Published in final edited form as:

Magn Reson Med. 2018 November ; 80(5): 2139–2154. doi:10.1002/mrm.27178.

Super-Resolution Musculoskeletal MRI Using Deep Learning

Akshay S. Chaudhari, PhD^{1,2,*†}, Zhongnan Fang, PhD^{3,†}, Feliks Kogan, PhD¹, Jeff Wood, MD¹, Kathryn J Stevens, MD^{1,4}, Eric K. Gibbons, PhD⁵, Jin Hyung Lee, PhD^{2,3,6,7}, Garry E. Gold, MD^{1,2,4}, and Brian A. Hargreaves, PhD^{1,2,7}

¹Department of Radiology, Stanford University, Stanford, California, USA

²Department of Bioengineering, Stanford University, Stanford, California, USA

³LVIS Corporation, Palo Alto, California, USA

⁴Department of Orthopaedic Surgery, Stanford University, Stanford, California, USA

⁵Department of Radiology and Imaging Sciences, University of Utah, Salt Lake City, Utah, USA

⁶Department of Neurology and Neurological Sciences, Stanford University, Stanford, California, USA

⁷Department of Electrical Engineering, Stanford University, Stanford, California, USA

Abstract

Purpose—To develop a super-resolution technique using convolutional neural networks for generating thin-slice knee MR images from thicker input slices, and to compare this method to alternative through-plane interpolation methods.

Methods—We implemented a 3D convolutional neural network entitled DeepResolve to learn residual-based transformations between high-resolution thin-slice images and lower-resolution thick-slice images at the same center locations. DeepResolve was trained using 124 double-echo in steady-state (DESS) datasets with 0.7mm slice thickness and tested on 17 patients. Ground-truth images were compared to DeepResolve, clinically utilized tricubic interpolation (TCI) and Fourier interpolation (FI) methods, along with state-of-the-art single image sparse-coding super-resolution (ScSR). Comparisons were performed using structural similarity (SSIM), peak signal-to-noise ratio (pSNR), and root-mean-square-errors (RMSE) image quality metrics for a multitude of thin-slice downsampling factors (DSFs). Two musculoskeletal radiologists ranked the three datasets and reviewed the diagnostic quality of the DeepResolve, TCI, and ground-truth images for sharpness, contrast, artifacts, signal-to-noise ratio, and overall diagnostic quality. Mann-Whitney U-Tests evaluated differences between the quantitative image metrics, reader scores, and rankings. Cohen's Kappa (κ) evaluated inter-reader reliability.

Results—DeepResolve had significantly better SSIM, pSNR, and RMSE than TCI, FI, and ScSR for all DSFs ($P < 0.05$, except 4x and 8x ScSR DSFs). In the reader study, DeepResolve

*Correspondence To: Akshay Chaudhari, Ph.D., Department of Radiology, Stanford University, The Lucas Center for Imaging, 1201 Welch Road PS055-B, Stanford, CA, 94305, USA, Phone: (714) 204-5705, akshaysc@stanford.edu.

†These authors contributed equally to this work

Disclosure:

Akshay Chaudhari has provided consulting services and is a part-time employee of Skope MR Inc.

significantly outperformed ($P<0.01$) TCI in all image quality categories and overall image ranking. Both readers had substantial scoring agreement ($\kappa=0.73$).

Conclusion—DeepResolve was capable of resolving high-resolution thin-slice knee MRI from lower-resolution thicker slices, achieving superior quantitative and qualitative diagnostic performance to both conventionally utilized and state-of-the-art methods.

Keywords

Super-resolution; deep-learning; isotropic MRI; interpolation; unsupervised sparsity learning; musculoskeletal MRI

Introduction

Clinical musculoskeletal imaging protocols typically include 2D fast-spin-echo (FSE) sequences scanned in various scan planes, often with the same contrast. While such sequences provide excellent in-plane resolution, there is always an associated risk for missing subtle lesions due to partial volume effects in slices with a high section thickness (usually 2.5-4mm) and with slice gaps. Image acquisitions with thick slices also preclude image reformations into arbitrary scan planes, which makes it challenging to interrogate tissues with oblique orientations, such as the anterior cruciate ligament and the articular cartilage of the femoral trochlea for example.

Several methods have been proposed recently for thin-section musculoskeletal MRI, with 3D fast spin-echo (using vendor products SPACE, CUBE, or VISTA) being a popular choice (1–5). However, 3D FSE sequences typically compromise in-plane resolution for through-plane resolution and suffer from substantial echo train blurring despite the use of variable flip angle methods (2,6,7). Consequently, despite acquiring thin-section images, the overall diagnostic quality of 3D FSE remains limited. 3D radial sequences have been used for isotropic musculoskeletal imaging but have not seen widespread adoption likely due to a different artifact appearance (8,9). Besides sequence-based approaches, compressed sensing (CS) and parallel imaging are also promising methods for accelerating thin-section MRI, however, neither is an ideal stand-alone method (2,10,11). CS MRI is uncommon in 2D FSE clinical protocols because only one dimension can be used to generate incoherent sampling, in addition to long reconstruction times. Parallel imaging acceleration is ultimately limited by SNR and g-factor (12). Additionally, the geometries of common knee coils (such as rigid 8-channel) are ill-suited for acceleration in all directions.

Single image super-resolution, an active field in image processing, can potentially boost MRI spatial resolution to generate thin-section MRI without compromising SNR or requiring additional MRI hardware or scan time (13,14). Interpolation, for example, is one naive implementation of super-resolution. With the practical acquisition challenges for thin-section imaging, MRI vendors offer retrospective interpolation of slices using zero-padded Fourier interpolations (FI) through options such as ‘ZIP2’ and ‘Interpolate’ for GE Healthcare and Siemens Healthineers MRI scanners, respectively. Similarly, medical image viewing platforms such as OsiriX use trilinear interpolation during image manipulation and multi-planar reformation of data (15). Low-order linear interpolation methods such as FI and

trilinear interpolation are widely used in clinical and research protocols for attempting to achieve thin-section imaging, however, neither produce images with high diagnostic quality (16,17).

