{p})$, the probability

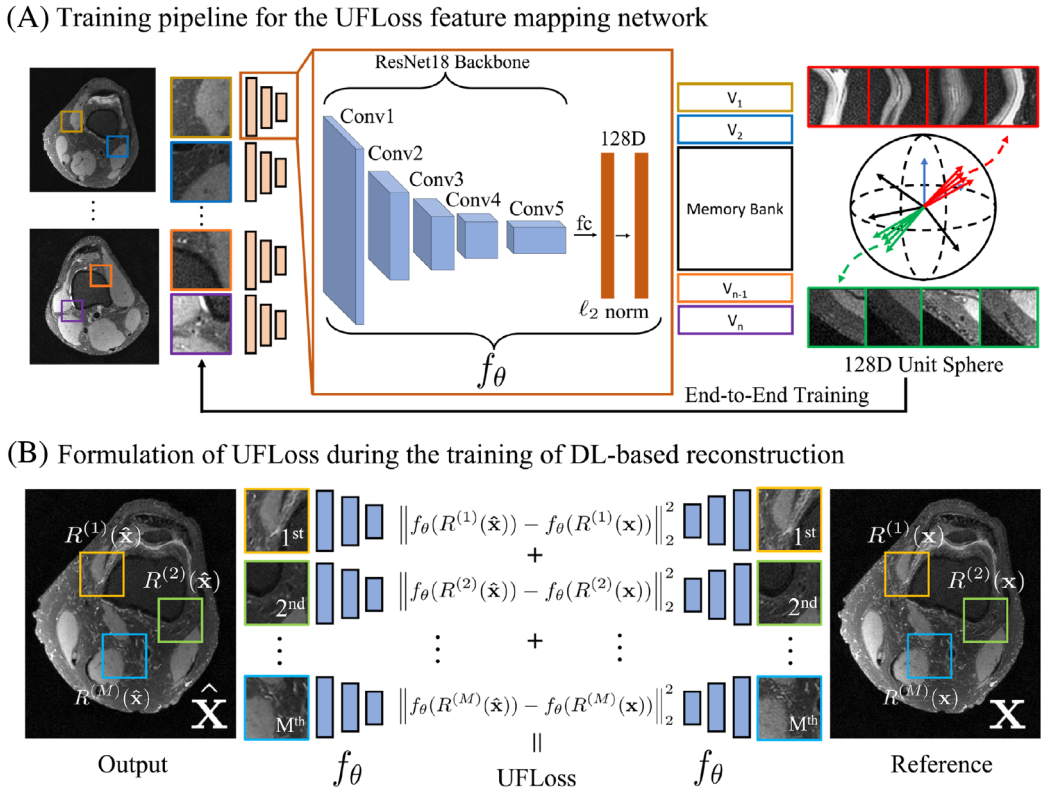


FIGURE 2 (A) Training pipeline for the Unsupervised Feature Loss (UFLoss) feature mapping network. Patches cropped from the fully sampled images are separately passed through a ResNet 18[29] backbone followed by an ℓ_2 normalization layer to map the patches to features on a low-dimensional unit sphere (128-dimension unit-norm features in this work). A memory bank is used to store the features from all the training patches to save computation when computing the softmax loss function (Equation 9). Then, end-to-end training is performed such that each patch is maximally separated from other patches in the 128D unit-norm feature space. Similar patches will naturally cluster in the low-dimensional space. (B) Detailed formulation of the proposed UFLoss during the training of the deep learning-based reconstruction. Operator R extracts a total of M patches from an image. These patches are extracted on a grid with a sliding window. Each patch from the reconstructed output and the fully sampled reference will go through a pretrained network f_θ and mapped to a low-dimensional feature space. The UFLoss corresponds to the sum of the ℓ_2 distance between the feature vectors from the output and the fully sampled reference

of it being identified as the ith patch under a linear classifier is:

$$P(i|\mathbf{v}) = \frac{\exp(\mathbf{w}_i^{\mathrm{T}}\mathbf{v})}{\sum_{j=1}^{N}\exp(\mathbf{w}_j^{\mathrm{T}}\mathbf{v})}, \tag{7}$$

where $\mathbf{w}_j$ is the weight vector of class j (or patch j), and $\mathbf{w}_j^{\mathrm{T}}\mathbf{v}$ shows how well the feature vector $\mathbf{v}$ matches the jth patch. However, the above formulation Equation (7) requires a class prototype $\mathbf{w}$ in addition to the patch feature itself, making direct comparison between patches infeasible. To address this problem, we follow the approach in Reference 28 to turn the instance-wise classification into a metric learning problem, where $\mathbf{w}_j^{\mathrm{T}}\mathbf{v}$ in Equation (7) is replaced with $\mathbf{v}_j^{\mathrm{T}}\mathbf{v}$. That is, the jth patch feature is its class prototype itself. The probability then becomes:

$$P(i|\mathbf{v}) = \frac{\exp(\mathbf{v}_i^{\mathrm{T}}\mathbf{v})/\tau}{\sum_{j=1}^{N}\exp(\mathbf{v}_j^{\mathrm{T}}\mathbf{v})/\tau}, \tag{8}$$

where τ is a temperature parameter that controls the extent of separation/concentration of the distribution in the feature space. The learning objective is set to maximize the joint probability $\Pi_{i=1}^{N}P_\theta(i|f_\theta(\mathbf{x}_i))$, which is equivalent to minimizing the negative log-likelihood over the training set:

$$J(\theta) = -\sum_{i=1}^{N}\log P(i|f_\theta(\mathbf{x_i})). \tag{9}$$

Note that in order to compute the probability $P(i|\mathbf{v})$ in Equation (8), features $\{\mathbf{v}_i\}$ from all the patches are required. Instead of exhaustively computing all the features every time, a memory bank $\mathbf{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_N\}$ is constructed to store all the feature vectors. During each training iteration, while the network parameters θ are optimized over the ith patch, the ith entry of the memory bank $\mathbf{v}_i$ is replaced by the output of the feature mapping network $\mathbf{f}_\theta(\mathbf{p}_i) \rightarrow \mathbf{v}_i$.

Once trained, the UFLoss network can be used as a perceptual loss term in other supervised reconstruction tasks, as described next.

2.3 | DL-based reconstruction with UFLoss

The UFLoss network is designed to maximally separate patches in the low-dimensional unit-sphere feature space. Perceptually similar patches are mapped to similar features.

Consider the under-sampled reconstruction using an unrolled network in Equation (5). Suppose we have the ground truth fully sampled image $\mathbf{x}_i$, and the output of the unrolled network $\hat{\mathbf{x}}_i = G_w(\mathbf{y}_i, \mathbf{E}_i)$. Since the inputs of the UFLoss network are image patches (Figure 2B), we first extract M overlapping image patches from both $\mathbf{x}_i$ and $\hat{\mathbf{x}}_i$, obtaining two patch groups: $\left\{\mathbf{p}_i^1, \mathbf{p}_i^2, ..., \mathbf{p}_i^M\right\}$ and $\left\{\hat{\mathbf{p}}_i^1, \hat{\mathbf{p}}_i^2, ..., \hat{\mathbf{p}}_i^M\right\}$. The patches are extracted on a grid with N_s pixel strides horizontally and vertically.

