UCLA
UCLA Previously Published Works

Title

Uncertainty-aware physics-driven deep learning network for free-breathing liver fat and R2* quantification using self-gated stack-of-radial MRI

Permalink

Journal

ISSN

Authors

Shih, Shu-Fu
Kafali, Sevgi Gokce
Calkins, Kara L
et al.

Publication Date

DOI

Peer reviewed

Uncertainty-Aware Physics-Driven Deep Learning Network for Free-Breathing Liver Fat and R_2* Quantification using Self-Gated Stack-of-Radial MRI

Shu-Fu Shih[1,2], **Sevgi Gokce Kafali**[1,2], **Kara L. Calkins**[3], **Holden H. Wu**[1,2,*]

[1] Department of Radiological Sciences, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

[2] Department of Bioengineering, University of California Los Angeles, Los Angeles, CA, USA

[3] Department of Pediatrics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

Abstract

Purpose: To develop a deep learning-based method for rapid liver proton-density fat fraction (PDFF) and R_2* quantification with built-in uncertainty estimation using self-gated free-breathing stack-of-radial MRI.

Methods: This work developed an uncertainty-aware physics-driven deep learning network (UP-Net) to (1) suppress radial streaking artifacts due to undersampling after self-gating, (2) calculate accurate quantitative maps, and (3) provide pixel-wise uncertainty maps. UP-Net incorporated a phase augmentation strategy, generative adversarial network architecture, and an MRI physics loss term based on a fat-water and R_2* signal model. UP-Net was trained and tested using free-breathing multi-echo stack-of-radial MRI data from 105 subjects. UP-Net uncertainty scores were calibrated in a validation dataset and used to predict quantification errors for liver PDFF and R_2* in a testing dataset.

Results: Compared with images reconstructed using compressed sensing (CS), UP-Net achieved structural similarity index >0.87 and normalized root mean squared error <0.18. Compared with reference quantitative maps generated using CS and graph-cut (GC) algorithms, UP-Net achieved low mean differences (MD) for liver PDFF (–0.36%) and R_2* (–0.37 s^{-1}). Compared with breath-holding Cartesian MRI results, UP-Net achieved low MD for liver PDFF (0.53%) and R_2* (6.75 s^{-1}). UP-Net uncertainty scores predicted absolute liver PDFF and R_2* errors with low MD of 0.27% and 0.12 s^{-1} compared to CS+GC results. The computational time for UP-Net was 79 ms/slice, while CS+GC required 3.2 min/slice.

Conclusion: UP-Net rapidly calculates accurate liver PDFF and R_2* maps from self-gated free-breathing stack-of-radial MRI. The pixel-wise uncertainty maps from UP-Net predict quantification errors in the liver.

*Correspondence to: Holden H. Wu, Ph.D., Department of Radiological Sciences, 300 UCLA Medical Plaza, Suite B119, Los Angeles, CA 90095, Phone: (310) 267-6843, Fax: (310) 825-9118, HoldenWu@mednet.ucla.edu.

1. INTRODUCTION

Chronic liver disease is a global health burden[1–3]. Liver disease is characterized by histological changes that include hepatic steatosis, inflammation, fibrosis, and iron deposition[4–10]. Progressive liver disease is associated with cirrhosis and hepatocellular carcinoma, and can culminate in liver failure[11]. Biopsy is considered the standard technique for diagnosing liver diseases. However, biopsy suffers from sampling bias, is invasive, and is associated with complications[12].

MRI evaluates hepatic steatosis and iron overload by quantifying proton-density fat fraction (PDFF) and R_2*, using chemical-shift-encoded multi-echo Dixon techniques that acquire and fit data to a signal model that accounts for the multi-peak fat spectrum and R_2* component[13]. Conventional Dixon techniques using a multi-echo 3D Cartesian sequence[14] are sensitive to motion and require breath-holding (10–20 sec). The breath-holding requirement limits the volumetric coverage and resolution, and can be challenging for patients[15]. Recently, a multi-echo 3D stack-of-radial technique[16,17] has been developed for free-breathing liver PDFF and R_2* quantification and demonstrated accurate results in subjects with non-alcoholic fatty liver disease (NAFLD)[18]. To compensate for respiratory motion in free-breathing radial data acquisition, self-gating is used to reconstruct images from a subset of data with consistent motion behavior. However, motion self-gating introduces radial undersampling artifacts in the images and quantitative maps. These artifacts can be mitigated by acquiring more radial spokes[17] or using constrained reconstruction[19], but these strategies require a longer acquisition and/or computational time.

Accurate and rapid signal fitting is another challenge in PDFF and R_2* quantification. Due to the non-convex structure of the signal model and ambiguities in resonant frequencies of water/fat protons with respect to B_0 field variations, signal fitting can converge to a local minimum solution and lead to fat-water swaps. State-of-the-art graph-cut (GC)-based methods[20,21] impose smoothness constraints on the field map and use optimization algorithms to reduce the occurrence of fat-water swaps. However, the GC-based algorithms are computationally expensive with computation time on the order of 10 sec/slice[21].

Compared with iterative constrained reconstruction methods for MRI, such as compressed sensing (CS)[22,23], deep learning (DL)-based methods can rapidly enhance or reconstruct images from undersampled data by leveraging datasets from prior scans. Previous studies have developed novel DL networks for MRI enhancement or reconstruction from undersampled Cartesian data[24,25]. Although there were DL networks developed for undersampled radial MRI[26–29], there is a lack of investigation regarding multi-echo radial MRI for PDFF and R_2* mapping. On the other hand, DL has also been used to replace the computationally expensive fat-water signal fitting process. Different network architectures[30–33] and loss functions[34] have been proposed to separate fat/water signals

or generate PDFF/R_2* maps. However, these methods only investigated fully-sampled Cartesian data and did not consider radial acquisition or data undersampling.

Developing, evaluating, and translating DL-based methods for quantitative MRI parameter mapping can be challenging because quantification errors can be difficult to detect by visual inspection. Confidence levels of quantification accuracy from the DL network outputs were not typically characterized in previous studies[30–34]. Recently, there have been initial developments in incorporating uncertainty estimation in the DL networks for MR image reconstruction[35–37]. These works showed promise by investigating the relationships between estimated uncertainty scores and reconstruction errors. A recent work showed promising results of using uncertainty estimation for quantitative MRI PDFF maps obtained from DL[38]. The study indicated that the uncertainty scores were related to the noise levels in the input data. However, the relationship between the uncertainty scores and quantification accuracy was not established.

In this work, we developed an uncertainty-aware physics-driven deep learning network (UP-Net) that can rapidly calculate accurate liver PDFF and R_2* maps using multi-echo images from undersampled self-gated free-breathing stack-of-radial MRI data. UP-Net simultaneously (1) suppressed radial streaking artifacts due to undersampling after self-gating, (2) calculated accurate quantitative liver PDFF and R_2* maps, and (3) provided pixel-wise uncertainty maps for each quantitative parameter within a rapid inference time <100 ms/slice. We calibrated the UP-Net uncertainty scores and demonstrated the ability to predict liver PDFF and R_2* quantification errors using the uncertainty scores.

2. METHODS

2.1 Uncertainty-Aware Physics-Driven Deep Learning Network (UP-Net)

We proposed UP-Net (Figure 1) to generate accurate quantitative maps from undersampled 2D multi-echo images and provide pixel-wise uncertainty maps which can be used to predict quantification errors. UP-Net contained two concatenated network modules for artifact suppression and parameter mapping. The first module took 2D multi-echo undersampled images x as the input and generated enhanced 2D multi-echo images $\hat{m}$ with suppressed undersampling artifacts. For x and $\hat{m}$, multi-echo images, including both the real and imaginary components, were stacked along the channel dimension. The second module transformed $\hat{m}$ to quantitative parameter maps $\hat{p}$ and their corresponding uncertainty maps $\hat{u}$. In our case of multi-parameter fitting, $\hat{p}$ and $\hat{u}$ are 3D tensors where different 2D quantitative maps are stacked along the channel dimension. UP-Net requires reference multi-echo images and reference quantitative maps for training. Details regarding reference data generation are described in section 2.5.

Convolutional neural networks (ConvNet) have been proposed to effectively suppress artifacts from undersampling[24–29]. Recently, there are works showing that generative adversarial networks (GAN) can improve the quality of the reconstructed images for radial MRI compared to conventional ConvNet[39]. GAN uses a discriminator to help generate images that closely resemble the reference images and match the intended data distribution, which can improve the image reconstruction or artifact suppression results.

We investigated a GAN architecture for the artifact suppression module and decided to use it for UP-Net based on results in an ablation study (see section 2.8). The generator was implemented using a 2D UNet architecture[40], and the discriminator was implemented using the architecture proposed in[41]. To deal with image contrast variation across subjects, instance normalization[42] was used in both the generator and the discriminator. A detailed diagram for the UP-Net implementation is presented in Supplementary Figure S1.

We considered the quantitative parameter output as distributions which can be characterized using pixel-wise means $\hat{p}$ and pixel-wise variances $\hat{u}$ from a Bayesian perspective[43,44]. We interpreted $\hat{p}$ and $\hat{u}$ as the quantitative maps and the corresponding uncertainty maps. For each pixel index j, a larger value of $\hat{u_j}$ indicates a wider spread of the distribution and therefore the associated $\hat{p_j}$ has higher uncertainty. By assuming a prior data distribution, the network can be trained to predict $\hat{p}$ and $\hat{u}$ simultaneously using the loss function introduced in section 2.2. In light of the deeply correlated nature of $\hat{p}$ and $\hat{u}$, we used a "bifurcated UNet" architecture (Figures 1 and S1) for the parameter mapping module. This architecture has one shared encoder that extracts features from multi-contrast images $\hat{m}$, and two separate decoders that generate parameter maps $\hat{p}$ and uncertainty maps $\hat{u}$. Because the uncertainty score, or the variance of a distribution, should always be nonnegative, a softplus layer ($Softplus(x) = \log(1 + e^x)$) was added prior to the uncertainty map output.