Besides interpolation, another primary method of MRI super-resolution imaging entails utilizing image sparsity with only single images required as inputs (18–20). The current state-of-the-art MRI single-image super-resolution algorithm is based on sparse-coding super-resolution (ScSR), which was initially developed for natural images, but later adapted to a handful of MRI applications (21–23). While promising, this super-resolution method has not been pervasive in medical imaging due to limited resolution improvements and slow execution speeds for 3D data. Recent advancements in high-performance computing hardware, deep-learning, and convolutional neural networks (CNNs) may help generalize super-resolution techniques used with natural 2D images for the application to 3D medical images (24–26). While the fundamental concepts of deep learning have existed for decades, improvements in computational power have spurred widespread usage (27,28). For example, now even a shallow three-layer CNN has been shown to outperform the state-of-the-art ScSR method for 2D natural images (24). Thus, there is potential that deep-learning super-resolution will enable superior performance for generating thin-section MRI compared to commonly utilized techniques, in clinically feasible algorithm execution times.

In this pilot study, we evaluate the feasibility of implementing deep-learning MRI super-resolution for generating thin-slice images and maintaining high in-plane resolution to reduce overall scan time. We term our proposed method ‘*DeepResolve*’ as it helps resolve high-resolution features from low-resolution inputs. Specifically, we train DeepResolve using publicly available datasets to generate high-resolution thin-slice knee MR images from slices at the same locations but with 2-8 times higher slice thickness. The goal of DeepResolve is not necessarily to generate identical images to the ground-truth; it is not yet clear that this is possible. Rather, we aim to enhance low-resolution images to make them more similar to the ground-truth as compared to commonly-utilized and state-of-the-art methods. We quantitatively compare the various resolution enhancement methods using image quality metrics and qualitatively compare the methods through a reader study to evaluate the diagnostic potential of DeepResolve.

Methods

DeepResolve Residual Learning

Given a set of i ground-truth high-resolution thin-slice images X_h^i and corresponding low-resolution thick-slice images interpolated to the thin-slice locations X_l^i , with the same field of view and matrix size, the DeepResolve network is a function $f = f(X_l^i)$ that calculates a residual difference image r^i between the ground-truth and interpolated low-resolution images ($r^i = X_h^i - X_l^i$). This function is trained by optimizing (with an L2-loss)

$$\operatorname{argmin}_f \|r^i - f(X_l^i)\|_2^2. \quad (1)$$

DeepResolve (Fig. 1a) is an unsupervised learning algorithm modeled as a cascade of convolutional filters. It is inspired by previous work on residual-based very deep super-resolution networks (VDSR), but unlike VDSR (which utilizes 2D convolutions), DeepResolve utilizes 3D convolutions that can provide additional spatial information to improve super-resolution performance (29–31).

After the DeepResolve network has been trained, the estimated residual $f(X_l^i)$ can be calculated directly and a high-resolution super-resolution image $\widetilde{X}_{sr}^l$ can be generated through

$$\widetilde{X}_{sr}^l = X_l^i + f(X_l^i). \quad (2)$$

Imaging Datasets

All data for this study were acquired from the Osteoarthritis Initiative (OAI), which is a longitudinal study for studying osteoarthritis progression (32). A total of 176 3D sagittal double-echo in steady-state (DESS) datasets from the OAI were utilized for DeepResolve experiments (relevant scan parameters: FOV=14cm, Matrix=384×307 (zero-filled to 384×384), TE/TR=5/16ms, 160 slices with a thickness of 0.7mm) (32). All patients had dual-knee weight-bearing radiographs performed to evaluate their Kellgren-Lawrence (KL) osteoarthritis grade (33). KL grades roughly correspond to no OA (KLG-0), minimal OA (KL-1), mild OA (KL-2), moderate OA (KL-3), and severe OA (KL-4). The 176 patients were split in 124 patients for training, 35 for validation, and 17 for testing (approximately 70%, 20%, 10%) while maintaining an equal distribution of KL grades in each group as shown in Fig. 1c-e.

Data Formatting

The magnitude-only DESS DICOM datasets were cropped from a size of 384×384×160 to 344×344×160 to only include relevant anatomy. Each dataset was further divided into 32×32×32-pixel patches with a stride of 16×16×16, yielding a total of 3,600 patches per patient and approximately 500,000 total training patches. Each image volume is divided by its maximum value and normalized from 0 to 1 prior to sub-dividing into patches in order to maintain relative image contrast between patches. The ratio of the ground-truth slice thickness and the downsampled low-resolution slice thickness was termed as downsampling factor (DSF). Thick-slice representation of the high-resolution thin-slice images was generated by anti-aliasing filtering and consequent downsampling of the thin-slice images in the left-right direction. A 1D finite impulse response, low-pass, 48th-order Hamming-windowed filter with a normalized pass-band of 1/DSF was generated in MATLAB (MathWorks, Natick, MA) for the retrospective downsampling (filter code provided in

supporting information). This simulated acquisition of thicker slices that have magnetization contributions from surrounding slices. The training dataset consequently consisted of the ground-truth high-resolution images and the simulated low-resolution images that were upscaled using 1D tricubic interpolation (TCI) in the slice direction (left-right) at the ground-truth slice locations (workflow described in Fig. 2). Separate single-scale DeepResolve networks were trained for DSFs of 1x, 2x, 3x, 4x, 6x, and 8x. A 1x DSF network was trained to evaluate whether DeepResolve introduces artifacts absent in the input images. We also implemented a multi-scale network that was trained with patches with mixed downsampling factors.

Network Implementation

We implemented DeepResolve using a cascade of convolutional filters paired with non-linear activations functions. The input images ($32 \times 32 \times 32$ pixel patches) were convolved with 3-dimensional filters with dimensions of $3 \times 3 \times 3$, stride of $1 \times 1 \times 1$, and a filter feature map length of 64. Consequently, in the first DeepResolve layer, the input patch was transformed into feature maps with dimensions of $32 \times 32 \times 32 \times 64$. Subsequent layers maintained identical dimensions except in the final layer where feature reduction was performed to generate a residual image of dimensions $32 \times 32 \times 32$. All convolutions were performed on a zero-padded input and the output was cropped to the size of the original input in order to maintain identical input and output dimensions.

