High-Fidelity Reconstruction with Instance-wise Discriminative Feature Matching Loss

Ke Wang[1], Jonathan I. Tamir[1,2], Stella X. Yu[1,3], and Michael Lustig[1]

[1]Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA, United States, [2]Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, United States, [3]International Computer Science Institute, University of California, Berkeley, Berkeley, CA, United States

Synopsis

Machine-learning based reconstructions have shown great potential to reduce scan time while maintaining high image quality. However, commonly used per-pixel losses for the training don't capture perceptual differences between the reconstructed and the ground truth images, leading to blurring or reduced texture. Thus, we incorporate a novel feature representation-based loss function with the existing reconstruction pipelines (e.g. MoDL), which we called Unsupervised Feature Loss (UFLoss). In-vivo results on both 2D and 3D reconstructions show that the addition of the UFLoss can encourage more realistic reconstructed images with much more detail compared to conventional methods (MoDL and Compressed Sensing).

Introduction

Deep learning-based reconstructions (including unrolled networks) have shown great potential for efficient image reconstruction from undersampled k-space measurements, well beyond the capabilities of compressed sensing (CS)[1] with fixed transform[2-6]. These reconstruction pipelines typically use the fully sampled images as the ground truth data to train the networks end-to-end with a mean-loss function over the training set. Unfortunately, the commonly used per-pixel losses (e.g., l_2, l_1 loss) don't necessarily capture perceptual or higher-order statistical differences between the reconstructed and the ground truth images. Recent work in computer vision has shown that perceptual losses can be used to further learn high-level image feature representations[7]. However, these perceptual losses are typically designed for instance-to-class discrimination (e.g. for classification), and not designed for instance-to-instance discrimination as in MR reconstruction. In this work, we incorporate a feature representation-based loss function with existing reconstruction pipelines, which we call Unsupervised Feature Loss (UFLoss). The UFLoss provides instance-level discrimination by mapping similar instances to similar low-dimensional feature vectors. A patch-wise mapping network is trained to map the image space to the feature space, where the UFLoss corresponds to the l_2 loss in the feature space. We compare MoDL[5] reconstruction results with and without our UFLoss on both 2D and 3D datasets, as well as with perceptual VGG loss. We show that the addition of the UFLoss can encourage more realistic reconstructed images with more details compared to conventional methods such as MoDL without UFLoss and CS.

Methods

In the learning-based image reconstruction workflow (Figure 1), we added a pre-trained feature loss on top of the per-pixel loss.

Training of the UFLoss mapping network:

The UFLoss mapping network is trained on patches of fully sampled images in an unsupervised fashion (Figure 2.a). The training is motivated by instance-level discrimination[8], where a feature mapping function is learned such that each patch is maximally separated from all other patches in the feature space. Suppose we have n patches $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}$ in the training sets and their unit-norm features $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}$ with $\mathbf{v_i} = f_\theta(\mathbf{x}) \in \mathbb{R}^d$. For a particular patch $\mathbf{x}$ with feature $\mathbf{v} = f_\theta(\mathbf{x})$, the probability of it being recognized as the i-th patch is given by the soft-max criterion:

$$P(i|\mathbf{v}) = \frac{\exp(\mathbf{v_i^T v})/\tau}{\sum_{\mathbf{j=1}}^{\mathbf{n}} \exp(\mathbf{v_i^T v})/\tau},$$

Where τ is a temperature parameter that controls the concentration level of the distribution. The learning objective is then 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:

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

After training, a target patch's nearest neighbors can be queried based on distance in the feature space (Figure 2.b). For training the feature embedding, we set $\tau = 1$, learning rate to 0.001 and the batch size to 16.

Training MoDL with UFLoss:

We apply our UFLoss to MoDL, a recently proposed unrolled iterative network which is trained end-to-end using ground-truth images. Instead of using only the per-pixel l_2 loss, we add our UFLoss based on the pre-trained network, resulting in the combined loss function:

$$\mathbf{loss}(\hat{x}) = \|x - \hat{x}\|_2^2 + \lambda \|f_\theta(x) - f_\theta(\hat{x})\|_2^2,$$

where f_θ is the feature mapping for the image, $x, \hat{x}$ are the ground truth and the reconstructed image respectively, and λ is the parameter to control the weights on the UFLoss. All the following experiments are implemented in Pytorch , with $\lambda = 1$, learning rate 0.0005 and batch size 1.