2.2 Loss Function for UP-Net Training

We constructed a loss function with 5 components for supervised training of UP-Net:

$$L_{UP-Net} = w_1 L_{imgMSE} + w_2 L_{imgGAN} + w_3 L_{mapMSE} + w_4 L_{physics} + w_5 L_{uncert} \tag{1}$$

An image mean squared error (MSE) loss was used to measure the errors between enhanced ($\hat{m}$) and reference (m) multi-echo images:

$$L_{imgMSE} = \frac{1}{N_j} \sum_j \left(\hat{m}_j - m_j\right)^2 \tag{2}$$

, where j represents the pixel index and N_j is the total number of pixels in the multi-echo images. We trained the GAN architecture using a Wasserstein GAN loss[45], which can be formulated as:

$$\min_G \max_D \mathbb{E}_{m \sim p_{train}(m)}[D(m)] - \mathbb{E}_{\hat{m} \sim p_G(\hat{m})}[D(G(\hat{m}))] \tag{3}$$

, where G represents the generator, D represents the discriminator. The loss for updating the generator G was:

$$L_{imgGAN} = \mathbb{E}_{\hat{m} \sim p_G(\hat{m})}[D(G(\hat{m}))] \tag{4}$$

We also used an MSE loss that measures the errors between quantitative maps from UP-Net ($\hat{p}$) and reference data (p):

$$L_{imgMSE} = \frac{1}{N_j} \sum_j (\hat{p}_j - p_j)^2 \tag{5}$$

To promote learning of the signal fitting process we used an MRI physics loss:

$$L_{physics} = \frac{1}{N_j} (\hat{m} - Q(\hat{p}))^2 \tag{6}$$

where Q represents an operator that transforms the quantitative maps to multi-echo images based on the MRI signal equation. In this work where we investigated PDFF and R_2* quantification, the operator Q we used was:

$$Q(\hat{p}) = Q(W, F, R_2^*, \varphi, TE) = \left(W + F \cdot \left(\sum_{m=1}^{M} a_m \cdot e^{i2\pi f_m TE} \right) \right) \cdot e^{-R_2^* TE} \cdot e^{i2\pi \varphi TE} \tag{7}$$

where W, F, R_2^*, φ represent the quantitative water maps, fat maps, R_2* maps, and B_0 field maps. A 7-peak fat model with amplitudes a_m and frequencies f_m were also included[46].

To predict quantitative parameter outputs with corresponding uncertainty scores, we used an uncertainty loss:

$$L_{uncert} = \frac{\|\hat{p} - p\|_1}{\hat{u}} + log(\hat{u}) \tag{8}$$

This uncertainty loss function is equivalent to performing maximum a posteriori (MAP) inference where a Laplace distribution[43] is assumed for each quantitative parameter in each pixel. We can also understand this loss function from a more intuitive perspective. First, in regions where the $\|\hat{p} - p\|_1$ error minimization is difficult (e.g., regions with lower signal-to-noise ratio), increased values of $\hat{u}$ can reduce the loss, therefore capturing uncertainty. Second, the $log(\hat{u})$ term can serve as a regularization term to avoid unconstrained increase in the uncertainty score.

The relative weights for each loss component in Equation 1 can impact the results. We chose the weight combination that achieved the lowest PDFF and R_2* quantification errors in the validation set: $w_1 = 0.2$, $w_2 = 0.2$, $w_3 = 0.2$, $w_4 = 0.3$, and $w_5 = 0.1$.

2.3 Training Strategy for UP-Net

To shorten the convergence time for training UP-Net, we used a step-by-step training strategy.

Step 1: Pre-train the artifact suppression module using pairs of input undersampled images x and reference images m as the training data and using only L_{imgMSE} and L_{imgGAN} for the loss function.

Step 2: Pre-train the parameter mapping module without the uncertainty estimation path using pairs of reference multi-contrast images m and reference quantitative maps p as the training data and using only L_{mapMSE} and $L_{physics}$ for the loss function.

Step 3: Load the weights trained from **Steps 1** and **2**, and then train the entire UP-Net end-to-end without the uncertainty path using L_{imgMSE}, L_{imgGAN}, L_{mapMSE} and $L_{physics}$ for the loss function (i.e., not including the uncertainty loss term yet).

Step 4: Train the entire UP-Net using training sets of undersampled images x, reference images m and reference quantitative maps p with the full loss function L_{UP-Net}.

For all training steps, data augmentation for images/maps was performed by mirroring (no flip or horizontal flip) and rotating by n*90 degrees (n=0,1,2,3). This augmented the training data size to 8 times that of the original training data. In addition, we used a "phase augmentation" strategy to further augment training data by adding a phase offset to the multi-echo input images, multi-echo reference images, and reference fat and water complex signals at the same time (Supplementary Figure S2). The signal magnitudes were not changed, and the relationship between images and quantitative maps were not modified. In each epoch during training, we generated 3 more instances for each 2D slice in this manner (i.e., in addition to the original data, 3 different phase offsets were applied to generate 3 more instances). The phase offsets were randomly selected between $0\sim2\pi$. This strategy aimed to improve robustness to phase variations, which is important in separating fat/water signals.

2.4 Data Acquisition

In a HIPAA-compliant and IRB-approved study, we acquired MR images from 105 subjects, including healthy subjects and subjects with suspected or confirmed NAFLD, at 3T (MAGNETOM Skyra or Prisma, Siemens Healthineers, Erlangen, Germany). Fifty-seven of the subjects were adults (34 females, 23 males; age 48.16±19.01 years; body mass index [BMI]: 26.98±5.94kg/m^2) and 48 of the subjects were children (19 females, 29 males; age 13.06±2.99 years; BMI: 22.85±8.41kg/m^2). Written informed consent, parental permission, and assent, if applicable, were obtained for all subjects prior to research procedures. We scanned using a prototype free-breathing multi-echo gradient-echo 3D stack-of-radial sequence with bipolar readout gradients (**FB Radial**, parameters in Table 1)[16,17]. To compare with standard breath-holding techniques, we acquired an additional breath-hold bipolar multi-echo gradient-echo 3D Cartesian sequence (**BH Cartesian**, parameters in Table 1)[14]. We separated the data into training (N=63), validation (N=21), and testing (N=21) datasets using a 3:1:1 ratio (Table 2). Subject information and data were entered into a secure database for management and analysis[47].

We trained and tested UP-Net using only FB Radial data, while the BH Cartesian data served as an external reference for evaluation of PDFF and R$_2$* quantification accuracy. For BH Cartesian data, images and the quantitative PDFF and R$_2$* maps were reconstructed using vendor-provided software on the scanner.

2.5 Reference Data Preparation for UP-Net Training

UP-Net was trained in a supervised approach, which demands high-quality multi-echo images and quantitative maps with minimal artifacts to serve as references. However, it is challenging to acquire fully-sampled data for reconstruction of motion-resolved 3D volumetric images in the abdomen. To satisfy the Nyquist sampling criteria after self-gating, longer acquisition time is needed, which may increase sensitivity to motion effects. Previous works have used CS to generate images and quantitative maps for DL network training in applications where a fully-sampled reference dataset is difficult to acquire[48,49]. Following a similar strategy, we acquired nominally fully-sampled stack-of-radial data before applying motion self-gating, and used CS to reconstruct motion self-gated images with suppressed undersampling artifacts.

The workflow for generating the training data for UP-Net is shown in Figure 2. Gradient delays were calibrated to correct the radial trajectory for FB Radial data reconstruction[16]. We extracted a projection-based self-navigator from the $k_x = k_y = 0$ line in k-space[23] to track respiratory motion along the z dimension. A sliding window approach was applied along the motion dimension to bin the k-space data into 6 respiratory motion states where each bin contained 40% of the entire k-space data (effective data undersampling factor = 2.5 in each state). We estimated coil sensitivity maps using a phased array beamforming technique[50,57], which has been shown to suppress radial artifacts resulting from hardware imperfections (e.g., gradient non-linearity and field inhomogeneity). We formulated the 2D CS reconstruction problem as[23]:

$$x* = argmin_x \left\| FSx - y \right\|_2^2 + \lambda_1 TV^{motion}(x) + \lambda_2 \sum\nolimits_{echo, state} \left\| Wavelet\left(x_{echo, \ state}\right) \right\|_1$$

, where F represents the non-uniform fast Fourier Transform (NUFFT) operator, S denotes coil sensitivity maps, x is the reconstructed multi-echo images, y is the acquired k-space data, and λ_1 and λ_2 are regularization parameters. The regularization parameters were chosen manually to balance between undersampling artifact reduction and image sharpness. After CS reconstruction, we calculated quantitative maps (including complex fat/water components, R_2* map, and B_0 field map) by fitting the CS-reconstructed multi-echo images to a 7-peak fat model with a single R_2* component[46] (same as Eq. 7) using GC-based algorithms[20,21]. Local fat-water swaps still occurred in certain slices and were difficult to correct using GC-based algorithms; we excluded these slices from the training dataset. We generated body masks from the first-echo CS-reconstructed magnitude images, and applied the body masks to the CS-reconstructed images and the corresponding quantitative maps for background artifact and noise suppression. We will refer to the reference CS-reconstructed self-gated free-breathing stack-of-radial images as FB+CS and the corresponding quantitative maps reconstructed by GC-based algorithms as FB+CS+GC.

The input images x to UP-Net were coil-combined 6-echo images using 40% of FB Radial data near the end-expiration state (Figure 1). The real and imaginary components from each echo were stacked along the channel dimension (6 echoes × real/imaginary components = 12 channels). The output from the artifact suppression module had the same data dimensions as the input images (12 channels), and were fed into the parameter mapping module. The

output from the parameter mapping module contained (1) complex-valued fat and water components, R_2* map, and field map stacked along the channel dimension and (2) three uncertainty maps for PDFF, R_2*, and field map stacked along the channel dimension. PDFF maps were generated from complex fat/water components for calculation of L_{mapMSE} and L_{uncert}, while the complex-valued fat/water components were directly used in $L_{physics}$.

Based on PDFF and R_2* quantification accuracy in the validation dataset, the hyperparameters for the end-to-end UP-Net training were chosen as: batch size=32, initial learning rates=0.0001, and epochs=150, using the Adam optimizer.

2.6 Evaluation of UP-Net Image Quality and Quantification Accuracy

We evaluated the performance of UP-Net in terms of image quality and quantification accuracy of the output images and maps in the testing dataset. For image quality, we compared the enhanced image results from UP-Net with the reference FB+CS images using normalized root mean squared error (NRMSE) and structure similarity index (SSIM). For quantification accuracy, we calculated differences in PDFF and R_2* quantification results using liver regions of interest (ROIs) for (1) FB+UP-Net versus FB+CS+GC and (2) FB+UP-Net versus BH Cartesian. ROIs with area of 5-cm^2 were placed in the right lobe of the liver by a trained researcher while avoiding large vessels and bile ducts[18]. A total of 3 ROIs were placed in the upper, middle and lower liver (one ROI at each level) for each subject. Bland-Altman analysis was performed to evaluate PDFF and R_2* accuracy by calculating the mean difference (MD) and 95% limits of agreement (LoA) between different methods.

2.7 Evaluation of UP-Net Uncertainty Estimation

We evaluated the performance of UP-Net uncertainty estimation in terms of its ability to predict quantification errors according to the following steps.

Step 1: Complete UP-Net training.