During each training step, random shifts between 0 to N_s pixels are applied with equal shifts to both $\mathbf{x}_i$ and $\hat{\mathbf{x}}_i$. This choice has the effect of averaging out the blocking artifacts and achieves the same performance as extracting all the patches.[4,30]

Since we use inner products to measure the distance in the hyperspherical feature space, the UFLoss can be formulated as the average of the negative inner products over all the patches. On top of that, we add a constant 1 in front of our loss function:

$$L_{\text{UFLoss}}(\mathbf{x}_i, \hat{\mathbf{x}}_i) = \frac{1}{M}\sum_j 1 - \langle f_\theta(\mathbf{p}_i^j), f_\theta(\hat{\mathbf{p}}_i^j)\rangle, \tag{10}$$

where $\langle\cdot,\cdot\rangle$ is the inner product operation between two unit-norm vectors and f_θ is the pretrained UFLoss mapping network. As both $f_\theta(\mathbf{p}_i^j)$ and $f_\theta(\hat{\mathbf{p}}_i^j)$ have unit norms, the above loss function can be also written as a mean-squared-error (MSE) in the feature space, or:

$$\begin{aligned} L_{\text{UFLoss}}(\mathbf{x}_i, \hat{\mathbf{x}}_i) &= \frac{1}{M}\sum_j 1 - \langle f_\theta(\mathbf{p}_i^j), f_\theta(\hat{\mathbf{p}}_i^j)\rangle \\ &= \frac{1}{2M}\sum_j \left\| f_\theta(\mathbf{p}_i^j)\right\|_2^2 - 2\langle f_\theta(\mathbf{p}_i^j), f_\theta(\hat{\mathbf{p}}_i^j)\rangle \\ &\quad + \left\| f_\theta(\hat{\mathbf{p}}_i^j)\right\|_2^2 \\ &= \frac{1}{2M}\sum_j \left\| f_\theta(\mathbf{p}_i^j) - f_\theta(\hat{\mathbf{p}}_i^j)\right\|_2^2. \end{aligned} \tag{11}$$

Following the per-pixel ℓ_2 loss and UFLoss mentioned above, the full objective loss function for the DL-based reconstruction can be written as:

$$\begin{aligned} L_{\text{Recon}} &= L_{\text{MSE-all}} + 2\mu L_{\text{UFLoss-all}} \\ &= \sum_i L_{\text{MSE}}(\mathbf{x}_i, \hat{\mathbf{x}}_i) + 2\mu\sum_i L_{\text{UFLoss}}(\mathbf{x}_i, \hat{\mathbf{x}}_i) \\ &= \sum_i \left\| G_w(\mathbf{y}_i, \mathbf{E}_i) - \mathbf{x}_i\right\|_2^2 \\ &\quad + \mu\sum_i \frac{1}{M}\sum_j \left\| f_\theta(\mathbf{p}_i^j) - f_\theta(\hat{\mathbf{p}}_i^j)\right\|_2^2, \end{aligned} \tag{12}$$

where μ is the weighting factor on the contribution of the UFLoss. End-to-end training is then performed on this total loss to optimize the reconstruction network G_w.

3 | METHODS

3.1 | Imaging datasets

We trained and evaluated our proposed UFLoss on both 2D and 3D fully sampled knee datasets with retrospective under-sampling. We used the fastMRI[31] high-resolution knee dataset for our 2D experiments. A total of 5700 fully sampled slices from 380 cases were split into 320 cases (6080 slices) for training, 40 cases (640 slices) for validation, and 20 cases (320 slices) for testing. Image normalization was performed such that the 95% percentile of the intensity values was scaled to 1 for each subject. The training dataset includes data from two different contrasts: proton-density with (PDFS) and without (PD) fat suppression. Relevant imaging parameters are described in the fastMRI[31] paper. For the unrolled reconstruction task, retrospective under-sampling was performed by applying a one-dimensional five times accelerated random under-sampling mask (20% sampling rate) with an 8% fully sampled k-space center. Sensitivity maps were computed using ESPIRiT[32] using Berkeley Advanced Reconstruction Toolbox (BART)[33] with a 24 × 24 calibration region.

We conducted our 3D experiments on 20 fully sampled 3D knee scans (available at mridata.org)[34] with retrospective under-sampling. The k-space data was acquired on a 3T GE Discovery MR 750, with an 8-channel HD knee coil. Scan parameters include a matrix size of 320 × 320 × 256, and TE/TR of 25 ms/1550 ms. A total of 5120 slices from 16 cases were used for training, 640 slices from two cases were used for validation, and 640 slices from the remaining two cases were used for testing. We normalized each 3D volume with respect to the 95% percentile of the intensity values for the entire volume. Each 3D volume was under-sampled with a different 8× Poisson-disk sampling mask (12.5% sampling rate) with a 24 × 24 calibration region. Sensitivity maps were computed using ESPIRiT[32] with a 24× 24 calibration region using BART.[33] Note that

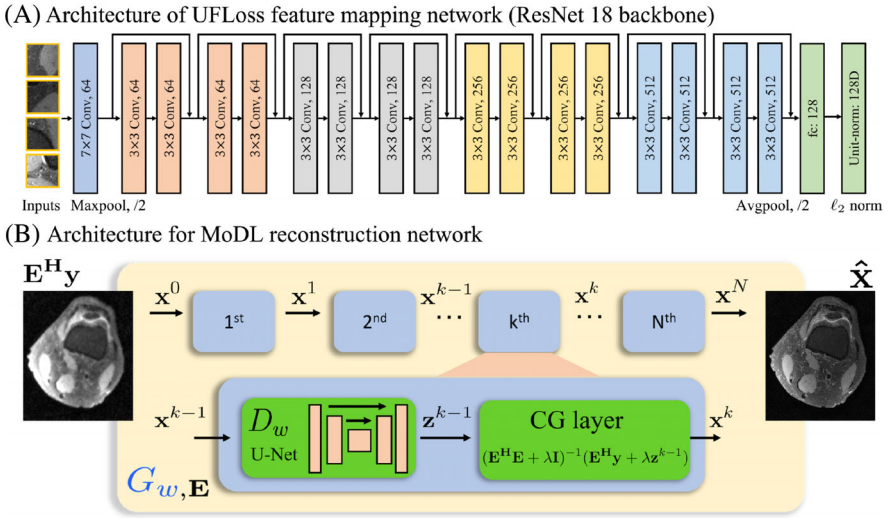


FIGURE 3 (A) The Unsupervised Feature Loss (UFLoss) feature mapping network is based on a ResNet 18 network structure[29] and followed by an ℓ_2 normalization layer to map the input patches to the 128D unit-norm feature space. (B) Architecture of the Model-based Deep Learning[11] reconstruction network. A data consistency Conjugate Gradient Descent module is inserted after a Convolutional Neural Network-based denoiser D_w. D_w follows the structure of U-Net[36] with two input channels that represent the real and imaginary parts of the complex-valued image data

we train both the UFLoss network and the DL-based reconstructions on the entire training set and use fully sampled coil-combined images as ground truth.

3.2 | Implementation of UFLoss feature mapping network

In all our networks, the input coil-combined complex-valued MR images/patches $\mathbf{x} \in \mathbb{C}^{\mathbf{N}}$ are converted into a two-channel representation $\mathbf{x} \in \mathbb{R}^{\mathbf{2N}}$, where the real and imaginary components are treated as two individual channels. As illustrated in Figure 3A, we implemented the UFLoss network using a ResNet 18[29] backbone followed by a ℓ_2 normalization layer to map the input patches to 128 dimension unit-norm features. Based on the field of view and resolution difference, the input patch sizes of the 2D fastMRI knee dataset and 3D knee dataset were set to 60×60 and 40×40 pixels, respectively. The UFLoss networks for the 2D fastMRI and 3D knee datasets were trained separately due to the differences in image content. Eighty patches were extracted from each slice at random locations, resulting in 409600 patches used to train the UFLoss network. Other hyperparameters include temperature τ of 1 (Equation 8), batch size 16, the number of epochs of 100, and the learning rate of 1e−4 with Adam[35] optimizer.

3.3 | Implementation of DL-based reconstruction with UFLoss

For the unrolled reconstruction network architecture, we used the structure from the MoDL paper,[11] where a CG block was inserted after a CNN-based denoiser, and unrolled with a fixed number of iterations. In this work, we used five unrolls and six CG steps. As shown in Figure 3B, a U-Net[36] architecture was adopted for the CNN-based denoiser D_w.