We utilized rectified linear unit (ReLU) non-linear activation functions of the form $R(x) = \max(0, x)$ for ensuring that the feature map outputs were non-linear representations of the input (34). The input low-resolution patch was transformed through a cascade of convolutional and ReLU pairs, except for the final layer that generated the residual image of size $32 \times 32 \times 32$. The final (20th) convolutional layer did not include a ReLU activation to ensure that the residual image could include positive and negative values. The residual image was added to the input image to approximate the high-resolution image and a mean-square-error L2-loss was compared to the original high-resolution image.

DeepResolve training was performed on 124 datasets using stochastic gradient descent using Keras and a Tensorflow backend (Google, Mountain View, CA) (35,36). Hyper-parameter optimization was performed by minimizing L2-loss on the 35 validation datasets with a learning rate of 0.0001, while maximizing the number of batches (50) into the GPU memory. The training was performed over 20 epochs by optimizing convolutional filter weights using 4 GeForce GTX Titan 1080 Ti GPUs (NVIDIA, Santa Clara, CA). All 3D convolutional layers were randomly initialized with a 'He' initialization (37). Testing was performed on 17 datasets (dimensions $344 \times 344 \times 160$) that were normalized with the same method as the training datasets. Due to limited GPU memory, each dataset divided into eight blocks of dimensions $172 \times 172 \times 80$ without overlap and these blocks were subsequently processed with DeepResolve and combined to create a full-sized dataset.

Resolution Enhancement Methods Comparison

To compare different methods of enhancing through-plane resolution, we generated images with common interpolation and state-of-the art super-resolution methods. In addition to the

TCI image inputs to DeepResolve, we also generated FI images from the simulated thick-slice images (38,39). For comparing against a state-of-the-art single MR image super-resolution method, we generated sparse-coding super-resolution (ScSR) images for the same $344 \times 344 \times 160$ imaging volume as DeepResolve (23). ScSR creates sparse residual images using a 2D patch-based dictionary approach that *iteratively* tries to enhance low-resolution features based on pairs of low-resolution and high-resolution image pairs (detailed description in supporting information). The ScSR parameters used were based on the original open-source MATLAB code, with revisions to account for 3D backprojection (23).

Overall, for each high-resolution slice from the ground-truth dataset, there were corresponding resolution-enhanced slices generated with DeepResolve, TCI, FI, and ScSR for all DSFs. Due to an image acquisition time of 11 minutes, the ground-truth DESS datasets are not practical for routine clinical acquisition. Thus, the goal of DeepResolve was not to replicate ground-truth images, but to enhance low-resolution images better than currently-utilized and state-of-the-art methods.

Quantitative Image Comparison

Quantitative image quality comparisons between the ground-truth images and all resolution enhancement methods for all DSFs were performed using root-mean-square-error (RMSE), peak signal-to-noise ratio (pSNR), and structural similarity (SSIM) over the full 3D volume. Similar images have lower RMSEs, higher pSNRs, and higher SSIM (bounded between 0 and 1) values. RMSE provides a measure between pixel-intensity-wise variations between two images being tested. pSNR evaluates the ratio between the maximum power of an input image and the power of features that distort the image. Unlike RMSE and pSNR that only calculate pixel-wise variations, SSIM can evaluate perceptual image quality (40).

Simulations

To evaluate the impact of DeepResolve on subtle pathologies, three pairs of artificial meniscal and cartilage lesions with varying signal intensity were simulated in a new DESS dataset with 3x DSF. Slicer3D was used to manually create lesions that were subtle spatially and in signal intensity (41). Quantitative SSIM, pSNR, and MSE values in a local 32×32 window around the lesions were calculated for DeepResolve and TCI, with respect to the ground-truth.

Reader Study

To assess diagnostic image quality, two board-certified musculoskeletal radiologists (with 17 and 2 years of experience respectively) concurrently assessed the ground-truth images along with TCI and DeepResolve images with 3x DSF for the 17 testing datasets. The readers loaded all three image-sets simultaneously using the '2D Viewer' in OsiriX (Pixmeo SARL, Bernex, Switzerland) and scrolled through all sagittal slices and the axial and coronal reformations. The readers were blinded to image type and the image ordering was randomized for each patient. Image quality metrics of sharpness, contrast, artifacts, signal-to-noise ratio, and overall quality, of each image set were scored on a five-point quality scale (1=non-diagnostic, 2=limited, 3=diagnostic, 4=good, 5=excellent). The readers also ranked the three image sets against one another (1=best, 3=worst).

Statistical Analysis

Notched-box plots were used to visualize the RMSE, pSNR, and SSIM variations between the ground-truth images compared with the DeepResolve, TCI, FI, and ScSR images for all DSFs. Mann-Whitney U-tests ($\alpha=0.05$) evaluated whether there were systematic differences between the DeepResolve metrics and metrics for the TCI, FI, and ScSR images. For the reader study, one-sided Mann-Whitney U-tests ($\alpha=0.05$) evaluated whether ground-truth scores were better than DeepResolve and TCI scores, and whether DeepResolve scores were better than TCI scores. A linearly-weighted Cohen's kappa (κ) and its confidence interval (CI) were calculated to evaluate inter-reader variability for the image quality ratings and the image rankings (42,43). All numerical analysis and data pre-processing was performed using MATLAB (MathWorks, Natick, MA).

Results

DeepResolve Characterization

The total duration required to train each single-scale and multi-scale DeepResolve network was 115 hours, which corresponded to approximately 6 hours per epoch. Network convergence was observed around epochs 10-12. The inference time required to apply the network model was approximately 11 seconds per each 3D dataset. Single-scale networks had slightly better quantitative outcomes than the multi-scale network and were used in the remainder of the study (additional information in supporting Fig. S1 and S2).

Example coronal images with varying DSFs and their SSIM maps (Fig. 3) indicated that as the DSF was increased, fine details of the medial collateral ligament (MCL) and the anterior cruciate ligament (ACL) were blurred out (arrows on SSIM maps). Moreover, the remnant femoral bone marrow signal due to imperfect fat-suppression, appeared overly-smoothed, which decreased the local SSIM. Sagittal, coronal, and axial images from one patient reconstructed with DeepResolve for all DSFs utilized in this study can be seen in supporting figure S3. The sagittal images (Fig. S3a-g) were least affected by blurring as the DSF increased, while the axial and coronal reformations (Fig. S3h-u) showed higher blurring with increasing DSFs.