Step 2: In the validation dataset, measure (1) quantification errors ($\sum_{i \in ROI} |\hat{p}_i - p_i|$) between UP-Net and reference FB+CS+GC results and (2) UP-Net uncertainty scores in the liver ROIs. Use a linear correlation model to generate "calibration curves" between quantification errors and UP-Net uncertainty scores for each quantitative parameter separately. Calculate Spearman correlation coefficients and test for statistical significance.

Step 3: Output UP-Net uncertainty scores for liver ROIs in the testing dataset. Transform UP-Net uncertainty scores to predicted quantification errors using the calibration curves.

Step 4: Perform Bland-Altman analysis on predicted quantification errors versus actual quantification errors in the testing dataset, with respect to FB+CS+GC results, for each quantitative parameter separately.

2.8 UP-Net Ablation Study

We performed an ablation study to assess the contributions of the key components used in UP-Net, including phase augmentation, GAN loss, MRI physics loss, uncertainty estimation, and joint end-to-end training strategy. After training each ablated model with the same training dataset, we compared the results in the testing dataset using NRMSE and SSIM for image quality, and absolute errors in liver PDFF and R_2^* for quantification accuracy. We used the Wilcoxon signed-rank test to evaluate if the performance of the ablated network models had significant difference versus the performance of UP-Net. $p < 0.01$ was considered significant.

3.   RESULTS

3.1   UP-Net Image Quality and Quantification Accuracy

Figure 3 shows representative first-echo magnitude and phase images from UP-Net input, UP-Net output results, and FB+CS results in two subjects with NAFLD (a 47-year-old male and a 17-year-old male). UP-Net suppressed the radial undersampling streaking artifacts in the liver and in the background (arrows in Figure 3), and achieved high SSIM compared with FB+CS. Figures 4 and 5 show representative quantitative PDFF, R_2^* and field map results in axial and coronal orientations and the corresponding uncertainty maps in two NAFLD subjects (same subjects as in Figure 3). UP-Net generated accurate PDFF/R_2^*/field maps compared with FB+CS+GC references in the liver ROIs. Most regions show low quantification errors (Figures 4 and 5) in all 3 quantitative parameters. Regions corresponding to air usually had large quantification errors. The uncertainty maps show high intensities (red arrows in Figures 4 and 5) and characterize the lower confidence in these regions in air.

Bland-Altman plots for liver PDFF and R_2^* quantification accuracy are shown in Figure 6. For PDFF quantification, FB+UP-Net achieved MD = −0.36% compared with FB+CS+GC, and MD = 0.53% compared with BH Cartesian. For R_2^* quantification, FB+UP-Net achieved MD = −0.37 s^{-1} compared with FB+CS+GC, and MD = 6.75 s^{-1} compared with BH Cartesian. LoA between FB+UP-Net versus FB+CS+GC was narrower than FB+UP-Net versus BH Cartesian for both PDFF and R_2^* quantification.

3.2   UP-Net Uncertainty Estimation

Linear correlation results comparing absolute quantification errors versus uncertainty scores of three quantitative parameters in liver ROIs in the validation dataset are shown in Figure 7a. The Spearman correlation coefficients for PDFF, R_2^* and field map were 0.358 ($p<0.05$), 0.466 ($p<0.01$), and 0.503 ($p<0.01$), respectively. These calibrated linear regression curves were used to convert uncertainty scores measured in the testing dataset to predicted quantification errors. The Bland-Altman plots for UP-Net predicted errors versus actual quantification errors in liver ROIs are shown in Figure 7b. MDs between UP-Net predicted errors versus actual absolute quantifications errors were 0.27%, 0.12 s^{-1}, and 0.19 Hz for PDFF, R_2^*, and field map, respectively. Note that the quantification errors were all generally low to begin with.

3.3 UP-Net Ablation Study Results

Table 3 shows the results of our ablation study. UP-Net achieved higher mean SSIM of 0.872 and lower mean NRMSE of 0.173 compared with the ablated UP-Net models without phase augmentation or GAN loss (all p<0.01). Compared with the ablated UP-Net model without the MRI physics loss, UP-Net achieved lower mean PDFF error of −0.36% and lower mean R_2* error of −0.37 s^{-1} (both p<0.01). UP-Net without uncertainty estimation did not have significant difference in image quality and quantification accuracy when compared with UP-Net (i.e., the addition of the uncertainty path did not degrade quantification accuracy). UP-Net without joint training achieved higher mean PDFF error of −0.46% (p<0.01) compared to UP-Net with joint training.

3.4 Processing/Reconstruction Time

The entire CS reconstruction was implemented in MATLAB 2021b (MathWorks, Natick, MA). Data preparation steps of gradient calibration, self-gating, NUFFT, and beamforming-based coil combination (Figures 1 and 2) required a total time of 30 sec/slice. Repeated forward and inverse NUFFT are the bottleneck for the CS reconstruction method. To improve computational performance, we implemented the CS reconstruction method using GPU-based NUFFT packages[58]. CS reconstruction took 3 min/slice on an Intel Xeon E5–2660 CPU with 128GB RAM and an NVIDIA v100 GPU with 32GB memory. We used the ISMRM fat-water toolbox[59] and code from previous works[21] for GC fitting algorithms, which required 15 seconds/slice on the same CPU. We implemented UP-Net using Python 3.8.10 and Pytorch 1.12.1. UP-Net required 28 hours to train on an NVIDIA v100 GPU with 32GB memory. With the prepared data as input, UP-Net took 79 msec/slice for network inference (using same hardware as network training). A more detailed analysis of the total operation counts for each method was provided in Supplementary Table S1.

4. DISCUSSION

We developed an uncertainty-aware physics-driven deep learning network that accurately quantifies liver PDFF and R_2* using undersampled self-gated free-breathing multi-echo stack-of-radial MRI. Compared with previous works on DL-based fat/water separation and/or R_2* mapping[30–34], our study has two main contributions. First, we investigated a DL approach for PDFF and R_2* mapping from undersampled radial MRI data. Unlike previously proposed networks that learned mapping from fully-sampled Cartesian images to fat/water signals or quantitative maps[30–34], UP-Net generates accurate quantitative maps from images impacted by radial undersampling artifacts. We incorporated artifact suppression and parameter mapping into one end-to-end network. This substantially reduced the computational time for image artifact suppression compared to time-intensive CS methods and fat-water signal fitting compared to GC algorithms. Second, our proposed network has built-in uncertainty estimation that generates pixel-wise uncertainty maps for different quantitative parameters. Uncertainty estimation to assess the confidence levels in DL-based MRI reconstruction and quantitative parameter mapping results is a nascent direction[35–38]. We specifically investigated the application of uncertainty estimation in DL-based PDFF and R_2* quantification and demonstrated that a calibration method for the UP-Net uncertainty scores can be used to predict absolute liver PDFF and R_2* quantification

errors in UP-Net parameter maps to within 1% and 3 s^{-1}, respectively, compared to actual errors with respect to reference methods.

To suppress radial undersampling streaking artifacts, we used UNet as the backbone architecture, which has been used in previous work for radial streaking reduction in 2D slices[26,27], dynamic 2D cardiac images[29], and 2D images from different respiratory phases[51]. In this work, we adapted the input/output dimensions of UNet to accommodate the 2D multi-echo images. We stacked the real and imaginary components from all of the multi-echo images along the channel dimension to preserve the consistency of the magnitude and phase input information for PDFF and R_2* quantification. We also adopted a GAN architecture and a phase augmentation strategy for image quality improvement. Due to limited memory on the GPU for network training, correlations between neighboring slices were not considered in this study. Networks that can efficiently process multi-echo 2D+slice or multi-echo 3D volumetric data could be further investigated. Although it is possible to omit the artifact suppression module and use one single network to generate quantitative maps directly from self-gated undersampled radial images, UP-Net with a modular architecture can provide more accurate quantitative maps with less radial streaking artifacts (example in Supplementary Figure S3).

In addition to rapid computational time, another potential advantage of using UP-Net or other DL-based methods for fat-water separation is reducing the occurrence of fat-water swaps. In our datasets, there were slices with local fat-water swaps (usually around the liver dome) using the GC methods. These slices required manually adjusting GC parameters, such as constraints on B_0 field map smoothness or range, to address the swaps. When training UP-Net, we excluded data with fat-water swaps and performed phase augmentation, which helped the network to learn reliable fat-water separation in the presence of B_0 field map variations. The use of UP-Net can potentially reduce the occurrence of fat-water swaps (Supplementary Figures S4) and avoid the extra time/effort needed to check and fix fat-water swaps.

One concern of DL-based fat-water separation is whether the network could perform accurate mapping for datasets with liver PDFF values outside the range in the training dataset. UP-Net has two advantages that could allow it to generalize to these cases. First, we used an MRI physics loss that will constrain the output to follow the fat-water signal model. Second, even though the training dataset we used in this work has a maximum liver PDFF around 30%, UP-Net still learned from the signal characteristics in fat-dominant tissues (e.g., subcutaneous adipose tissue) with PDFF up to 90%. To investigate this, we created synthetic testing datasets with higher liver PDFF, and used UP-Net to perform PDFF mapping. A representative example in Supplementary Figure S5 shows that UP-Net can indeed quantify higher liver PDFF values (e.g., >40%) that were not included in the training dataset. Another concern of DL-based fat-water separation is whether the network can be adapted to several different body parts. Although different body parts may have different B_0 field map ranges and variations, their signal characteristics are described by the same fat-water signal model. After training UP-Net on a certain dataset (e.g., liver and upper abdomen), the fat-water signal model is implicitly learned. Through transfer learning and fine tuning, UP-Net can potentially be applied to other body parts (e.g., lower abdomen).

Fully-sampled free-breathing motion-resolved volumetric abdominal stack-of-radial MRI data is often impractical to acquire. We used CS to generate images and quantitative maps with suppressed radial streaking artifacts. CS methods have already been validated for PDFF quantification using undersampled Cartesian MRI[52] and for PDFF and R_2* quantification using undersampled radial MRI data[19]. For complete evaluation of our UP-Net quantification accuracy, we also compared UP-Net results with standard 3D BH Cartesian MRI. Many previous DL-based fat-water signal fitting methods were only evaluated on individual fat and water maps[30–33]. We evaluated our results on quantitative PDFF maps. Similar to a previous DL-based method for joint PDFF and R_2* mapping using Cartesian MRI[34], we also achieved low biases in PDFF, R_2*, and field map values versus reference methods. Notably, we trained and tested our method on a larger dataset (105 subjects). In contrast, a previous report considered 31 subjects[34]. Previous DL methods did not investigate their results in NAFLD subjects, while our UP-Net was trained and evaluated in a population including healthy subjects and subjects with suspected or confirmed NAFLD. In our Bland-Altman analysis of PDFF and R_2* quantification accuracy, the MD between FB+UP-Net and FB+CS+GC was smaller than the MD between FB+UP-Net and BH Cartesian. This was expected because UP-Net was trained using reference data from the FB+CS+GC method. The MD and LoA of PDFF and R_2* quantification comparing FB+UP-Net versus BH Cartesian are similar to results in previous studies comparing self-gated FB stack-of-radial MRI with BH Cartesian[18].