The training of MoDL was performed by minimizing the proposed loss function L_{Recon} (Equation 12) over the training set for 50 epochs, with an empirical weighting parameter $\mu = 1.5$, and Adam[35] optimizer with a learning rate of 1e−4.

To compute the UFLoss, patches are extracted on a grid across the image with five-pixel strides in both vertical and horizontal directions. At each training step, both output and reference images are randomly shifted from 0 to 5 pixels in the vertical and horizontal directions to eliminate blocking artifacts. In this work, we chose the weighting parameter to balance the values of $L_{\text{MSE-all}}$ and $L_{\text{UFLoss-all}}$ so that they are on par after the training converges. During inference, a zero-filled reconstruction is passed through the MoDL reconstruction network. Note that training with UFLoss does not change the network architecture, so the inference time remains the same as MoDL with pure ℓ_2 loss.

All the proposed algorithms were implemented using Pytorch 1.2,[37] and were run on 12 GB Nvidia Titan Xp graphics processing units (GPUs).

3.4 | Evaluation of the proposed UFLoss

3.4.1 | UFLoss as valid loss function

To evaluate whether UFLoss is also a valid loss function for comparing two images at the intensity level, we study how the UFLoss changes with different sizes of perturbations in two representative types:

1. Additive white Gaussian noise.

A perturbed image $\mathbf{x_p}$ is generated from the original image $\mathbf{x_o}$ by adding different levels of additive Gaussian noise $\mathbf{n_\sigma}$:

$$\mathbf{x_p} = (1-\beta)\mathbf{x_o} + \beta\mathbf{n_\sigma}, \tag{13}$$

where β is the noise level parameter in the range of 0 − 10%, and noise $\mathbf{n_\sigma}$ follows normal distribution: $\mathbf{n_\sigma} \sim \mathcal{N}(0,1)$. We study how $L_{\text{UFLoss}}(\mathbf{x_o}, \mathbf{x_p})$ changes as β increases.

2. Image blurring.

A perturbed low-resolution image $\mathbf{x_p}$ is generated by cropping and zero-padding the k-space of the original image $\mathbf{x_o}$. The k-space cropping rate $\mathbf{R}$ ranges from 1-4. $\mathbf{R} = 4$ indicates that only 25% of k-space samples in both horizontal and vertical dimensions are kept. A higher $\mathbf{R}$ corresponds to more blurring and a coarser resolution. We study how $L_{\text{UFloss}}(\mathbf{x_o}, \mathbf{x_p})$ varies with different $\mathbf{R}$'s.

In addition, we evaluate whether, by minimizing the objective UFLoss between the original and perturbed images $L_{\text{UFloss}}(\mathbf{x_o}, \mathbf{x_p})$, we are able to guide the perturbed version toward the original version without falling into local minima. The starting perturbed image $\mathbf{x_{p-0}}$ is generated by image blurring where $\mathbf{R} = 4$. We update it per gradient descent with respect to $L_{\text{UFloss}}(\mathbf{x_o}, \mathbf{x_{p-k}})$ in an iterative fashion:

$$\mathbf{x_{p-k+1}} = \mathbf{x_{p-k}} - \alpha \frac{\partial L_{\text{UFLoss}}(\mathbf{x_o}, \mathbf{x_{p-k}})}{\partial \mathbf{x_{p-k}}}, \tag{14}$$

where $\mathbf{x_{p-k}}$ is the perturbed image after $\mathbf{k}$ steps of gradient descent.

3.4.2 | Perceptual similarity

In order to better interpret and understand the perceptual features learned for the UFLoss, we performed a patch retrieval experiment to evaluate and show patch pairs with high and low UFLoss feature similarities. First, we constructed a feature database (memory bank) by running all training patches through the pretrained UFLoss network. Then, given an input patch from the testing set, we passed it through the network and queried its neighbors from the training patches based on their distances (inner products) in the feature space. We picked and visualized patches of the highest feature inner products with the input patch and also counter-examples with relatively low inner products.

To further evaluate the UFLoss sensitivity and perceptual similarity for different anatomies and contrasts, we constructed correlation maps by computing the feature correlation (inner product) between a source patch and all patches in different images and visualized them as heatmaps. This experiment helps us better understand how anatomy and structure similarities relate to UFLoss feature similarities.

Specifically, we first extracted a source patch from a source image. Then, we computed the feature correlations between the source patch and all patches on a grid from (1) the same source image; (2) the target image with the same contrast but from a different subject; and (3) the target image with different contrast and also from a different subject. Patches closer to the source patch in the feature space correspond to higher inner products. We evaluated this experiment on both PDFS and PD scans. For comparisons, we also conducted the same experiments for the SSIM feature, where we computed the SSIM score between the source patch and all patches from different images.

3.4.3 | Unrolled reconstructions with UFLoss

To quantitatively evaluate our proposed UFLoss on under-sampled MRI reconstruction, we implemented both PICS[4] and MoDL.[11] In the unrolled reconstruction experiments, MoDL with our proposed UFLoss was compared with PICS and with MoDL using only per-pixel ℓ_2 loss. The PICS method was implemented using the BART Toolbox[33] with wavelets as the sparse transform. In order to further demonstrate the performance of our UFLoss, MoDL with ℓ_2 + perceptual VGG loss[21] was also included in our comparisons. To compute the perceptual VGG loss, both the real and imaginary parts are scaled from 0 to 255 and duplicated three times to serve as the inputs of the pretrained VGG network. The VGG network is pretrained on ImageNet classification. VGG loss corresponds to the ℓ_2 distance between the relu_22 features from the output and the ground truth image.

For all the experiments, reconstruction performance was evaluated using different quantitative metrics, which reflect different aspects of image quality. The normalized root mean squared error (NRMSE) was used to measure the overall pixel-wise errors. SSIM[13] was used to assess the local image similarity with respect to the fully sampled reference. At the same time, we also computed our proposed UFLoss between the reconstructed images and the fully sampled references.

4 | RESULTS

4.1 | UFLoss as a valid loss function

Figure 4 indicates that our proposed UFLoss could be used as a valid loss function by itself. As shown in Figure 4A,