Resolution Enhancement Methods Comparison

A comparison of the DeepResolve, TCI, FI, and ScSR methods for a DSF of 3x, and the ground-truth images, can be seen in Figure 5. DeepResolve maintained fine features in the coronal reformation and was most comparable in terms of qualitative visual quality and quantitative structural similarity measures to the original ground-truth image. High-resolution features such as the MCL (solid arrow), a small osteophyte on the lateral tibial plateau (dashed arrow), and inflammation with sharp features (dotted arrow) were easily visualized on the DeepResolve image, however, visualization was far more challenging with the other methods (Fig. 4a-e). This was further indicated through the difference maps (Fig. 4g-j) and pixel-by-pixel structural similarity maps (Fig. 4k-n) where unlike the DeepResolve method, TCI, FI, and ScSR all had lower similarity around the medial collateral ligament.

Comparing the DeepResolve-generated residuals with the actual residuals (between the ground-truth and TCI) showed that DeepResolve generated the most high-resolution details (Fig. 5a-e). Most of the differences in the residuals were either in the background or in the bone marrow, where the fat-suppression was imperfect. The corresponding k-space (Fig. 5f-h) and average projections of the k-space frequencies in the superior-inferior (SI) direction (Fig. 5i) demonstrated that DeepResolve resolved additional higher-resolution features than TCI.

Quantitative Image Comparison

A comparison of image quality metrics for all resolution enhancement methods to the original ground-truth images demonstrated the superior performance of DeepResolve (Fig. 6). SSIM for DeepResolve images was significantly higher ($P<0.0005$) than all other methods for all DSFs, indicating that the DeepResolve method maintained optimal perceptual quality compared to ground-truth images. DeepResolve also significantly ($P<0.001$) outperformed the other methods in the metrics of pSNR and RMSE for DSF's, except for ScSR images with DSFs of 4x and 6x (P -values of 0.08 and 0.11 respectively). These results demonstrate that DeepResolve quantifiably enhanced image quality for a multitude of metrics compared to other resolution enhancement methods.

Simulations

Sagittal views of the simulated meniscal and cartilage lesions demonstrated that DeepResolve could resolve the very subtle lesions moderately well, compared to the ground-truth (Fig. 7). DeepResolve had better SSIM, pSNR, and MSE values in the local 32×32 window compared to the TCI reconstructions. Coronal images of the lesion and the quantitative metrics, which showed a similar trend, are included in supporting figure S4.

Reader Study

In the reader study (Fig. 8), DeepResolve images were rated significantly better ($P<0.01$) than TCI images for all image quality metrics assessed. Comparing the ground-truth images and the DeepResolve showed that the ground-truth images were significantly better ($P<0.05$) for the image assessment metrics of sharpness, SNR, and overall image quality. All DeepResolve metrics indicated diagnostic image quality (or better) since the average reader scores exceeded 3, which was not the case for TCI images (dotted line, Fig. 8).

The image ranking for the three datasets paralleled the image quality assessments. The ground-truth images were on average ranked the best (score 1.2 ± 0.2), while DeepResolve images were neutral (1.9 ± 0.5) and the TCI images were the worst (2.9 ± 0.4). There was substantial agreement between the two readers for the image quality assessments ($\kappa=0.73$, 0.64-1.00 CI) and excellent agreement for image rankings ($\kappa=0.83$, 0.45-0.99 CI) (42).

Derangement Examples

An example of a meniscus tear of the lateral body for all three sets of images showed that the sagittal TCI image exhibited more blurring than the DeepResolve image, however, the coronal TCI reformation was considerably blurrier than DeepResolve (Fig. 9). Contours of fine structures such as the cartilage, meniscus, and the joint inflammation in the

DeepResolve images appeared more comparable to the ground-truth images. In contrast to the gross meniscal tear, a very subtle finding of a grade 2A lateral patellar chondromalacia (according to a modified Noyes scale) appeared similarly on all three sagittal image-sets, but the axial DeepResolve reformations had better image quality than TCI (Fig. 10) (44,45). The fine contours of the cartilage and the chondral signal heterogeneity were adequately depicted on the ground-truth image but appeared blurrier on DeepResolve and TCI images. However, DeepResolve did maintain higher image quality than TCI.

Discussion

In this study, we implemented a convolutional neural network entitled DeepResolve, which was capable of transforming low-resolution thick-slice knee MRI slices into high-resolution thin-slice images. In a comparison with commonly-used and state-of-the-art super-resolution methods, DeepResolve was found to have the best performance for all downsampling factors tested using quantitative image quality metrics. Additionally, in a reader study with two musculoskeletal radiologists, DeepResolve images with a downsampling factor of 3x were able to maintain diagnostic image quality and significantly outperformed tricubic interpolation. DeepResolve was able to resolve high-resolution image features for a dataset consisting of 160 low-resolution slices in only 10 seconds. Acquiring the ground-truth images used in this study that had low-section thickness and high in-plane resolution is impractical in clinically feasible scan times. In such cases, DeepResolve can enhance the quality of high section-thickness images and easily outperforms interpolation schemes utilized commonly on MR scanners and image viewers.

Quantitative Image Comparison

Based on image similarity metrics such as SSIM, pSNR, and RMSE, DeepResolve demonstrated the superiority of a deep neural network to perform 3D MRI super-resolution, compared to other resolution enhancement techniques. The increase in SSIM was of particular importance since it has been shown to compare images not just on pixel-by-pixel variations, but rather, using perceptual quality metrics derived from the human visual system. However, despite the superior performance and the ease of implementing DeepResolve, it is important to analyze *why* the network functions well.

Methods such as TCI or FI inherently treat high-resolution features to be resolved as low-order representations of the low-resolution features already present. While an acceptable assumption in natural images, it may not be valid in medical images due to the complex morphology, textures, and low SNR of tissues. Additionally, such interpolation methods consistently perform the interpolation over a specified length scale - normally a few pixels surrounding the pixel being interpolated. This behavior may also not be ideal for medical images, especially in musculoskeletal MRI, since different tissues exist with features over varying spatial lengths. Consequently, nearest-neighbor interpolations will fail to account for morphological heterogeneity. Moreover, such naïve interpolation methods do not rely on any priors to perform the interpolation. As a result, there is no way to differentiate whether any given pixel truly represents the anatomy or whether that pixel is affected by noise, or artifacts such as motion or aliasing.