Datasets for the experiments:

We trained and evaluated our proposed UFLoss on both 2D and 3D knee datasets. We used the fastMRI[10] data for the 2D experiments. A total of 3200 slices from 200 cases were used for training. A total of 60 60×60 patches were cropped from each slice (~19,000 in total) and used to train the UFLoss mapping network ($d = 128$ for output). We conducted our 3D experiments on 20 Fully sampled 3D knee datasets (from mridata.org). A total of 5120 slices from 16 cases were used for training, and 40 40×40 patches from each slice (~20,400 in total) were used to train the UFLoss mapping network ($d = 128$ for output). In order to demonstrate the feasibility of our UFLoss on prospectively undersampled scans, we conducted 3D FSE experiments at 3T with 8x Poisson-disk sampling mask, and then used our pre-trained model to reconstruct the images. We compared the MoDL results with and without UFLoss, with the VGG loss (pre-trained on ImageNet), as well as CS reconstruction on the testing set. Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index (SSIM) were used as the metrics.

Results

Figure 3 and Figure 4 show a representative comparison of CS, MoDL, MoDL with VGG, and MoDL with UFLoss on retrospectively undersampled 2D and 3D reconstructions respectively. The reconstructions using MoDL with UFLoss show sharper edges, higher PSNR, and higher SSIM compared to other methods with the same training epochs.

Figure 5 shows the reconstruction results on prospectively undersampled 3D knee scans, using the pre-trained model. Detailed texture and sharper edges can be seen more clearly on the reconstruction results using VGG and UFLoss.

Conclusions

In this work, we presented UFLoss, an unsupervised learned feature loss, which can be easily trained on fully sampled images offline. With the addition of UFLoss, we were able to recover texture, small features and sharper edges with higher overall image quality under the learning-based reconstruction (MoDL).

Acknowledgements

No acknowledgement found.

References

1. Lustig, M., Donoho, D., & Pauly, J. M. (2007). Sparse MRI: The application of compressed sensing for rapid MR imaging. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 58(6), 1182-1195.

2. Diamond, S., Sitzmann, V., Heide, F., & Wetzstein, G. (2017). Unrolled optimization with deep priors. arXiv preprint arXiv:1705.08041.

3. Schlemper, J., Caballero, J., Hajnal, J. V., Price, A., & Rueckert, D. (2017, June). A deep cascade of convolutional neural networks for MR image reconstruction. In International Conference on Information Processing in Medical Imaging (pp. 647-658). Springer, Cham.

4. Hammernik, K., Klatzer, T., Kobler, E., Recht, M. P., Sodickson, D. K., Pock, T., & Knoll, F. (2018). Learning a variational network for reconstruction of accelerated MRI data. Magnetic resonance in medicine, 79(6), 3055-3071.

5. Aggarwal, H. K., Mani, M. P., & Jacob, M. (2018). Modl: Model-based deep learning architecture for inverse problems. IEEE transactions on medical imaging, 38(2), 394-405.

6. Tamir, J. I., Stella, X. Y., & Lustig, M. (2019). Unsupervised deep basis pursuit: Learning reconstruction without ground-truth data. In Proceedings of the 27th Annual Meeting of ISMRM.

7. Johnson, J., Alahi, A., & Fei-Fei, L. (2016, October). Perceptual losses for real-time style transfer and super-resolution. In European conference on computer vision (pp. 694-711). Springer, Cham.

8. Wu, Z., Xiong, Y., Yu, S., & Lin, D. (2018). Unsupervised feature learning via non-parametric instance-level discrimination. arXiv preprint arXiv:1805.01978.

9. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., ... & Lerer, A. (2017). Automatic differentiation in pytorch.

10. Zbontar, J., Knoll, F., Sriram, A., Muckley, M. J., Bruno, M., Defazio, A., ... & Zhang, Z. (2018). fastmri: An open dataset and benchmarks for accelerated mri. arXiv preprint arXiv:1811.08839.

11. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Figures
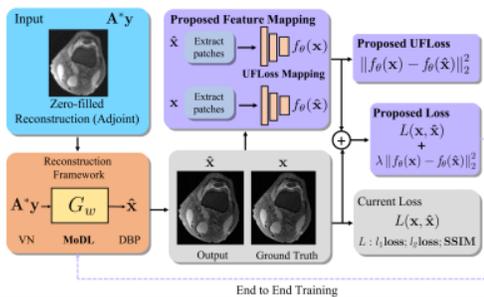
Figure 1. Overview of the proposed UFLoss under a general reconstruction framework (Here we use MoDL for the experiments). The current loss function (in grey) corresponds to the conventional loss (e.g. l1 loss, l2 loss, SSIM loss), while the proposed UFLoss (in purple) corresponds to the l2 loss in the feature space. End-to-end training is performed with the addition of our proposed UFLoss.
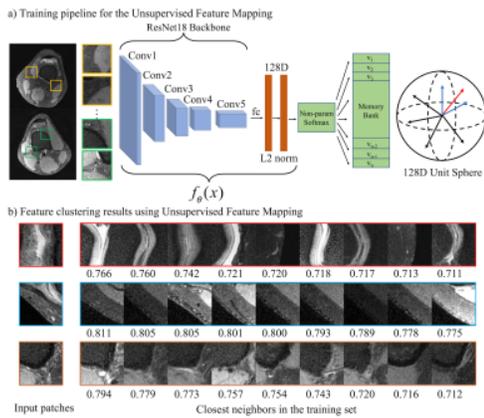
Figure 2. a) Training pipeline for the UFLoss mapping: the patches cropped from the images are passed through a ResNet 18[11] Backbone and an l_2 norm layer to map the images to the features on a 128D Unit Sphere. A memory bank is used to store the features from all the patches to compute the loss function given by the soft-max criterion. b) Feature clustering results using UFLoss mapping, where a target patch's nearest neighbors can be queried based on distance in the feature space. The inner products between the target patch and the neighbor patches in the feature space are shown under each patch.
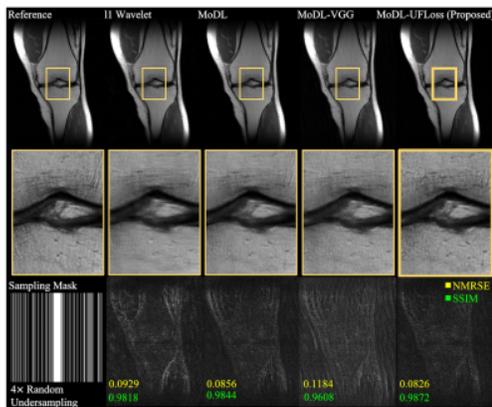
Figure 3. A representative comparison of CS, MoDL, MoDL with VGG, and MoDL with UFLoss on 2D reconstruction. 4 times variable density sampling mask was applied to the fully sampled data. Zoom-in details and the error maps show sharper edges and improved SNR/SSIM with our UFLoss compared to other methods. Sequence parameters: Echo train length: 4, matrix size: 320 x 320, in-plane resolution: 0.5mm x 0.5 mm, slice thickness: 3mm, TR: 2200 to 3000 ms and TE: 27 to 34 ms.
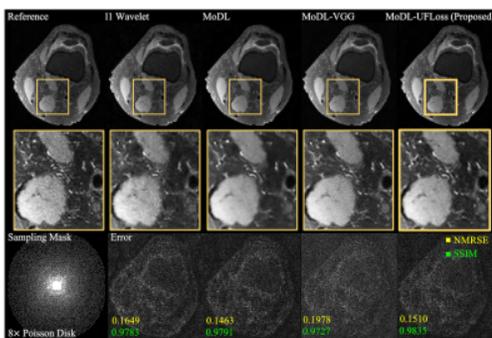
Figure 4. Comparison on different methods (l1 Wavelet, MoDL, MoDL with VGG and MoDL with UFLoss) on 3D knee reconstruction. All the data were acquired on a 3T GE Discovery MR 750, with an 8-channel HD knee coil. 8X Poisson Disk was used to mask the fully sampled k-space. Reconstruction using MoDL with UFLoss shows sharper structures and finer image details compared to other methods. Scan parameters included a matrix size of 320x320x256, and TE/TR of 25ms/1550ms.
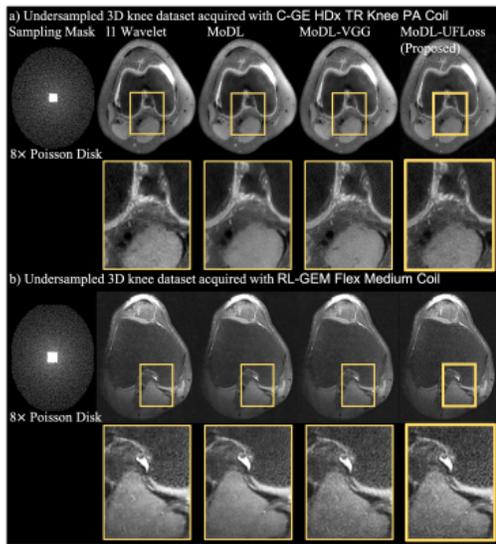
Figure 5. Reconstruction results on prospectively undersampled 3D knee scans with pre-trained model using different methods. a) Underdampled 3D FSE knee dataset on mridata.org. (volume size: 320x288x236, TR/TE = 1400/20.46 ms, different from training set). The zoom-in details show some finer structures in MoDL with UFLoss and MoDL with VGG compared to other methods. b) Undersampled 3D FSE knee dataset with the same scan parameters as the training set. Some detailed texture and sharper edges can be seen more clearly on the reconstruction results using MoDL with VGG and UFLoss.