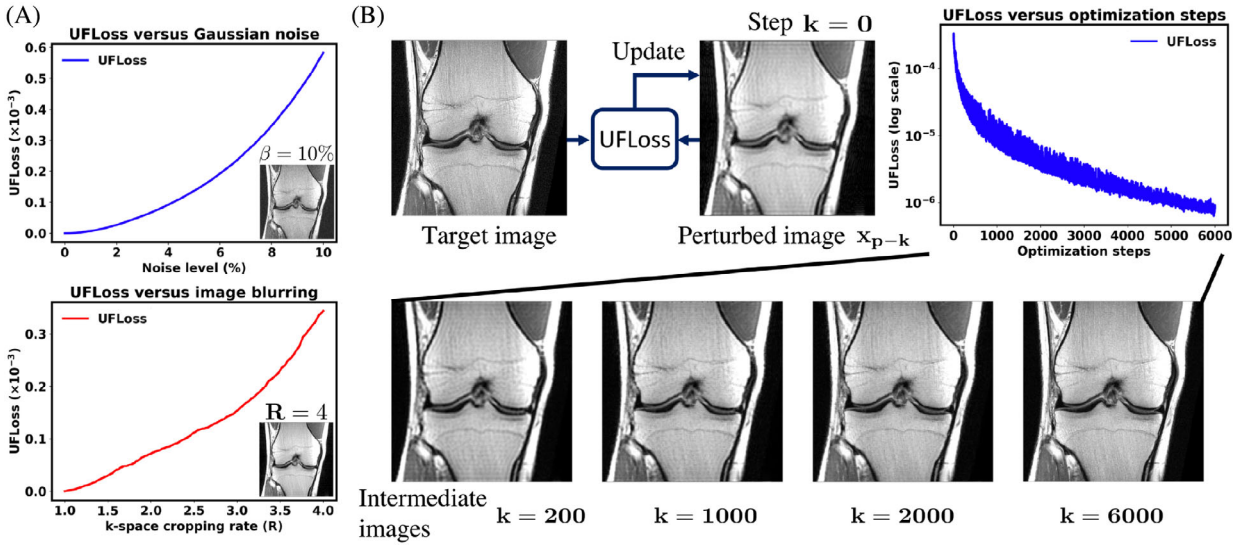


FIGURE 4 Unsupervised Feature Loss (UFLoss) can be used as a valid loss function. (A) Evaluation of UFLoss with different levels of perturbations. **Upper**: additional Gaussian noise, **Lower**: image blurring through k-space cropping. UFLoss evolution curves indicate that UFLoss increases in a convex way with respect to more Gaussian noise and increases in a near-convex way with respect to more blurring. (B) Evaluation of UFLoss in guiding a blurred image $\mathbf{x}_{\mathbf{p-0}}$ to the target high resolution image. Gradient descent is performed on $\mathbf{x}_{\mathbf{p-k}}$ to reduce the UFLoss with respect to the target image in an iterative way. Intermediate images show that UFLoss is able to gradually guide the blurred image to the target without falling into any local minimum

UFLoss between the perturbed and original clean images increases in a convex way with respect to more Gaussian noise and increases in a near-convex way with respect to more blurring. Even though the UFLoss feature mapping network is not specifically trained for any such perturbations, it learns low-level perceptual similarities between images, where a larger intensity perturbation corresponds to a larger UFLoss. On the other hand, Figure 4B indicates that by minimizing the UFLoss between the perturbed and target images, we are able to successfully restore the blurred image toward the clean one without falling into any local minimum. Intermediate deblurred image samples are shown in the figure along with the UFLoss evolution curve.

4.2 | Perceptual similarity

Figure 5A shows the feature similarity results using the UFLoss feature. The feature space inner products between the input patch and the retrieved patches are shown as different colors of the borders. As seen in the figure, patches with similar perceptual structures (e.g., edges, bone structures) are mapped closer to each other in the feature space.

Figure 5B (PDFS) and Supporting Information Figure S1 (PD) show the feature correlation maps (UFLoss and SSIM) between different patches. Two source patches, indicated with green and blue edges, were chosen from each source image in the left column. The heatmaps under to each image, with corresponding green and blue edges, show the corresponding maps for each source patch from the source image. For the UFLoss results, we only show the positive inner products for visualization purposes, while in principle, the inner products range from −1 to 1. As shown in the UFLoss feature correlation maps, patches containing meniscus from both the same contrast and different contrast show high correlations with the input patch of the meniscus (blue border) while, on the other hand, patches from other anatomy show low correlation with it. These UFLoss feature correlation maps indicate that our unsupervised feature mapping is able to capture the perceptual structure similarities across different subjects and across different contrasts. In contrast, SSIM feature correlation maps do not successfully capture perceptual similarities across anatomies and contrasts (e.g., meniscus). More specifically, as shown in Supporting Information Figure S2, patch with the highest UFLoss feature correlation (top) shows very similar anatomical textures of the meniscus compared to the source patch. At the same time, because SSIM focuses more on the local signal statistics instead of high-level perceptual similarity, the patch with the highest SSIM (bottom) has totally different textures from a different anatomical region.

4.3 | Unrolled reconstructions with UFLoss

Figure 6 shows reconstruction comparisons between different methods (PICS, MoDL, MoDL with VGG,

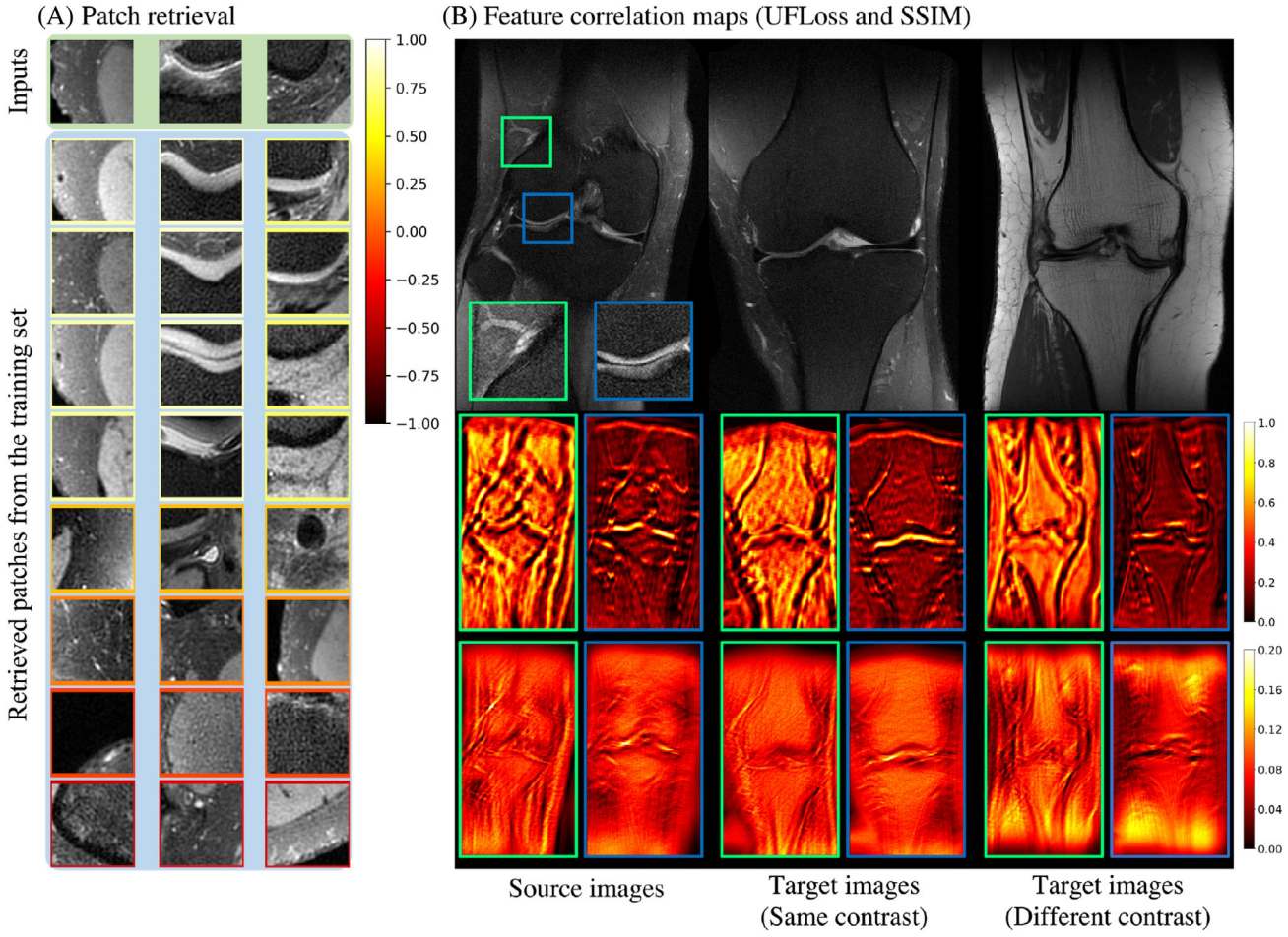


FIGURE 5 UFLoss is able to capture perceptual similarities across anatomies and contrasts. (A) Feature clustering results using UFLoss feature mapping where, given an input patch, neighbor patches from the training set can be queried based on their feature space distance. The top four patches are the closest neighbors with the input patch and have the highest inner products. At the same time, we also show four counterexamples with relatively low inner products with the input patch. The feature space inner products between the input patch and the retrieved patches are shown as different colors of the borders. The color bar on the right indicates that a brighter border corresponds to a higher correlation while a darker border corresponds to a lower correlation. (B) Feature correlations between different patches. The heat maps under a certain image show the feature correlations (feature space inner products for Unsupervised Feature Loss [UFLoss]) between all the patches from the image and the reference patches from the source image (first column). The heat maps with green/blue borders correspond to different source patches whose borders have the same colors. The correlation results for proton density with fat suppression contrast using UFLoss and structural similarity index features are shown in the top and bottom rows, respectively