DeepResolve was also able to outperform the state-of-the-art ScSR method. ScSR can be viewed as a single-layer shallow CNN which performs patch-based convolutions on $N \times N$ windows to create a single-layer of weights undergoing non-linear activations (25). This is conceptually similar to DeepResolve, but since ScSR effectively utilizes only one layer of abstraction, it can only discern features of one length scale (24). ScSR also implements an iterative approach to perform the mapping between high-resolution and low-resolution, which is not as efficient as the backpropagation approach utilized in DeepResolve. Additionally, ScSR is based on linear sparse decomposition methods, but DeepResolve induces higher system non-linearities because of ReLU activations, which may help describe its superior performance. However, it should be noted that the solution of DeepResolve is not guaranteed to exist within the null-space of the downsampling operator during the testing. To overcome this, implementing a 3D backprojection (as performed in ScSR) may enhance data fidelity by ensuring data consistency between the predicted super resolution image and the under-sampled input. While no obvious artifacts were observed along the edges of the eight testing patches used to create the full-sized image (as seen in supporting Fig S5), an additional backprojection may also help mitigate the discontinuity created by DeepResolve at patch edges.

Reader Study

DeepResolve performed consistently better than TCI, showing that the DeepResolve method has potential for diagnostic use. While the overall reader scores for ground-truth images were the best, it is impractical to acquire such 11-minute high-resolution images in clinical settings. However, the DeepResolve scores indicated that DeepResolve is a time-efficient method and a significant advancement over commonly utilized interpolation methods for enhancing through-plane resolution. This was well demonstrated with the fact that a healthy MCL usually has sub-millimeter thickness in the coronal plane and DeepResolve was able to maintain fine MCL detail in an acquisition that simulated a slice thickness of 2.1mm. One concern regarding applying deep learning to medical images is its impact on very subtle features that may indicate pathology. In the example of the fine chondromalacia, while DeepResolve did not have similar image quality as the ground-truth image, it demonstrated irregular contours and signal heterogeneity better than the TCI image. Since DeepResolve may indicate both gross and subtle pathologic changes, perhaps not as well as the ground-truth images but considerably better than TCI images, applying DeepResolve to research and clinical operations is promising.

DeepResolve was able to generate images with 'diagnostic quality' or above (average reader score above 3), which can be useful for interrogating other fine musculoskeletal structures such as the demonstrated meniscal tears and fine cartilage lesions. DeepResolve may be promising for clinical use for acquiring slices at 2-3mm slice thicknesses and subsequently transforming them into thinner slices for multi-planar reformations. Such methods could also be especially useful for newer implementations of DESS that enable simultaneous T_2 relaxometry, morphometry, and semi-quantitative radiological assessment (48). The thicker slices could be used for generating high-SNR for quantitative T_2 measurements, while the thin slices could be used for accurate morphometry and semi-quantitative whole-joint

assessment (49). Bilateral knee imaging methods that acquire several hundred slices could also benefit from DeepResolve (50).

The training data for DeepResolve consisted of 34% of patients with Kellgren-Lawrence OA grade of 2 and 59% of patients of grade 3. As a result, the training was exposed to healthy and degenerate features of the knee. The reader study may have benefitted from a comparison of all resolution enhancement methods, however, only the ground-truth, DeepResolve, and TCI images were chosen so as to prevent reader fatigue. TCI was chosen over FI since TCI generated perceivably better images. ScSR images were excluded because although they achieved moderate resolution enhancement, their slow reconstruction speeds were impractical for routine use. Similarly, only a DSF of 3x was chosen for the reader study as an intermediate benchmark for the foundation for additional studies.

Network Characterization

A small convolution filter size of $3 \times 3 \times 3$ was chosen in DeepResolve in order to avoid averaging together high frequency image and convolutional details. It has been shown that filters with larger supports can be effectively decomposed into several smaller filters, introducing unnecessary redundancy during the network training process (51). As a result, a small convolution kernel efficiently maximized high-resolution feature extraction. This is the same reason a pooling layer was not utilized, since it inherently averages together high-resolution details for memory optimization. All feature maps were also zero-padded prior to convolutions and cropped to the original input-size following the convolutions in order to maintain equivalent patch-input and convolution-output sizes to ensure that the details at the edges of patches would still maintain fidelity. Moreover, a patch overlap of 50% of the patch size was used during the network training phase while the testing data was divided into 8 large patches without overlap to mitigate network effects caused at patch edges.

The DeepResolve network depth of 20 was far deeper than many other neural networks used in image classification and super-resolution. This was chosen with regard to the convolution filter and image patch sizes used in this experiment. For example, after the first $3 \times 3 \times 3$ convolution, the receptive field for each weight in the second layer is $3 \times 3 \times 3$. After the second $3 \times 3 \times 3$ convolution, the total receptive field for the next weight is $5 \times 5 \times 5$, which can be generalized to $(2D+1)^3$, where D is the current network depth. Thus, for a patch size of $32 \times 32 \times 32$, the receptive field for a weight in the 16th layer corresponds to features encountered in the *entire* image patch. Additional deeper layers can provide even higher-level abstractions. Since super-resolution is by nature an ill-posed problem, providing spatial cues at varying length-scales may help algorithm convergence.

The DeepResolve network depth of 20 was chosen based on the initial 2D super-resolution implementation while the patch size of $32 \times 32 \times 32$ was chosen to incorporate the entire patch in the convolutional receptive field (29). The batch size of 50 was consequently chosen to maximize the number of patches that could fit into memory. The high-dimensionality of the network and hyperparameter space will likely require careful optimizations in future studies, however. Previous research has demonstrated that increasing the depth of the CNN may improve super-resolution quality (29). Moreover, deep-learning models have empirically

shown to follow a power-law loss relationship, which may help determining ideal training data size (52).