We carefully examined the contributions from key components in UP-Net, including phase augmentation, GAN loss, and MRI physics loss. Among these components, MRI physics loss was especially important for accurate parameter quantification. From our ablation study, the network without MRI physics loss generated larger biases in both PDFF and R_2* quantification. MRI physics loss considered the relationship between multi-echo signals and quantitative parameters and did not require reference quantitative maps. However, the MRI physics loss alone may not provide sufficient information to resolve fat-water swaps. By adding an MSE loss for quantitative maps and training with reference non-swapped maps, we directed UP-Net to learn the spatial distribution of fat- or water- dominant pixels and reduce fat-water swaps. In applications that focus on the signal magnitude, phase information in DL results is often discarded or overlooked. Our use of a phase augmentation strategy strengthened UP-Net's ability to learn complex-valued signal relationships by including images with the same magnitude but different phase. This strategy can also be applied in applications that need accurate phase, including temperature mapping and quantitative susceptibility mapping. GAN architectures for imaging tasks is an active research topic. In this work, we used Wasserstein loss in our GAN architecture, which has previously been used in Cartesian MRI reconstruction[53]. More complicated GAN architectures and loss functions designed for medical images[54] could be investigated further.

The "black box" nature of DL-based methods for MRI is an important concern and potential barrier to clinical translation. Uncertainty estimation in DL networks[35–38] presents a promising approach to provide context and assess confidence in DL outputs for clinical applications that demand a high level of numerical accuracy, including the use of quantitative maps for diagnostic decisions. In this study, we showed that with calibration, UP-Net uncertainty scores predicted quantification errors in a separate testing

dataset. These promising results have some potential applications. For example, confidence masks can be generated by thresholding the uncertainty scores and then overlaid on the UP-Net quantitative parameter maps. Radiologists can avoid making measurements and decisions in areas with higher uncertainty scores and have more confidence in using DL-generated images and quantitative maps. A recent study has demonstrated that by passing uncertainty information in concatenated tasks, the performance of the downstream task (e.g., segmentation or detection) can be improved[55]. The uncertainty maps generated by UP-Net can potentially provide information and improve subsequent automatic liver MRI analysis, such as DL-based liver segmentation and disease classification.

This study has limitations. First, we did not investigate the influence of different data undersampling factors on UP-Net performance. We used a 40% data acceptance window (2.5-fold undersampling) on nominally fully-sampled data, as suggested in previous studies[17,18]. Self-gating data acceptance rates can be further reduced to improve motion fidelity. The number of acquired radial spokes can also be reduced to investigate additional scan acceleration. However, higher undersampling factors pose more difficulties in both generating high-quality reference data and training UP-Net. Adjustments such as adding k-space consistency layers[56] might be required for UP-Net to address higher undersampling factors. Second, we trained and tested UP-Net using data with specific number of echoes, TE, TR, and flip angle. These sequence parameters were closely related to PDFF/R_2* accuracy in the data we used. Our current analysis on the quantification accuracy and uncertainty prediction may not be directly applicable in other datasets with different sequence parameters. Third, the calculation of the UP-Net uncertainty loss term required reference quantitative maps. Therefore, the UP-Net uncertainty values reflect differences between UP-Net results and results from reference methods. Fourth, we calibrated the PDFF and R_2* uncertainty estimation in the validation dataset only using ROIs in the liver. This approach required CS and GC reference reconstruction results for calibration. In addition, different calibration curves may be needed to quantify DL uncertainty in other tissues, such as subcutaneous and visceral adipose tissues. Fifth, we used linear regression to investigate the relationship between UP-Net uncertainty scores and quantification errors. However, this approach may not be sufficient to characterize all the factors at play. PDFF measurements are results of relative amounts of two chemical shift species and can have different inherent uncertainty at different PDFF levels. R_2* uncertainty depends on combinations of the number of echoes, the chosen echo times, and the underlying true R_2* values. Because PDFF and R_2* are incorporated together in the fat-water signal model, multi-variate models can also be considered to improve uncertainty characterization and calibration in the future.

5. CONCLUSION

In this study, we developed an uncertainty-aware physics-driven deep learning network that rapidly calculates accurate liver PDFF and R_2* maps from undersampled free-breathing self-gated multi-echo stack-of-radial images and provides pixel-wise uncertainty maps. GAN architecture, phase augmentation, and MRI physics loss improved the UP-Net image quality and quantification accuracy for liver PDFF and R_2*. We demonstrated that UP-Net uncertainty scores can be used to predict absolute quantification errors in liver PDFF and R_2*.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

DATA AVAILABILITY STATEMENT

To support the findings in this work, our code for UP-Net implementation is available at https://github.com/HoldenWuLab/UP-Net.

REFERENCES

1. Williams R Global challenges in liver disease. Hepatology 2006;44(3):521–526. doi: 10.1002/hep.21347. [PubMed: 16941687]

2. Rinella ME. Nonalcoholic fatty liver disease: a systematic review. JAMA 2015;313(22):2263–2273. doi:10.1001/jama.2015.5370. [PubMed: 26057287]

3. Vernon G, Baranova A, Younossi Z. Systematic review: the epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults. Aliment Pharmacol Ther 2011;34(3):274–285. doi: 10.1111/j.1365-2036.2011.04724.x. [PubMed: 21623852]

4. Powell EE, Jonsson JR, Clouston AD. Steatosis: co-factor in other liver diseases. Hepatology 2005;42(1):5–13. doi: 10.1002/hep.20750. [PubMed: 15962320]

5. Batts KP. Iron overload syndromes and the liver. Mod Pathol 2007;20(1):S31–S39. doi: 10.1038/modpathol.3800715. [PubMed: 17486050]

6. Yokoo T, Browning JD. Fat and iron quantification in the liver: past, present, and future. Top Magn Reson Imaging 2014;23(2):73–94. doi: 10.1097/RMR.0000000000000016. [PubMed: 24690618]

7. Aigner E, Weiss G, Datz C. Dysregulation of iron and copper homeostasis in nonalcoholic fatty liver. World J Hepatol 2015;7(2):177. doi: 10.4254/wjh.v7.i2.177. [PubMed: 25729473]

8. Dongiovanni P, Fracanzani AL, Fargion S, Valenti L. Iron in fatty liver and in the metabolic syndrome: a promising therapeutic target. J Hepatol 2011;55(4):920–932. doi: 10.1016/j.jhep.2011.05.008. [PubMed: 21718726]

9. Milic S, Mikolasevic I, Orlic L, Devcic E, Starcevic-Cizmarevic N, Stimac D, Kapovic M, Ristic S. The role of iron and iron overload in chronic liver disease. Med Sci Monit 2016;22:2144. doi: 10.12659/MSM.896494. [PubMed: 27332079]

10. Wood MJ, Skoien R, Powell LW. The global burden of iron overload. Hepatol Int 2009;3(3):434–444. doi: 10.1007/s12072-009-9144-z. [PubMed: 19669241]

11. Zoller H, Tilg H. Nonalcoholic fatty liver disease and hepatocellular carcinoma. Metabolism 2016;65(8):1151–1160. doi: 10.1016/j.metabol.2016.01.010. [PubMed: 26907206]

12. Sumida Y, Nakajima A, Itoh Y. Limitations of liver biopsy and non-invasive diagnostic tests for the diagnosis of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. World J Gastroenterol 2014;20(2):475. doi: 10.3748/wjg.v20.i2.475. [PubMed: 24574716]

13. Hines CD, Frydrychowicz A, Hamilton G, Tudorascu DL, Vigen KK, Yu H, McKenzie CA, Sirlin CB, Brittain JH, Reeder SB. T1 independent, T2* corrected chemical shift based fat–water separation with multi-peak fat spectral modeling is an accurate and precise measure of hepatic steatosis. J Magn Reson Imaging 2011;33(4):873–881. doi: 10.1002/jmri.22514. [PubMed: 21448952]

14. Zhong X, Nickel MD, Kannengiesser SA, Dale BM, Kiefer B, Bashir MR. Liver fat quantification using a multi-step adaptive fitting approach with multi-echo GRE imaging. Magn Reson Med 2014;72(5):1353–1365. doi: 10.1002/mrm.25054. [PubMed: 24323332]

15. Armstrong T, Ly KV, Murthy S, Ghahremani S, Kim GHJ, Calkins KL, Wu HH. Free-breathing quantification of hepatic fat in healthy children and children with nonalcoholic fatty liver disease using a multi-echo 3-D stack-of-radial MRI technique. Pediatr Radiol 2018;48(7):941–953. doi: 10.1007/s00247-018-4127-7. [PubMed: 29728744]

16. Armstrong T, Dregely I, Stemmer A, Han F, Natsuaki Y, Sung K, Wu HH. Free-breathing liver fat quantification using a multiecho 3 D stack-of-radial technique. Magn Reson Med 2018;79(1):370–382. doi: 10.1002/mrm.26693. [PubMed: 28419582]

17. Zhong X, Armstrong T, Nickel MD, Kannengiesser SA, Pan L, Dale BM, Deshpande V, Kiefer B, Wu HH. Effect of respiratory motion on free-breathing 3D stack-of-radial liver relaxometry and improved quantification accuracy using self-gating. Magn Reson Med 2020;83(6):1964–1978. doi: 10.1002/mrm.28052. [PubMed: 31682016]

18. Armstrong T, Zhong X, Shih S-F, Felker E, Lu DS, Dale BM, Wu HH. Free-breathing 3D stack-of-radial MRI quantification of liver fat and R2* in adults with fatty liver disease. Magn Reson Imaging 2022;85:141–152. doi: 10.1016/j.mri.2021.10.016. [PubMed: 34662702]

19. Schneider M, Benkert T, Solomon E, Nickel D, Fenchel M, Kiefer B, Maier A, Chandarana H, Block KT. Free-breathing fat and R2* quantification in the liver using a stack-of-stars multi-echo acquisition with respiratory-resolved model-based reconstruction. Magn Reson Med 2020;84(5):2592–2605. doi: 10.1002/mrm.28280. [PubMed: 32301168]

20. Hernando D, Kellman P, Haldar J, Liang ZP. Robust water/fat separation in the presence of large field inhomogeneities using a graph cut algorithm. Magn Reson Med 2010;63(1):79–90. doi: 10.1002/mrm.22177. [PubMed: 19859956]

21. Cui C, Wu X, Newell JD, Jacob M. Fat water decomposition using globally optimal surface estimation (GOOSE) algorithm. Magn Reson Med 2015;73(3):1289–1299. doi: 10.1002/mrm.25193. [PubMed: 24604689]

22. Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. Magn Reson Med 2007;58(6):1182–1195. doi: 10.1002/mrm.21391. [PubMed: 17969013]

23. Feng L, Axel L, Chandarana H, Block KT, Sodickson DK, Otazo R. XD-GRASP: golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing. Magn Reson Med 2016;75(2):775–788. doi: 10.1002/mrm.25665. [PubMed: 25809847]

24. Lee D, Yoo J, Tak S, Ye JC. Deep residual learning for accelerated MRI using magnitude and phase networks. IEEE Trans Biomed Eng 2018;65(9):1985–1995. doi: 10.1109/TBME.2018.2821699. [PubMed: 29993390]

25. Schlemper J, Caballero J, Hajnal JV, Price A, Rueckert D. A deep cascade of convolutional neural networks for MR image reconstruction. International Conference on Information Processing in Medical Imaging; 2017. Springer. p 647–658. doi: 10.1007/978–3-319–59050-9_51.