MoDL with UFLoss) for a representative 3D knee scan with under-sampling rate of $R = 8$. Quantitative metrics (NRMSE, SSIM) are shown under the images. As indicated in the zoomed images and error maps, MoDL with UFLoss shows finer structural details, sharper edges, and higher perceptual agreement with the fully sampled reference images compared to the other reconstruction methods. Without our UFLoss, pure ℓ_2 loss at this under-sampling rate leads to blurring and perceptual quality degradation. MoDL with the VGG perceptual loss[21] shows higher perceptual quality compared with MoDL, but generates unintended checkerboard structured artifacts, which is consistent with findings in References 38,39. In terms of the training time and GPU memory cost for 3D reconstruction experiments, under the same setup, MoDL with UFLoss takes 92 min for a single epoch using 8.1 GB GPU memory, while MoDL with ℓ_2 loss takes 58 minutes using 5.5 GB and MoDL with perceptual VGG loss takes 61 min using 5.7 GB. In inference time, it takes around 25 ms and 0.9 GB for all methods.

Figure 7 shows the comparison of different reconstruction methods for a representative 2D PD slice from the fastMRI dataset.[31] The retrospective 2D under-sampling rate is 5, where around 20% of the k-space data is sampled. At this acceleration rate, PICS failed to effectively recover the fine bone structures, and MoDL with ℓ_2 loss alone also suffers blurring artifacts. In contrast, MoDL with UFLoss demonstrates more realistic reconstruction

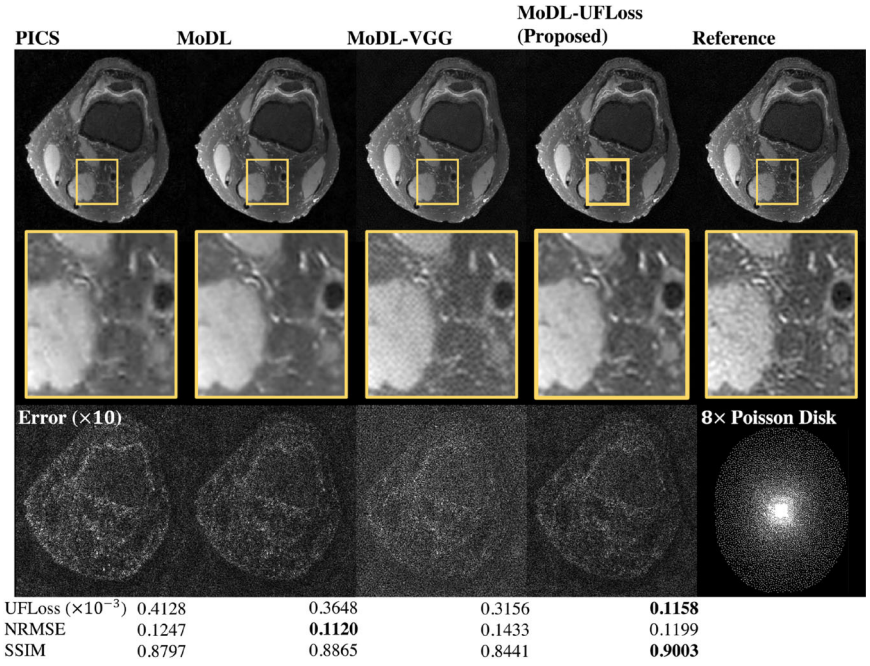


FIGURE 6 Representative three-dimensional knee reconstruction results from different methods. A fully sampled scan is retrospectively under-sampled with a Poisson under-sampling mask by a factor of 8. From left to right are reconstructions by: combined PI and CS (PICS), model-based Deep Learning (MoDL) with ℓ_2 loss, MoDL with ℓ_2+perceptual VGG loss, and MoDL with ℓ_2+our proposed unsupervised Feature Loss (UFLoss). Normalized root mean squared error, structural similarity index, and UFLoss for each method are computed with respect to the fully sampled reference and shown under the image for reference. As shown in the zoomed images and error maps, our proposed MoDL with UFLoss showed sharper edges and more detailed structures with high perceptual similarity compared to the reference image

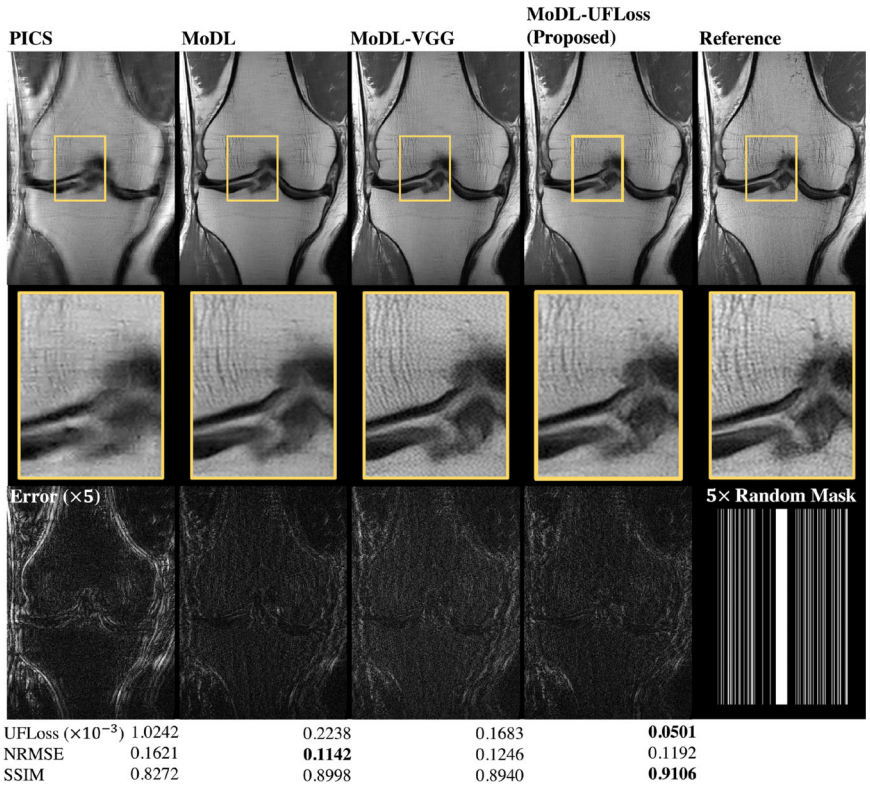


FIGURE 7 Representative examples of two-dimensional proton-density knee reconstruction results using different methods. A fully sampled slice is retrospectively randomly under-sampled by a factor of 5. From left to right are reconstructions by combined PI and CS, MoDL with ℓ_2 loss, model-based Deep Learning (MoDL) with perceptual VGG loss, and MoDL with our proposed unsupervised Feature Loss (UFLoss). Normalized root mean squared error, structural similarity index, and UFLoss for each method are shown below the figure for references. As shown in the zoom-in views and error maps, our proposed MoDL with UFLoss can provide more realistic and natural-looking textures, while MoDL with ℓ_2 loss alone tends to blur out some high-frequency textures

performance with more detailed texture everywhere, including the bone.