Unlike PI and CS techniques that could be simulated on numerical phantoms, making such comparisons with CNNs may not be feasible. CNNs are extensively trained on a specific anatomy with specific contours, contrasts, and noise-distributions, which makes it challenging to apply them to different anatomies or phantoms due to the varied feature space. This is why we simulated artificial features that had a subtle structure as well as subtle signal in DESS images, rather than using a numerical phantom. In future scenarios, transfer learning methodologies could be used to increase generalizability to different anatomies. Transfer learning utilizes knowledge from a pre-trained network to fine-tune a different, but related, network, and it may also allow training with smaller dataset.

DeepResolve Practicality

In deep-learning methods, most of the computational burden is undertaken during training-time in order to improve test-time performance. Compared to popular networks such as VGG-19 and AlexNet that train 60 million and 144 million parameters respectively, DeepResolve only requires 2 million parameters for training (53,54). Such a compact network enabled efficient training and full-3D reconstruction (dimensions $344 \times 344 \times 160$) in only 10 seconds. Comparatively, ScSR, which enabled moderate resolution enhancement, required approximately three hours of reconstruction time due to its iterative behavior, which is not pragmatic for routine use. TCI and FI required short computation times, but they did not produce images with diagnostic image quality. Despite the inferior performance of these methods, FI is commonly used on MRI scanners while TCI is commonly used on PACS viewers. DeepResolve addresses the limitations described above by generating super-resolution images with high image fidelity and diagnostic quality in short implementation times, which may be practical for routine clinical and research applications.

In this study, DeepResolve was only trained for 20 epochs to prevent network overfitting. While training over additional epochs may decrease training loss, it may also restrict the network to only learn features of the training data, which can reduce generalizability. For future studies, inclusion of dropout and sparsity-enforced norms on network weights may help increase generalizability (54,55). DeepResolve could also be generalized to in-plane resolution enhancements in addition to through-plane resolution enhancements. It is also not specific to musculoskeletal MRI or DESS contrasts only. Such a post-processing-based method can easily be combined with traditional parallel imaging and compressed sensing (PI+CS) schemes as well as novel deep-learning-based reconstruction methods for further accelerating MR imaging to jointly optimize all acceleration paradigms such as PI+CS (56). In addition, since DeepResolve is an unsupervised learning solution, it does not require labeled data, which overcomes one of the largest challenges of implementing machine learning solutions.

Limitations

Despite the promising early results of DeepResolve super-resolution, there were certain limitations to this study that should be addressed in future studies. An L2 loss function was

chosen for DeepResolve. Additional loss functions such as the L1-norm, pSNR, and SSIM could be explored. While DeepResolve in this study was performed on magnitude data, utilizing complex and/or multi-channel data may result in higher fidelity outputs by imposing mutually-dependent constraints on the complex and coil data channels. The reader study undertaken in this manuscript only assessed image quality as a first-pass for whether DeepResolve images may serve diagnostic value. Follow-up studies systematically demonstrating the relative ability of DeepResolve to identify internal derangement of the knee will be essential in determining the overall efficacy of the method, especially in a larger cohort of symptomatic patients.

Conclusion

In this study, we have demonstrated DeepResolve – a deep-learning-based super-resolution method capable of resolving high-resolution thin-slice features from slices originally considerably thicker. DeepResolve expectedly did not match the image quality of the high-resolution ground-truth images but it outperformed other resolution enhancement methods, through quantitative image quality metrics and a qualitative reader study. DeepResolve shows potential in both research and clinical settings, to reduce image acquisition time, imaging costs, and motion-artifacts by acquiring fast low-resolution images and retrospectively increasing image resolution.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grant Support:

Contract grant sponsor: National Institutes of Health (NIH); contract grant numbers NIH R01 AR063643, R01 EB002524, K24 AR062068, and P41 EB015891. Contract grant sponsor: GE Healthcare (research support). Image data was acquired from the Osteoarthritis Initiative (OAI). The OAI is a public-private partnership comprised of five contracts (N01-AR-2-2258; N01-AR-2-2259; N01-AR-2-2260; N01-AR-2-2261; N01-AR-2-2262) funded by the National Institutes of Health, a branch of the Department of Health and Human Services, and conducted by the OAI Study Investigators. Private funding partners include Merck Research Laboratories; Novartis Pharmaceuticals Corporation, GlaxoSmithKline; and Pfizer, Inc. Private sector funding for the OAI is managed by the Foundation for the National Institutes of Health. This manuscript was prepared using an OAI public use data set and does not necessarily reflect the opinions or views of the OAI investigators, the NIH, or the private funding partners.

References

1. Kijowski R, Davis KW, Woods MA, Lindstrom MJ, De Smet AA, Gold GE, Busse RF. Knee joint: comprehensive assessment with 3D isotropic resolution fast spin-echo MR imaging—diagnostic performance compared with that of conventional MR imaging at 3.0 T. *Radiology*. 2009; 252:486–495. DOI: 10.1148/radiol.2523090028 [PubMed: 19703886]
2. Kijowski R, Rosas H, Samsonov A, King K, Peters R, Liu F. Knee imaging: Rapid three-dimensional fast spin-echo using compressed sensing. *J Magn Reson Imaging*. 2017; 45:1712–1722. DOI: 10.1002/jmri.25507 [PubMed: 27726244]
3. Bao S, Tamir JJ, Young JL, Tariq U, Uecker M, Lai P, Chen W, Lustig M, Vasanawala SS. Fast comprehensive single-sequence four-dimensional pediatric knee MRI with T2 shuffling. *J Magn Reson Imaging*. 2016; :1–12. DOI: 10.1002/jmri.25508