26. Han Y, Yoo J, Kim HH, Shin HJ, Sung K, Ye JC. Deep learning with domain adaptation for accelerated projection-reconstruction MR. Magn Reson Med 2018;80(3):1189–1205. doi: 10.1002/mrm.27106. [PubMed: 29399869]

27. Kofler A, Dewey M, Schaeffter T, Wald C, Kolbitsch C. Spatio-temporal deep learning-based undersampling artefact reduction for 2D radial cine MRI with limited training data. IEEE Trans Med Imaging 2019;39(3):703–717. doi: 10.1109/TMI.2019.2930318. [PubMed: 31403407]

28. Lv J, Chen K, Yang M, Zhang J, Wang X. Reconstruction of undersampled radial free-breathing 3D abdominal MRI using stacked convolutional auto-encoders. Med Phys 2018;45(5):2023–2032. doi: 10.1002/mp.12870. [PubMed: 29574939]

29. El-Rewaidy H, Fahmy AS, Pashakhanloo F, Cai X, Kucukseymen S, Csecs I, Neisius U, Haji-Valizadeh H, Menze B, Nezafat R. Multi-domain convolutional neural network (MD-CNN) for radial reconstruction of dynamic cardiac MRI. Magn Reson Med 2021;85(3):1195–1208. doi: 10.1002/mrm.28485. [PubMed: 32924188]

30. Goldfarb JW, Craft J, Cao JJ. Water–fat separation and parameter mapping in cardiac MRI via deep learning with a convolutional neural network. J Magn Reson Imaging 2019;50(2):655–665. doi: 10.1002/jmri.26658. [PubMed: 30701621]

31. Andersson J, Ahlström H, Kullberg J. Separation of water and fat signal in whole-body gradient echo scans using convolutional neural networks. Magn Reson Med 2019;82(3):1177–1186. doi: 10.1002/mrm.27786. [PubMed: 31033022]

32. Cho J, Park H. Robust water–fat separation for multi-echo gradient-recalled echo sequence using convolutional neural network. Magn Reson Med 2019;82(1):476–484. doi: 10.1002/mrm.27697. [PubMed: 30790344]

33. Liu K, Li X, Li Z, Chen Y, Xiong H, Chen F, Bao Q, Liu C. Robust water–fat separation based on deep learning model exploring multi-echo nature of mGRE. Magn Reson Med 2021;85(5):2828–2841. doi: 10.1002/mrm.28586. [PubMed: 33231896]

34. Jafari R, Spincemaille P, Zhang J, Nguyen TD, Luo X, Cho J, Margolis D, Prince MR, Wang Y. Deep neural network for water/fat separation: Supervised training, unsupervised training, and no training. Magn Reson Med 2021;85(4):2263–2277. doi: 10.1002/mrm.28546. [PubMed: 33107127]

35. Edupuganti V, Mardani M, Vasanawala S, Pauly J. Uncertainty quantification in deep MRI reconstruction. IEEE Trans Med Imaging 2020;40(1):239–250. doi: 10.1109/TMI.2020.3025065. [PubMed: 32956045]

36. Zhang Z, Romero A, Muckley MJ, Vincent P, Yang L, Drozdzal M. Reducing uncertainty in undersampled MRI reconstruction with active acquisition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019. p 2049–2058.

37. Schlemper J, Castro DC, Bai W, Qin C, Oktay O, Duan J, Price AN, Hajnal J, Rueckert D. Bayesian deep learning for accelerated MR image reconstruction. International Workshop on Machine Learning for Medical Image Reconstruction; 2018. Springer. p 64–71.

38. Shih S-F, Kafali SG, Armstrong T, Zhong X, Calkins KL, Wu HH. Deep learning-based parameter mapping with uncertainty estimation for fat quantification using accelerated free-breathing radial MRI. 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI); 2021. IEEE. p 433–437. doi: 10.1109/ISBI48211.2021.9433938.

39. Liu F, Samsonov A, Chen L, Kijowski R, Feng L. SANTIS: sampling-augmented neural network with incoherent structure for MR image reconstruction. Magnetic resonance in medicine 2019;82(5):1890–1904. [PubMed: 31166049]

40. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention; 2015. Springer. p 234–241. doi: 10.1007/978–3-319–24574-4_28.

41. Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z. Photo-realistic single image super-resolution using a generative adversarial network. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p 4681–4690.

42. Ulyanov D, Vedaldi A, Lempitsky V. Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:160708022 2016.

43. Kendall A, Gal Y. What uncertainties do we need in bayesian deep learning for computer vision? arXiv preprint arXiv:170304977 2017.

44. Abdar M, Pourpanah F, Hussain S, Rezazadegan D, Liu L, Ghavamzadeh M, Fieguth P, Cao X, Khosravi A, Acharya UR. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Inf Fusion 2021. doi: 10.1016/j.inffus.2021.05.008.

45. Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. International conference on machine learning; 2017. PMLR. p 214–223.

46. Ren J, Dimitrov I, Sherry AD, Malloy CR. Composition of adipose tissue and marrow fat in humans by 1H NMR at 7 Tesla. J Lipid Res 2008;49(9):2055–2062. doi: 10.1194/jlr.D800010-JLR200. [PubMed: 18509197]

47. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing

translational research informatics support. J Biomed Inform 2009;42(2):377–381. doi: 10.1016/j.jbi.2008.08.010. [PubMed: 18929686]

48. Fu Z, Mandava S, Keerthivasan MB, Li Z, Johnson K, Martin DR, Altbach MI, Bilgin A. A multi-scale residual network for accelerated radial MR parameter mapping. Magn Reson Imaging 2020;73:152–162. doi: 10.1016/j.mri.2020.08.013. [PubMed: 32882339]

49. Shao J, Ghodrati V, Nguyen KL, Hu P. Fast and accurate calculation of myocardial T1 and T2 values using deep learning Bloch equation simulations (DeepBLESS). Magn Reson Med 2020;84(5):2831–2845. [PubMed: 32416010]

50. Mandava S, Keerthivasan MB, Martin DR, Altbach MI, Bilgin A. Radial streak artifact reduction using phased array beamforming. Magn Reson Med 2019;81(6):3915–3923. doi: 10.1002/mrm.27689. [PubMed: 30756432]

51. Freedman JN, Gurney-Champion OJ, Nill S, Shiarli A-M, Bainbridge HE, Mandeville HC, Koh D-M, McDonald F, Kachelrieß M, Oelfke U. Rapid 4D-MRI reconstruction using a deep radial convolutional neural network: Dracula. Radiother Oncol 2021;159:209–217. doi: 10.1016/j.radonc.2021.03.034. [PubMed: 33812914]

52. Doneva M, Börnert P, Eggers H, Mertins A, Pauly J, Lustig M. Compressed sensing for chemical shift-based water–fat separation. Magn Reson Med 2010;64(6):1749–1759. [PubMed: 20859998]

53. Jiang M, Yuan Z, Yang X, Zhang J, Gong Y, Xia L, Li T. Accelerating CS-MRI reconstruction with fine-tuning Wasserstein generative adversarial network. IEEE Access 2019;7:152347–152357. doi: 10.1109/ACCESS.2019.2948220.

54. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: A review. Med Image Anal 2019;58:101552. doi: 10.1016/j.media.2019.101552. [PubMed: 31521965]

55. Mehta R, Christinck T, Nair T, Bussy A, Premasiri S, Costantino M, Chakravarty M, Arnold DL, Gal Y, Arbel T. Propagating uncertainty across cascaded medical imaging tasks for improved deep learning inference. IEEE Trans Med Imaging 2021;41(2):360–373. doi: 10.1109/TMI.2021.3114097.

56. Liang D, Cheng J, Ke Z, Ying L. Deep magnetic resonance image reconstruction: Inverse problems meet neural networks. IEEE Signal Process Mag 2020;37(1):141–151. doi: 10.1109/MSP.2019.2950557. [PubMed: 33746470]

57. Shih S-F, Wu HH. Beamforming-based coil combination method with phase constraint for streaking artifact reduction in Radial MRI. In Proceedings of the 30th annual meeting of ISMRM, London, United Kingdom, 2022. #1697.