Figure 8 shows the reconstruction comparisons for a representative 2D PDFS slice from the fastMRI dataset.[31] Quantitative comparisons are shown at the bottom of the figure. Due to the suppression of the fat signal, the signal-to-noise-ratio of the data is relatively low, where high-frequency features can be mixed up with the noise. The zoomed-in views and the corresponding error maps indicate that PICS results in a high level of artifacts. Meanwhile, MoDL with ℓ_2 loss alone misses fine detailed structures. Similar to the analysis above, MoDL with the VGG feature loss is capable of recovering subtle structures but generates unintended structured artifacts. In contrast,

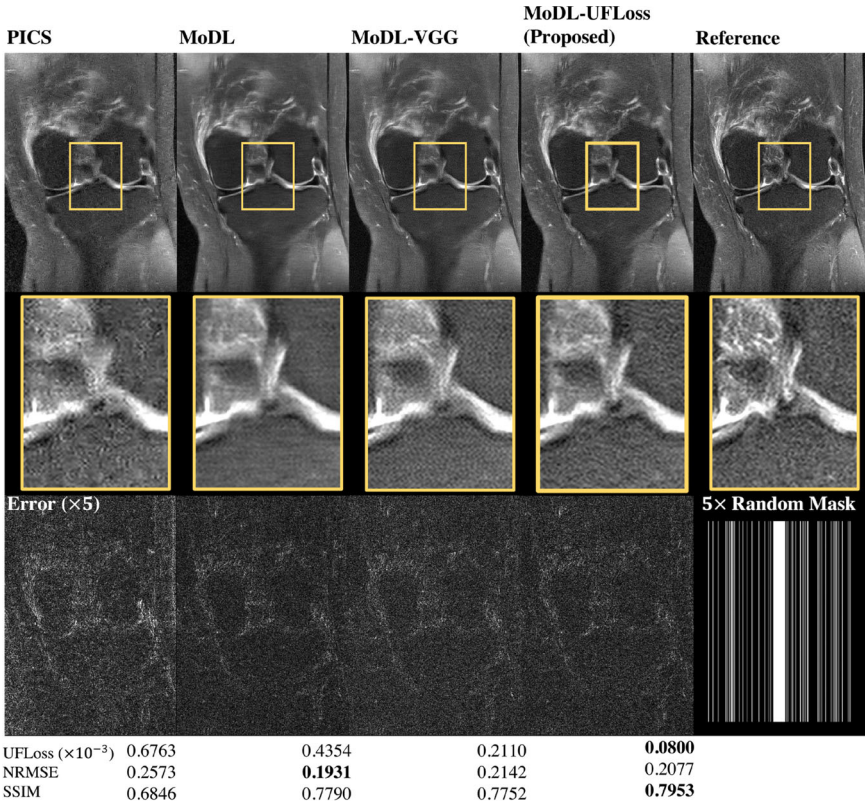


FIGURE 8 Representative examples of two-dimensional proton density with fat suppression (PDFS) knee reconstruction results using different methods at under-sampling rate $R = 5$. Normalized root mean squared error, structural similarity index (SSIM), and unsupervised Feature Loss (UFLoss) for each method are shown in the figure. Quantitative metrics indicate that model-based Deep Learning (MoDL) with UFLoss has the highest SSIM and the lowest UFLoss, as well as the highest perceptual quality of the reconstructed image. Meanwhile, as shown in the zoom-in images and error maps, our proposed MoDL with UFLoss reconstruction looks more natural with a more faithful contrast than other methods

MoDL with UFLoss can effectively recover the detailed texture and have the most realistic reconstructions. In terms of the training time and GPU memory cost for 2D fastMRI experiments, under the same setup, MoDL with UFLoss takes 143 min for a single epoch using 11.9 GB GPU memory, while MoDL with ℓ_2 loss takes 104 min using 7.3 GB and MoDL with perceptual VGG loss takes 108 min using 7.5 GB. In inference time, it takes around 40 ms and 1.4 GB for all methods.

So far, for all of our experiments, we used a fixed UFLoss weighting factor ($\mu = 1.5$) for Equation (12). Supporting Information Figure S3 shows two representative reconstruction results with different UFLoss weighting factors during the training. We can clearly see that neither pure ℓ_2 loss nor pure UFLoss achieves the best image quality. By combining these two terms, our model is able to take advantage of both the per-pixel intensity information and patch-level perceptual similarities.

Figure 9 shows the quantitative metric (NRMSE, SSIM, UFLoss) comparisons for the 2D unrolled reconstruction experiments. For both (a) PD and (b) PDFS experiments, 10 representative testing scans with 15 slices each are used to calculate the quantitative metrics. As indicated in the figure, for both contrasts, MoDL with UFLoss outperforms both PICS and MoDL with ℓ_2 loss in terms of SSIM and UFLoss and can achieve comparable performance in terms of NRMSE.

5 | DISCUSSION

In this work, we presented a novel patch-based perceptual loss function, which we call UFLoss. UFLoss corresponds to the ℓ_2 distance in a low-dimensional feature space. Feature vectors are mapped from image patches through a pretrained mapping network. The mapping network aims to maximally separate all the patches in the feature space, where similar patches become closer to each other, capturing high-level perceptual similarities. As indicated in Figure 5, unlike ℓ_2 distance, which focuses on the pixel-wise values, our proposed UFLoss agrees better with human visual judgment, where similar-looking patches have lower UFLoss in the feature space. By incorporating UFLoss into the training of DL-based reconstructions, we are able to recover finer textures, smaller features, and sharper edges with higher overall image quality compared to conventional per-pixel losses. By leveraging a memory bank to store all the features, the training of our mapping network becomes feasible for a large dataset: The UFLoss network training required less than 500 MB GPU memory and was easily trained within 2 hours. In terms of computation costs, our UFLoss imposes about 50% increase in training time and memory requirements during training. However, in inference time, the UFLoss has no penalty at all on the reconstruction time as well as the memory requirements.

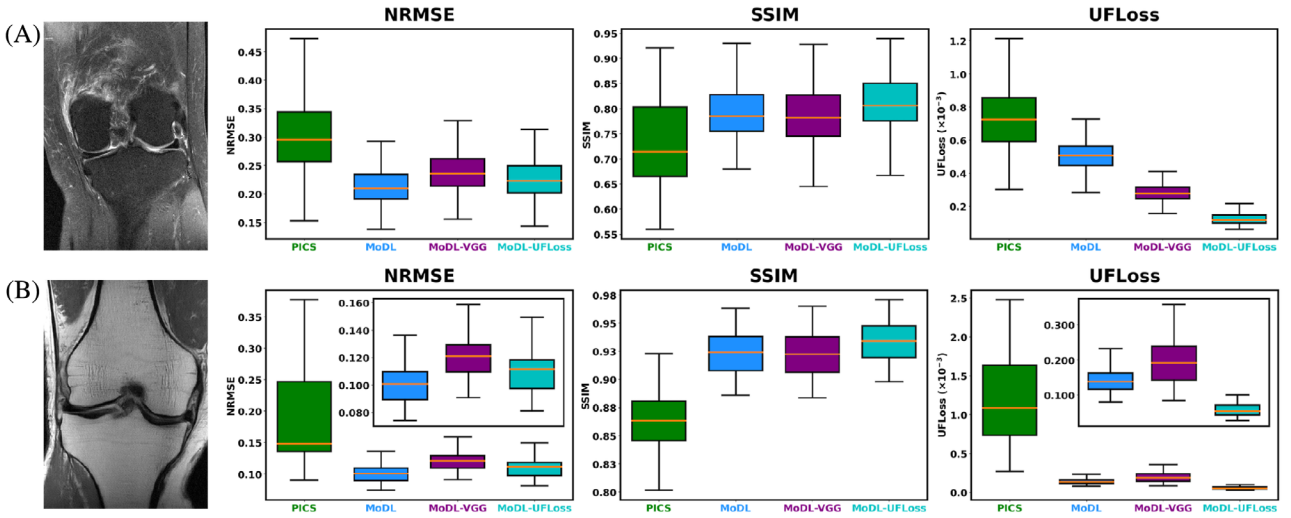
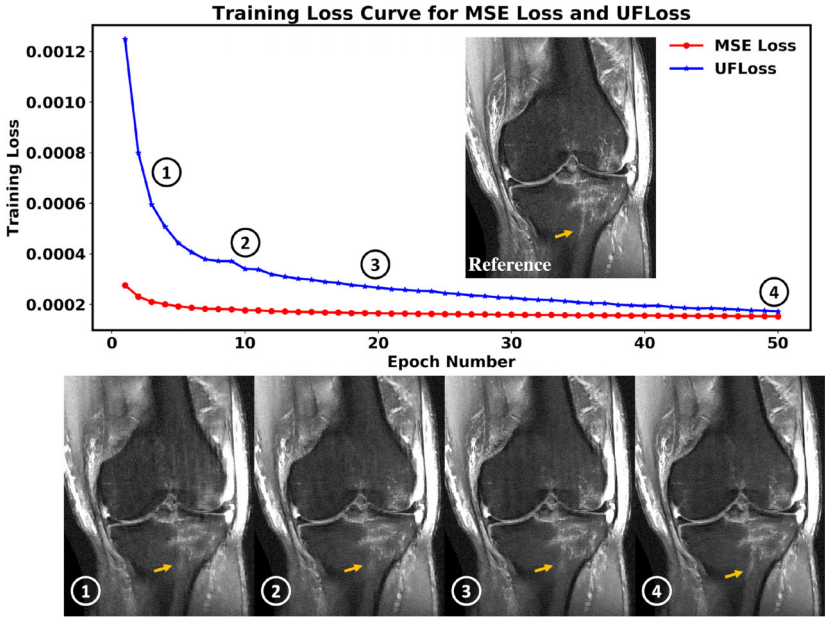