4. Gold GE, Busse RF, Beehler C, Han E, Brau AC, Beatty PJ, Beaulieu CF. Isotropic MRI of the knee with 3D fast spin-echo extended echo-train acquisition (XETA): initial experience. *AJR Am J Roentgenol.* 2007; 188:1287–1293. DOI: 10.2214/AJR.06.1208 [PubMed: 17449772]
5. Fritz J, Fritz B, Thawait GG, Meyer H, Gilson WD, Raithel E. Three-Dimensional CAIPIRINHA SPACE TSE for 5-Minute High-Resolution MRI of the Knee. *Invest Radiol.* 2016; 0:1.doi: 10.1097/RLL.0000000000000287
6. Busse RF, Hariharan H, Vu A, Brittain JH. Fast spin echo sequences with very long echo trains: design of variable refocusing flip angle schedules and generation of clinical T2 contrast. *Magn Reson Med.* 2006; 55:1030–1037. DOI: 10.1002/mrm.20863 [PubMed: 16598719]
7. Kijowski R, Davis KW, Woods MA, Lindstrom MJ, De Smet AA, Gold GE, Busse RF. Knee joint: comprehensive assessment with 3D isotropic resolution fast spin-echo MR imaging—diagnostic performance compared with that of conventional MR imaging at 3.0 T. *Radiology.* 2009; 252:486–495. DOI: 10.1148/radiol.2523090028 [PubMed: 19703886]
8. Kijowski R, Blankenbaker DG, Klaers JL, Shinki K, De Smet AA, Block WF. Vastly undersampled isotropic projection steady-state free precession imaging of the knee: diagnostic performance compared with conventional MR. *Radiology.* 2009; 251:185–194. DOI: 10.1148/radiol.2511081133 [PubMed: 19221057]
9. Chaudhari AS, Sveinsson B, Moran CJ, McWalter EJ, Johnson EM, Zhang T, Gold GE, Hargreaves BA. Imaging and T2 relaxometry of short-T2 connective tissues in the knee using ultrashort echo-time double-echo steady-state (UTEDESS). *Magn Reson Med.* 2017; 78:2136–2148. DOI: 10.1002/mrm.26577 [PubMed: 28074498]
10. Stevens KJ, Busse RF, Han E, Brau ACS, Beatty PJ, Beaulieu CF, Gold GE. Ankle: Isotropic MR Imaging with 3D-FSE-Cube—Initial Experience in Healthy Volunteers. *Radiology.* 2008; 249:1026–1033. DOI: 10.1148/radiol.2493080227 [PubMed: 19011194]
11. Kijowski R, Gold GE. Routine 3D magnetic resonance imaging of joints. *J Magn Reson Imaging.* 2011; 33:758–771. DOI: 10.1002/jmri.22342 [PubMed: 21448939]
12. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: Sensitivity encoding for fast MRI. *Magn Reson Med.* 1999; 42:952–962. DOI: 10.1002/(SICI)1522-2594(199911)42:5<952::AID-MRM16>3.0.CO;2-S [PubMed: 10542355]
13. Plenge E, Poot DHJ, Bernsen M, Kotek G, Houston G, Wielopolski P, Van Der Weerd L, Niessen WJ, Meijering E. Super-resolution methods in MRI: Can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time? *Magn Reson Med.* 2012; 68:1983–1993. DOI: 10.1002/mrm.24187 [PubMed: 22298247]
14. Van Reeth E, Tham IWK, Tan CH, Poh CL. Super-resolution in magnetic resonance imaging: A review. *Concepts Magn Reson Part A.* 2012; 40:306–325.
15. Rosset A, Spadola L, Ratib O. OsiriX: An open-source software for navigating in multidimensional DICOM images. *J Digit Imaging.* 2004; 17:205–216. DOI: 10.1007/s10278-004-1014-6 [PubMed: 15534753]
16. Dyrby TB, Lundell H, Burke MW, Reisle NL, Paulson OB, Pito M, Siebner HR. Interpolation of diffusion weighted imaging datasets. *Neuroimage.* 2014; 103:202–213. DOI: 10.1016/j.neuroimage.2014.09.005 [PubMed: 25219332]
17. Van Reeth E, Tan CH, Tham IWK, Poh CL. Isotropic reconstruction of a 4-D MRI thoracic sequence using super-resolution. *Magn Reson Med.* 2015; 73:784–793. DOI: 10.1002/mrm.25157 [PubMed: 24478231]
18. Park SC, Park MK, Kang MG. Super-resolution image reconstruction: A technical overview. *IEEE Signal Process Mag.* 2003; 20:21–36. DOI: 10.1109/MSP.2003.1203207
19. Wang Y-H, Qiao J, Li J-B, Fu P, Chu S-C, Roddick JF. Sparse representation-based MRI super-resolution reconstruction. *Measurement.* 2014; 47:946–953.
20. Plenge E, Poot DHJ, Bernsen M, Kotek G, Houston G, Wielopolski P, van der Weerd L, Niessen WJ, Meijering E. Super-resolution methods in MRI: Can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time? *Magn Reson Med.* 2012; 68:1983–1993. [PubMed: 22298247]
21. McDonagh S, Hou B, Kamnitsas K, Oktay O, Alansary A, Rutherford M, Hajnal JV, Kainz B. Context-Sensitive Super-Resolution for Fast Fetal Magnetic Resonance Imaging. 2017:1–11.