58. Knoll F, Schwarzl A, Diwoky C, Sodickson DK. gpuNUFFT - an open source GPU library for 3D regridding with direct Matlab interface. In Proceedings of the 22nd annual meeting of ISMRM, Milan, Italy, 2014. #4297.
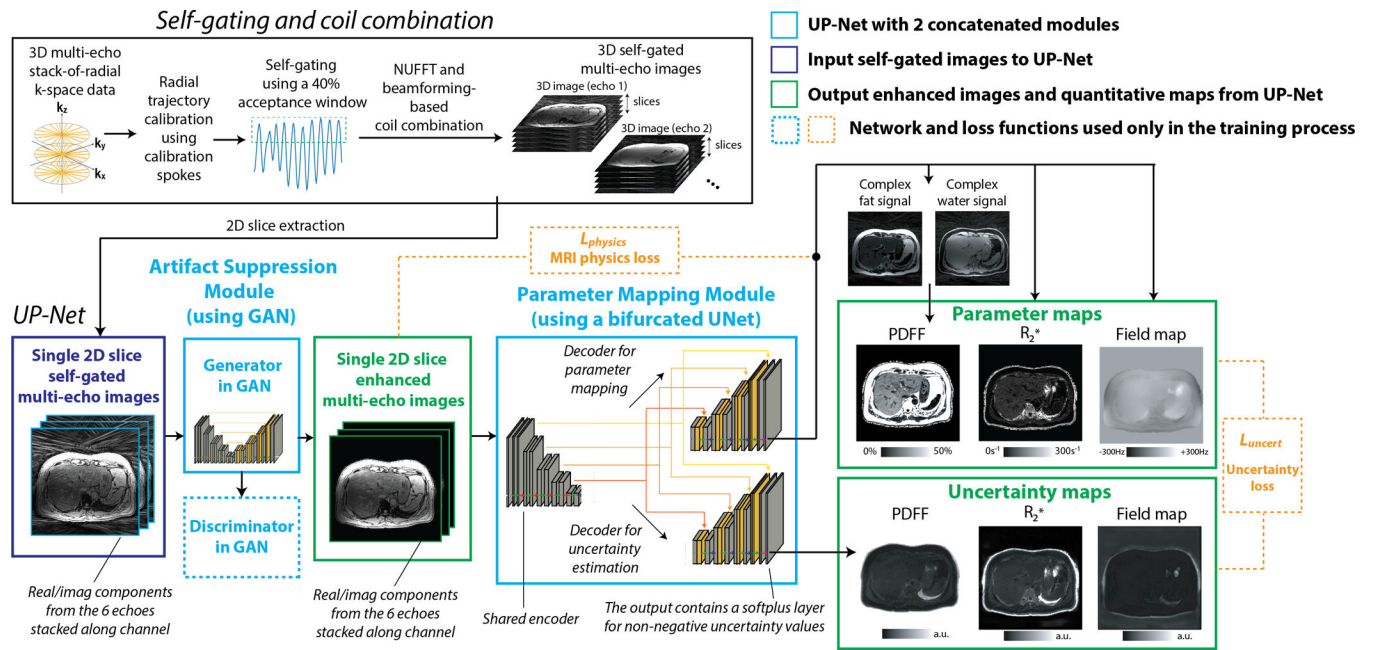
59. ISMRM fat water toolbox. 2012. (http://ismrm.org/workshops/FatWater12/data.htm)

Figure 1.
The proposed uncertainty-aware physics-driven deep learning network (UP-Net) for rapid free-breathing proton-density fat fraction (PDFF) and R_2^* quantification from self-gated multi-echo stack-of-radial MR images. The artifact suppression module used a generative adversarial network (GAN) architecture to reduce the radial undersampling artifacts due to self-gating. The parameter mapping module used a bifurcated UNet structure, which had a shared encoder and two decoders, to calculate parameter maps (pixel-wise means) and uncertainty maps (pixel-wise variances). See Supplementary Figure S1 for more details. NUFFT: non-uniform fast Fourier transform.
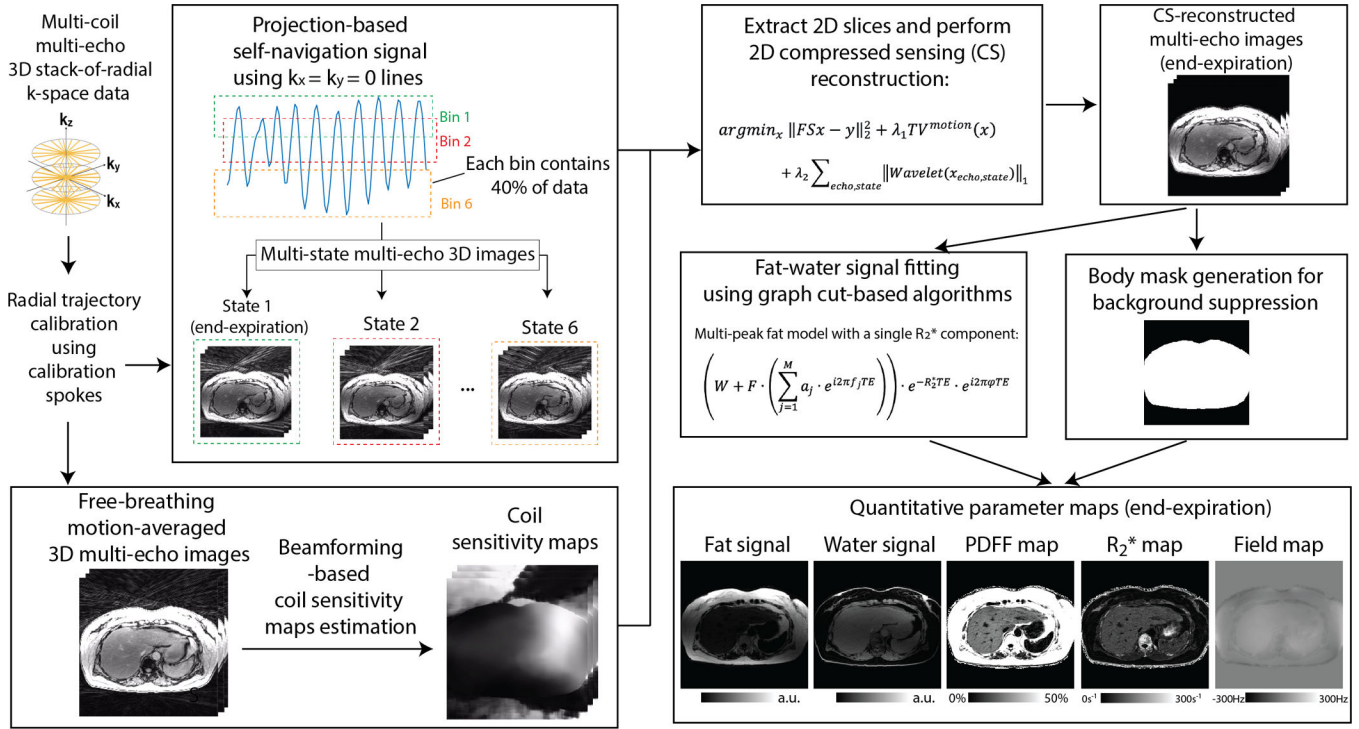
Figure 2.