FIGURE 9 Model-based Deep Learning (MoDL) with Unsupervised Feature Loss (UFLoss) shows competitive results in the metric comparisons for both (A) proton density (PD) and (B) PD with fat suppression (PDFS) experiments. Two representative fully sampled scans (10 PD and 10 PDFS) with 15 slices each are randomly under-sampled by a factor of 5 and reconstructed using combined PI and CS, MoDL, MoDL with perceptual VGG loss, and MoDL with UFLoss. normalized root mean squared error (NRMSE), structural similarity index (SSIM), and UFLoss are calculated with respect to fully sampled reference images and shown in the plot. We use zoomed-in plots to show more clear comparisons for some sub-plots. For both contrasts, MoDL with UFLoss outperforms both PICS and MoDL with ℓ_2 loss in terms of SSIM and UFLoss and can achieve comparable performance in terms of NRMSE

As we mentioned before, another important class of feature losses for DL-based reconstruction is adversarial loss or GAN loss.[15] Adversarial losses have shown great success in capturing perceptual properties of ground-truth images and could be used to improve the reconstruction quality. However, due to the min-max loss function, the convergence of GANs is generally underdetermined, and it is difficult to determine the stop criterion for GANs' training.[40] In contrast, the convergence and stop criterion of training with UFLoss is clear and straightforward, simply when the loss function (pixel loss + UFLoss) converges. Another important distinction is that GAN loss is an instance-to-set loss, which means that so long as the reconstruction is similar to any of those ground-truth training images, the loss would be small, which is undesirable for reconstruction.[19,20,41,42] In comparison, UFLoss is an instance-wise discriminative loss, comparing the reconstruction to the specific ground truth image in the feature space, which provides clear guidance and is more constraining during the training.

In this study, UFLoss can be viewed as a separate module and be easily incorporated into other learning frameworks. The performance of UFLoss was demonstrated for accelerating 2D and 3D knee imaging by comparing the reconstruction results with respect to fully sampled references. The in vivo results show that the addition of UFLoss during the network's training allows realistic texture recovery and improves overall image quality compared to a reconstruction network trained without UFLoss. Our UFLoss network trained on specific anatomy and contrast may yield suboptimal results when applied to a different contrast/anatomy. Therefore, in the ideal case, one may want to use different networks for different types of images. Fortunately, the UFLoss can be trained on the same ground truth images that are used to train reconstruction networks, therefore it does not require additional datasets to do so. Finally, the training of a UFLoss network takes less than 2 h to train, so the overhead is negligible.

Another interesting finding of the UFLoss comes from how the training losses evolve, as shown in Figure 10. The total loss consists of two different components, the per-pixel ℓ_2 MSELoss and our proposed UFLoss, which are shown in the top subfigure as red and blue curves, respectively. The bottom subfigure shows the testing reconstruction results at different epochs. As indicated from the curve, the MSELoss remains almost constant after ten epochs, while our proposed UFLoss still decreases continuously. Inspecting the reconstructed images at different training epochs, we can see that the image quality continues to improve with the further reduction of the UFLoss. At the same time, the quantitative metrics indicate that those reconstructed images have very similar NRMSE compared with the fully sampled reference but a much more significant difference in their UFLoss values. A low UFLoss value corresponds to better image quality. These results indicate that using the ℓ_2 MSE loss alone is not optimal. Therefore, the UFLoss can be potentially used as a better perceptual comparison criterion and help further improve the reconstruction quality.



	1	2	3	4
UFLoss ($\times 10^{-3}$)	0.1788	0.1269	0.1005	0.0885
NRMSE	0.2160	0.2069	0.2012	0.1989
Epochs	4	10	20	50

FIGURE 10 Training loss curves for the l_2 mean squared error (MSE) loss and our proposed unsupervised Feature Loss (UFLoss). A two-dimensional fully sampled slice is randomly under-sampled by a factor of 5 and reconstructed at different training epochs. Normalized root mean squared error and UFLoss are shown as quantitative metrics under each reconstructed image. Yellow arrows point at the same representative textures at different reconstructions. UFLoss continues improving the reconstructed image quality after l_2 MSE loss converged

Limitations of this study include: (1) The training of DL-based reconstructions with UFLoss is time-consuming (around 1.5×) and memory-inefficient (around 1.5×) due to the extraction and feed-forwarding of a large number of patches within a single step. This can be potentially improved by using fully convolutional image-scale networks, GPU parallel computing, and efficient memory-time trade-off.[43] (2) In this work, we have not thoroughly investigated the sensitivity of different hyperparameters (e.g., patch size, temperature parameter, UFLoss network depth) to the training and final reconstructions. Supporting Information Figure S3 demonstrates how UFLoss weighting parameter contributes to the reconstruction results. A more thorough parameter search and analysis will be explored in the future. (3) Even though empirical evidence for both 2D and 3D knee results has demonstrated that UFLoss can effectively encourage finer texture and sharper edges, we have not investigated the theoretical performance guarantee of UFLoss on enhancing the texture sharpness and image quality in this paper; however, our observation is supported by other perceptual loss methods in the literature.[21,22]

6 | CONCLUSION

In summary, a novel patch-based feature loss, UFLoss, is proposed, and it can be easily incorporated into the training of any existing DL-based reconstruction frameworks without any modification to the model architecture. UFLoss is based on an unsupervised pretrained feature mapping network without any external supervision. With the addition of our proposed UFLoss, we are able to reconstruct high fidelity images with sharper edges, more faithful contrasts, and better image quality overall.

ACKNOWLEDGEMENTS

The authors thank Anja Brau, Sangtae Ahn, Graeme C McKinnon, Marc Lebel, Xucheng Zhu, Gopal Nataraj and Efrat Shimron for their helpful suggestions, and Efrat Shimron for her help with the paper editing. We also acknowledge support from NIH grants R01EB026136, R01HL136965, R01EB009690, U01EB029427 and GE Healthcare.

CONFLICT OF INTEREST

Our group receives research support from GE Healthcare. Uri Wollner and Rafi Brada are employees of GE Global Research.