The workflow for generating reference data (multi-echo images and quantitative maps) for training UP-Net. Nominally fully-sampled stack-of-radial k-space data were binned into 6 respiratory motion states using projection-based self-navigators. A 2D compressed sensing (CS) framework with beamforming-based coil sensitivity maps was used to reconstruct multi-echo images with reduced undersampling streaking artifacts. Quantitative maps were generated by fitting the multi-echo images to a fat-water signal model with a single R_2^* component. Body masks were generated from the CS-reconstructed first-echo images for background suppression.
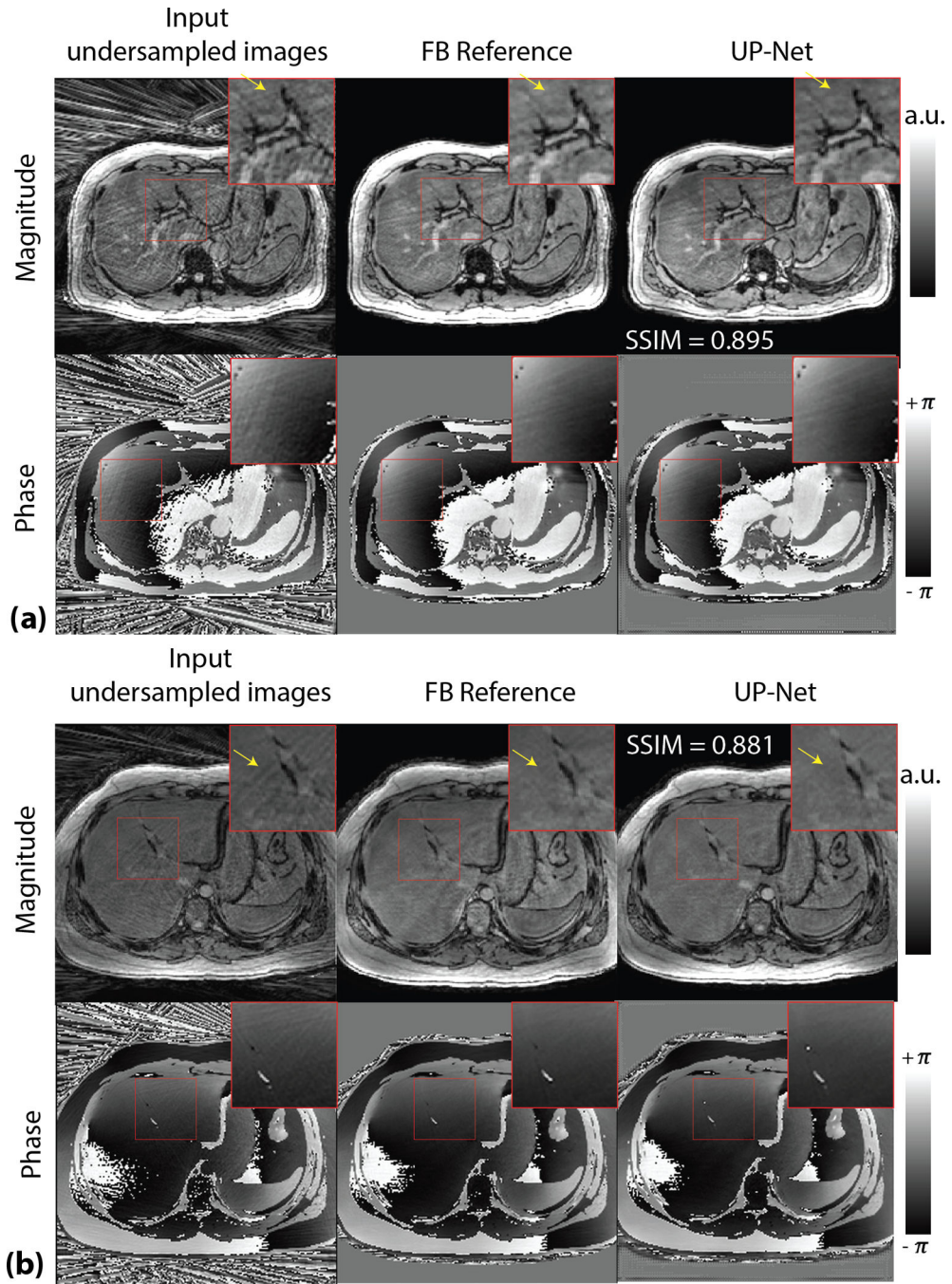
Figure 3.
Representative free-breathing (FB) stack-of-radial first-echo images from self-gated input images, UP-Net output results, and reference images reconstructed by compressed sensing (CS). **(a)** Results from a 47-year-old male (BMI=28.0kg/m^2) in the testing set. **(b)** Results from a 17-year-old male (BMI=30.4kg/m^2) in the testing set. Structural similarity index (SSIM) values comparing UP-Net output with reference images are shown. Arrows point to streaking artifacts in the self-gated images that are suppressed in the UP-Net output and reference images. BMI: body mass index.
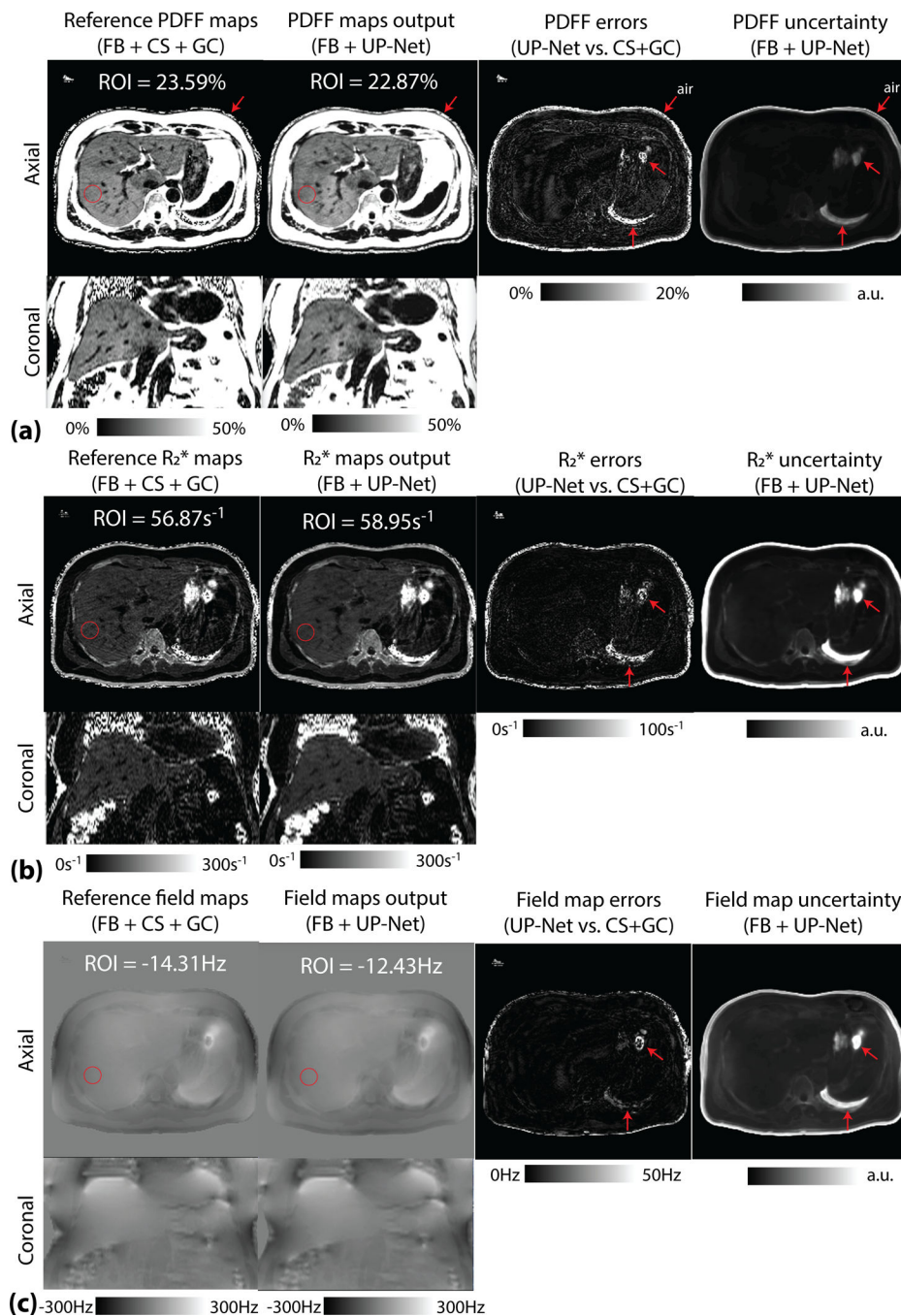
Figure 4.
Representative free-breathing (FB) **(a)** proton-density fat fraction (PDFF), **(b)** R_2* and **(c)** field maps and corresponding uncertainty maps from the same subject in Figure 3a. PDFF, R_2* and field map errors were generally low when comparing UP-Net results with reference maps reconstructed using compressed sensing (CS) and graph-cut (GC) algorithms. In regions with larger quantification errors, higher UP-Net uncertainty scores were observed in all 3 quantitative maps (red arrows). Errors and higher uncertainty scores around the body were in regions corresponding to air.
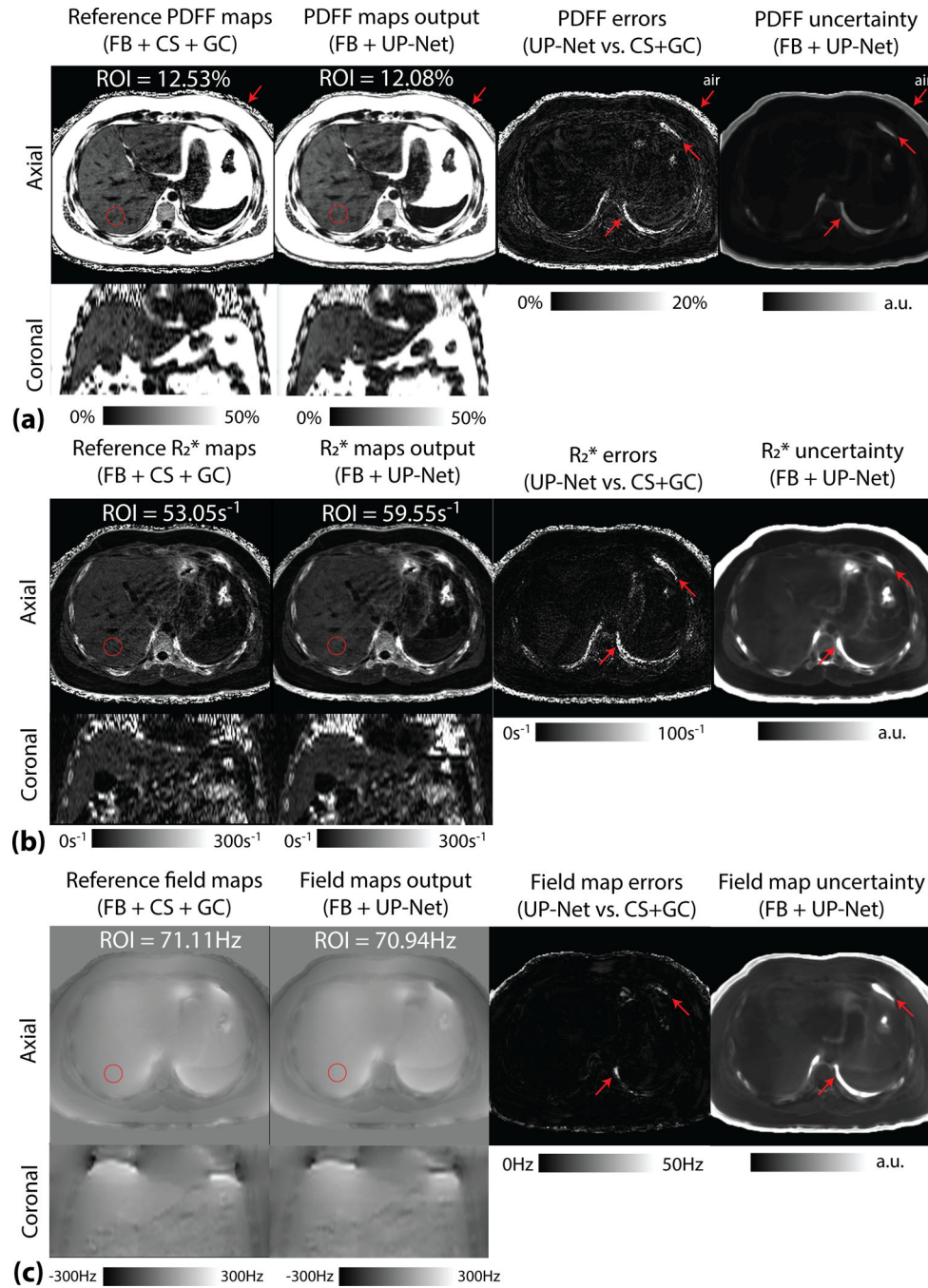
Figure 5.

Representative free-breathing (FB) **(a)** proton-density fat fraction (PDFF), **(b)** R$_2$* and **(c)** field maps and corresponding uncertainty maps from the same subject in Figure 3b. PDFF, R$_2$* and field map errors were generally low when comparing UP-Net results with reference maps reconstructed using compressed sensing (CS) and graph-cut (GC) algorithms. In regions with larger quantification errors, higher UP-Net uncertainty scores were observed in all 3 quantitative maps (red arrows). Errors and higher uncertainty scores around the body were in regions corresponding to air.
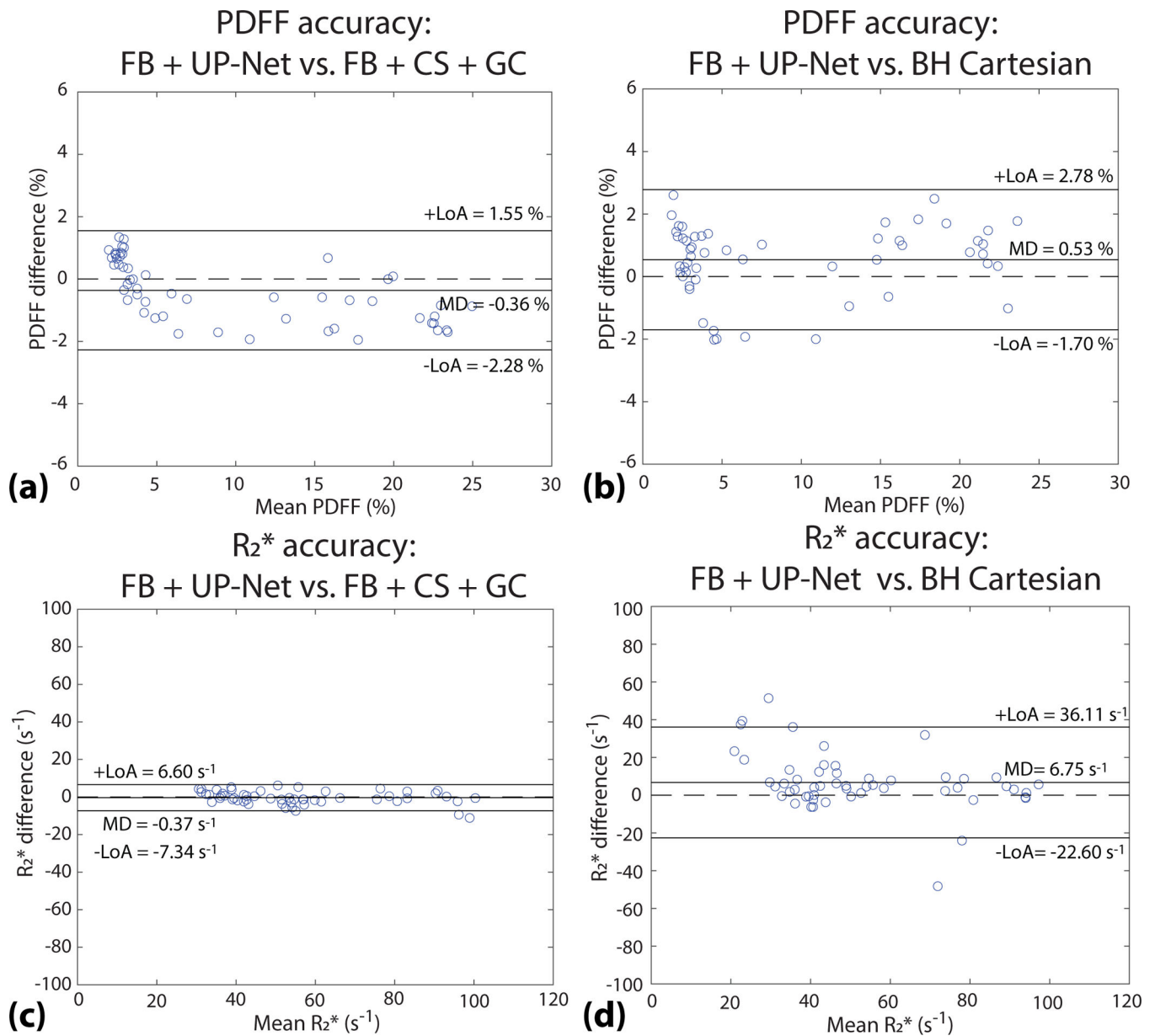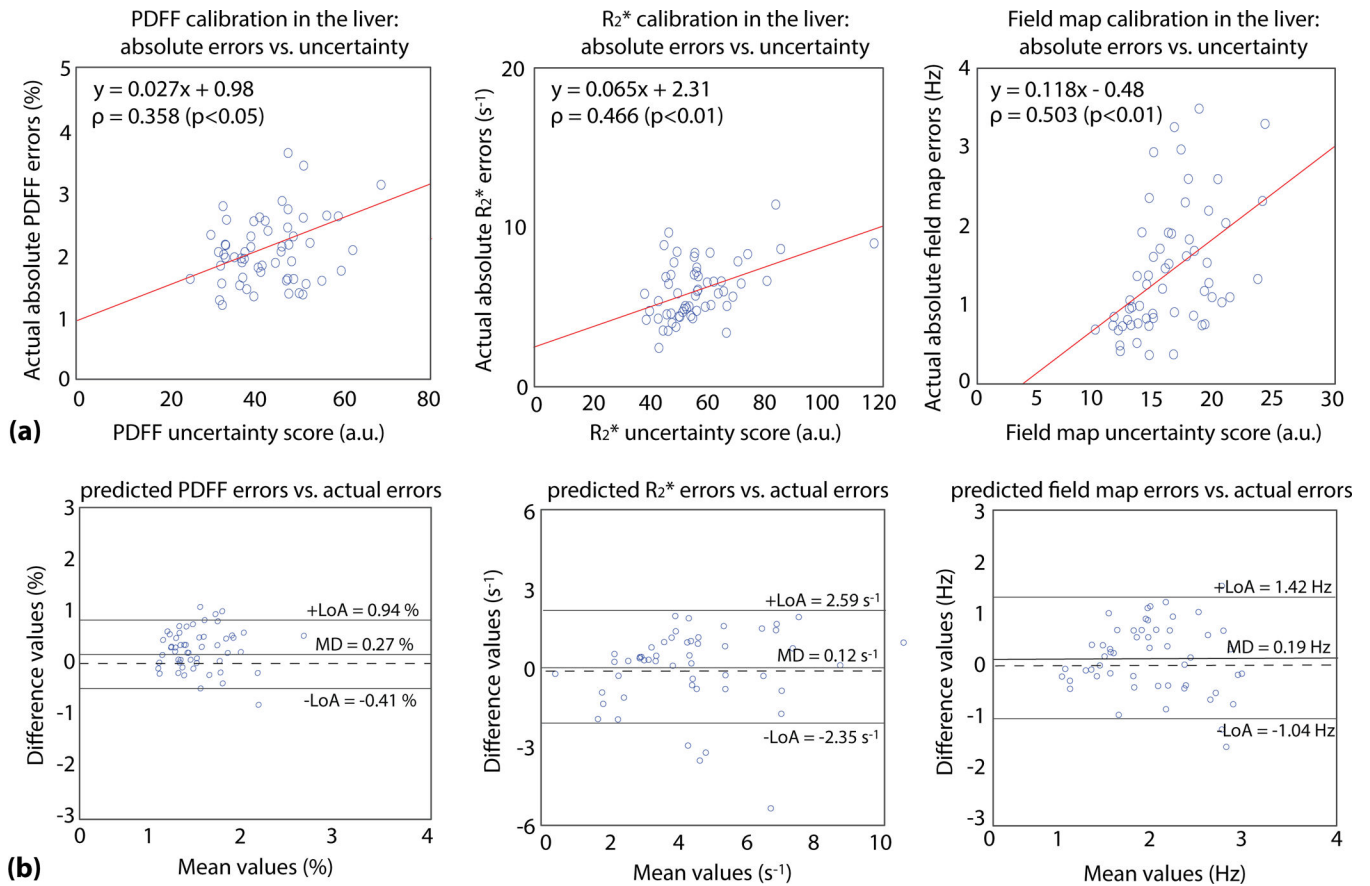
Figure 6.
(a-b) Bland-Altman plots comparing liver proton-density fat fraction (PDFF) values from UP-Net output maps versus free-breathing (FB) reference maps and breath-holding (BH) Cartesian maps. **(c-d)** Bland-Altman plots comparing liver R_2* values from UP-Net output maps versus FB reference maps and BH Cartesian maps. The dashed lines represent zero difference. The solid lines represent mean differences (MD) and 95% limits of agreements (LoA).

Figure 7.
(a) Correlation plots between absolute quantification errors (UP-Net outputs versus reference compressed sensing and graph-cut fitting results) and UP-Net uncertainty scores in liver regions of interest (ROIs) in the validation dataset. Linear regression was performed to calibrate PDFF, R_2^* and field map uncertainty scores with respect to the absolute errors. **(b)** Bland-Altman plots comparing the errors predicted from UP-Net uncertainty scores versus the actual absolute quantification errors in PDFF, R_2^* and field map in liver ROIs in the testing dataset. The black dashed lines represent zero difference. The black solid lines represent mean differences (MD) and 95% limits of agreements (LoA).

Table 1.

Representative sequence parameters for free-breathing 3D stack-of-radial (FB Radial) and breath-holding (BH) 3D Cartesian axial MRI scans at 3T. N/A: not applicable.

Sequence parameters	FB Radial	BH Cartesian
TE (ms)	1.23, 2.46, 3.69, 4.92, 6.15, 7.38	
TR (ms)	8.85	
Flip angle (°)	5	5
Field of view	$360–440 \times 360–440$ mm^2	$360–440 \times 360–440$ mm^2
Slice thickness (mm)	5	5
Matrix size (x, y, z)	224–288, 224–288, 40–72	224–288, 224–288, 30–40
Acceleration factor	N/A	R=4 (parallel imaging)
Radial spokes	354–454	N/A
Scan time (min:sec)	2:28 – 4:49 [*]	0:19 [#]
Retrospective undersampling	R=2.5 (40% self-gating acceptance rate)	N/A

[*] Radial gradient calibration time was not included.

[#] Prescan calibration time was not included.

Table 2.

Dataset characteristics. NAFLD: non-alcoholic fatty liver disease. Std: standard deviation.

Datasets	Total number of subjects	Adult subjects	Pediatric subjects	Total number of 2D slices	Range of liver PDFF values
Training set	63	24 NAFLD, 11 healthy	18 NAFLD[#], 10 healthy	2528[*]	Min: 0.4% Max: 33.4% Mean: 10.5% Std: 9.7%
Validation set	21	7 NAFLD, 4 healthy	6 NAFLD[#], 4 healthy	812	Min: 0.9% Max: 28.4% Mean: 9.2% Std: 8.8%
Testing set	21	7 NAFLD, 4 healthy	5 NAFLD[#], 5 healthy	860	Min: 0.6% Max: 25.2% Mean: 9.8% Std: 8.4%

[*] Before performing data augmentation. See text in section 2.3 for details about data augmentation.

[#] Suspected or confirmed NAFLD.

Table 3.

Ablation study in the testing dataset for different components used in UP-Net. Structural similarity index (SSIM) and normalized root mean squared error (NRMSE) were evaluated on magnitude images, with respect to compressed sensing results. Proton-density fat fraction (PDFF) and R_2^* quantification errors were evaluated in liver regions of interest, with respect to compressed sensing and graph-cut fitting results. Results are reported as mean ± standard deviation.

Network	Component					Metric			
	Phase Augmentation	GAN Loss	Physics Loss	Uncertainty Estimation	Joint Training	SSIM	NRMSE	PDFF Errors	R_2^* Errors
1		✓	✓	✓	✓	0.851±0.055[*]	0.182±0.048[*]	−0.92% ±0.95%[*]	−0.54s^{-1}±3.02s^{-1}[*]
2	✓		✓	✓	✓	0.858±0.067[*]	0.194±0.053[*]	−0.31% ±1.05%	−0.68s^{-1}±3.46s^{-1}[*]
3	✓	✓		✓	✓	0.870±0.049	0.178±0.050	−1.69% ±1.49%[*]	−2.50s^{-1}±5.03s^{-1}[*]
4	✓	✓	✓		✓	0.877±0.048	0.176±0.042	−0.29% ±0.88%	−0.34s^{-1}±3.81s^{-1}
5	✓	✓	✓	✓		0.884±0.050	0.168±0.061	−0.46% ±1.47%[*]	−0.41s^{-1}±3.02s^{-1}
UP-Net	✓	✓	✓	✓	✓	0.872±0.053	0.173±0.059	−0.36% ±0.98%	−0.37s^{-1}±3.56s^{-1}

[*] represents statistically significant difference (p<0.01, Wilcoxon signed-rank test) compared with UP-Net.