DATA AVAILABILITY STATEMENT

In the spirit of reproducible research, our source code can be found at https://github.com/mikgroup/UFLoss to reproduce most of the results in this paper.

ORCID

Ke Wang https://orcid.org/0000-0001-5951-1727

TWITTER

Ke Wang @KewangKe

REFERENCES

1. Sodickson DK, Manning WJ. Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. *Magn Reson Med*. 1997;38:591-603.

2. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magn Reson Med Offic J Int Soc Magn Res Med*. 1999;42:952-962.
3. Griswold MA, Jakob PM, Heidemann RM, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med Offic J Int Soc Magn Res Med*. 2002;47:1202-1210.
4. Lustig M, Donoho D, Pauly JM. Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn Reson Med Offic J Int Soc Magn Res Med*. 2007;58:1182-1195.
5. Ravishankar S, Bresler Y. MR image reconstruction from highly undersampled K-space data by dictionary learning. *IEEE Trans Med Imaging*. 2010;30:1028-1041.
6. Hammernik K, Klatzer T, Kobler E, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med*. 2018;79:3055-3071.
7. Chen F, Taviani V, Malkiel I, et al. Variable-density single-shot fast spin-echo MRI with deep learning reconstruction by using variational networks. *Radiology*. 2018;289:366-373.
8. Mardani M, Gong E, Cheng JY, et al. Deep generative adversarial neural networks for compressive sensing MRI. *IEEE Trans Med Imaging*. 2018;38:167-179.
9. Quan TM, Nguyen-Duc T, Jeong WK. Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. *IEEE Trans Med Imaging*. 2018;37:1488-1497.
10. Schlemper J, Caballero J, Hajnal JV, Price AN, Rueckert D. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans Med Imaging*. 2017;37:491-503.
11. Aggarwal HK, Mani MP, Jacob M. MoDL: model-based deep learning architecture for inverse problems. *IEEE Trans Med Imaging*. 2018;38:394-405.
12. Tamir JI, Yu SX, Lustig M. Unsupervised deep basis pursuit: learning inverse problems without ground-truth data; 2019. arXiv preprint arXiv:1910.13110.
13. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similar ity. *IEEE Trans Image Process*. 2004;13:600-612.
14. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: a review. *Med Image Anal*. 2019;58:101552.
15. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. *Adv Neural Inf Process Syst*. 2014;27:2672-2680.
16. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017; IEEE.
17. Mirza M, Osindero S. Conditional generative adversarial nets; 2014. arXiv preprint arXiv:1411.1784.
18. Liu F, Samsonov A, Chen L, Kijowski R, Feng L. SANTIS: sampling-augmented neural network with incoherent structure for MR image reconstruction. *Magn Reson Med*. 2019;82:1890-1904.
19. Sandino CM, Cheng JY, Chen F, Mardani M, Pauly JM, Vasanawala SS. Compressed sensing: from research to clinical practice with deep neural networks: shortening scan times for magnetic resonance imaging. *IEEE Signal Process Mag*. 2020;37:117-127.
20. Muckley MJ, Riemenschneider B, Radmanesh A, et al. Results of the 2020 fastMRI challenge for machine learning MR image reconstruction. *IEEE Trans Med Imaging*. 2021;40:2306-2317.
21. Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. Proceedings of the European Conference on Computer Vision; 2016:694-711; Springer, New York, NY.
22. Zhang R, Isola P, Efros AA, Shechtman E, Wang O. The unreasonable effectiveness of deep features as a perceptual metric. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018:586-595; IEEE.
23. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition; 2014. arXiv preprint arXiv:1409.1556.
24. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: a large-scale hierarchical image database. Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2009:248-255; IEEE.
25. Wang K, Tamir JI, Stella XY, Lustig M. High-fidelity reconstruction with instance-wise discriminative feature matching loss. *Proc Int SocMagn Reson Med*. 2020;0994-0994.
26. Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J Imaging Sci*. 2009;2:183-202.
27. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found Trends Mach Learn*. 2011;3:1-122.
28. Wu Z, Xiong Y, Yu SX, Lin D. Unsupervised feature learning via non-parametric instance discrimination. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018:3733-3742; IEEE.
29. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016:770-778.
30. Tamir JI, Uecker M, Chen W, et al. T2 shuffling: sharp, multicontrast, volumetric fast spin-echo imaging. *Magn Reson Med*. 2017;77:180-195.
31. Zbontar J, Knoll F, Sriram A, et al. fastMRI: an open dataset and benchmarks for accelerated MRI; 2018. arXiv preprint arXiv:1811.08839.
32. Uecker M, Lai P, Murphy MJ, et al. ESPIRiT-an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. *Magn Reson Med*. 2014;71:990-1001.
33. Uecker M, Ong F, Tamir JI, et al. Berkeley advanced reconstruction toolbox. *Proc Intl Soc Mag Reson Med*. 2015;23:2468-2468.
34. Sawyer AM, Lustig M, Alley M, et al. Creation of fully sampled MR data repository for compressed sensing of the knee.
35. Kingma DP, Ba J. Adam: a method for stochastic optimization; 2014. arXiv preprint arXiv:1412.6980.
36. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention; 2015:234-241; Springer, New York, NY.
37. Paszke A, Gross S, Chintala S, et al. Automatic differentiation in pytorch; 2017. https://openreview.net/pdf?id=BJJsrmfCZ
38. Sugawara Y, Shiota S, Kiya H. Super-resolution using convolutional neural networks without any checkerboard artifacts. Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP); 2018:66-70; IEEE.
39. Odena A, Dumoulin V, Olah C. Deconvolution and checkerboard artifacts. *Distill*. 2016;1:e3.

40. Kodali N, Abernethy J, Hays J, Kira Z. On convergence and stability of gans. arXiv preprint. arXiv:1705.07215; 2017.
41. Cohen JP, Luck M, Honari S. Distribution matching losses can hallucinate features in medical image translation. Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention; 2018:529-536; Springer, New York, NY.
42. Edupuganti V, Mardani M, Vasanawala S, Pauly J. Uncertainty quantification in deep MRI reconstruction. *IEEE Trans Med Imaging*. 2020;40:239-250.
43. Wang K, Kellman M, Sandino CM, Zhang K, Vasanawala SS, Tamir JI. Memory-efficient Learning for High-Dimensional MRI Reconstruction. *In International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham; 2021:461-470.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Figure S1. UFLoss is able to capture perceptual similarities across anatomies and contrasts. The heat maps under a certain image show the feature correlations between all the patches from the image and the source patches from the source image (first column). The heat maps with green/blue borders correspond to different source patches whose borders have the same colors. The correlation results for PD contrasts using UFLoss and SSIM features are shown in the top and bottom rows, respectively

Figure S2. UFLoss retrieves patches with closer structural similarity compared to SSIM across different contrasts. The heat maps alongside the PD image show the feature correlation values between all the patches from the PD image and the source patch from the PDFS image (first column). The correlation results using UFLoss and SSIM features are shown on the right. Patches with the highest UFLoss and SSIM feature correlations in the PD image are visualized as zoomed-in patches with light blue borders. Feature correlation value are shown under each patch

Figure S3. Representative examples of 2D PD and 2D PDFS knee reconstruction with different UFLoss weighting factors during the training. Fully-sampled slices are retrospectively randomly under-sampled by a factor of 5, and reconstructed using MoDL with different weights of UFLoss. Pure ℓ_2 loss, combined ℓ_2 and UFLoss with $\mu = 0.5,1.5,4$, and pure UFLoss are included for evaluations. Zoomed-in details are shown along with each image

How to cite this article: Wang K, Tamir JI, De Goyeneche A, et al. High fidelity deep learning-based MRI reconstruction with instance-wise discriminative feature matching loss. *Magn Reson Med*. 2022;1-16. doi: 10.1002/mrm.29227