22. Ota J, Umehara K, Ishimaru N, Ishida T. Application of Sparse-Coding Super-Resolution to 16-Bit DICOM Images for Improving the Image Resolution in MRI. *Open J Med Imaging*. 2017; 7:144.
23. Wang YH, Qiao J, Li JB, Fu P, Chu SC, Roddick JF. Sparse representation-based MRI super-resolution reconstruction. *Meas J Int Meas Confed*. 2014; 47:946–953. DOI: 10.1016/j.measurement.2013.10.026
24. Dong C, Loy CC, He K, Tang X. Learning a Deep Convolutional Network for Image Super-Resolution. *Comput Vision–ECCV 2014*. 2014; 8689:184–199. DOI: 10.1007/978-3-319-10593-2_13
25. Dong C, Loy CC, He K, Tang X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans Pattern Anal Mach Intell*. 2016; 38:295–307. DOI: 10.1109/TPAMI.2015.2439281 [PubMed: 26761735]
26. Lim B, Son S, Kim H, Nah S, Lee KM. Enhanced Deep Residual Networks for Single Image Super-Resolution. 2017; doi: 10.1109/CVPRW.2017.151
27. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, Jackel LD. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput*. 1989; 1:541–551. DOI: 10.1162/neco.1989.1.4.541
28. LeCun Y, Boser BE, Denker JS, Henderson D, Howard RE, Hubbard WE, Jackel LD. Handwritten Digit Recognition with a Back-Propagation Network. In: Touretzky DS, editor *Advances in Neural Information Processing Systems 2*. Morgan-Kaufmann; 1990. 396–404.
29. Kim J, Lee JK, Lee KM. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. *Cvpr*. 2016; 2016:1646–1654. DOI: 10.1109/TPAMI.2015.2439281
30. Wang Z, Liu D, Yang J, Han W, Huang T. Deep Networks for Image Super-Resolution with Sparse Prior. 2015; doi: 10.1109/ICCV.2015.50
31. Han YS, Yoo J, Ye JC. Deep Learning with Domain Adaptation for Accelerated Projection Reconstruction MR. 2017
32. Peterfy CG, Schneider E, Nevitt M. The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee. *Osteoarthr Cartil*. 2008; 16:1433–1441. DOI: 10.1016/j.joca.2008.06.016 [PubMed: 18786841]
33. Kellgren JH, Lawrence JS. Osteo-arthritis and disk degeneration in an urban population. *Ann Rheum Dis*. 1958; 17:388–397. DOI: 10.1136/ard.17.4.388 [PubMed: 13606727]
34. Nair V, Hinton GE. Rectified Linear Units Improve Restricted Boltzmann Machines. *Proc 27th Int Conf Mach Learn*. 2010:807–814. doi: 10.1.1.165.6419.
35. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. 2014. 1–15. <http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503>.
36. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE*. 1998; 86:2278–2323. DOI: 10.1109/5.726791
37. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proc IEEE Int Conf Comput Vis*. 2015; :1026–1034. 2015 Inter: DOI: 10.1109/ICCV.2015.123
38. Bernstein MA, Fain SB, Riederer SJ. Effect of windowing and zero-filled reconstruction of MRI data on spatial resolution and acquisition strategy. *J Magn Reson Imaging*. 2001; 14:270–280. DOI: 10.1002/jmri.1183 [PubMed: 11536404]
39. Du YP, Parker DL, Davis WL, Cao G. Reduction of partial volume artifacts with zero filled interpolation in three dimensional MR angiography. *J Magn Reson Imaging*. 1994; 4:733–741. DOI: 10.1002/jmri.1880040517 [PubMed: 7981519]
40. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Wavelets for Image Image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Process*. 2004; 13:600–612. DOI: 10.1109/TIP.2003.819861 [PubMed: 15376593]
41. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging*. 2012; 30:1323–41. DOI: 10.1016/j.mri.2012.05.001 [PubMed: 22770690]
42. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–74. [PubMed: 843571]

43. Cohen J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull.* 1968; 70:213–220. DOI: 10.1037/h0026256 [PubMed: 19673146]
44. Noyes FR, Stabler CL. A system for grading articular cartilage lesions at arthroscopy. *Am J Sports Med.* 1989; 17:505–513. DOI: 10.1177/036354658901700410 [PubMed: 2675649]
45. Duc SR, Pfirrmann CWA, Koch PP, Zanetti M, Hodler J. Internal knee derangement assessed with 3-minute three-dimensional isovoxel true FISP MR sequence: preliminary study. *Radiology.* 2008; 246:526–35. DOI: 10.1148/radiol.2462062092 [PubMed: 18227545]
46. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative Adversarial Networks. 2014; :1–9. DOI: 10.1001/jamainternmed.2016.8245
47. Ledig C, Theis L, Huszar F, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. 2016; doi: 10.1109/CVPR.2017.19
48. Chaudhari AS, Black MS, Eijgenraam S, Wirth W, Maschek S, Sveinsson B, Eckstein F, Oei EHG, Gold GE, Hargreaves BA. Five-minute knee MRI for simultaneous morphometry and T2 relaxometry of cartilage and meniscus and for semiquantitative radiological assessment using double-echo in steady-state at 3T. *J Magn Reson Imaging.* 2017; doi: 10.1002/jmri.25883
49. Sveinsson B, Chaudhari A, Gold G, Hargreaves B. A simple analytic method for estimating T2 in the knee from DESS. *Magn Reson Imaging.* 2017; 38:63–70. DOI: 10.1016/j.mri.2016.12.018 [PubMed: 28017730]
50. Kogan F, Levine EG, Chaudhari AS, Monu UD, Epperson K, Oei EHG, Gold GE, Hargreaves BA. Simultaneous Bilateral-Knee MR Imaging. *Magn Reson Med.* 2017; doi: 10.1002/mrm.27045
51. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2014; :1–14. DOI: 10.1016/j.infsof.2008.09.005
52. Hestness J, Narang S, Ardalani N, Diamos G, Jun H, Kianinejad H, Patwary MMA, Yang Y, Zhou Y. Deep Learning Scaling is Predictable, Empirically. 2017;1–19. doi: 1712.00409.
53. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014 arXiv Prepr. arXiv1409.1556.
54. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks; *Adv Neural Inf Process Syst.* 2012. 1–9. .
55. Zhu Y, Bradic J. Breaking the curse of dimensionality in regression. 2017; 18:1–53.
56. Hammernik K, Klatzer T, Kobler E, Recht MP, Sodickson DK, Pock T, Knoll F. Learning a Variational Network for Reconstruction of Accelerated MRI Data. 2017; 0:1–17. DOI: 10.1002/mrm.26977

Figure 1.

The DeepResolve network computes a residual image from an input low-resolution (LR) image in order to generate the corresponding high-resolution (HR) image (a). For illustrative purposes, in-plane residual variations between the LR and HR images are highlighted, however, in the study, most residuals existed in the through-plane (left-right) direction. DeepResolve consists of 20 layers of paired convolution and rectified linear unit (ReLU) non-linear activation blocks (b). The first and last layer of DeepResolve has dimensions of $32 \times 32 \times 32$ while all other layers have dimensions of $32 \times 32 \times 32 \times 64$. The final layer did not include an activation function to ensure that the generated residual included positive and negative values. Each input 3D dataset was divided into isotropic $32 \times 32 \times 32$ pixel patches and all patches from all patients were used during the DeepResolve training process (c). A

breakdown of the Kellgren-Lawrence (KL) osteoarthritis grade of the patients utilized in training (d), validation (e), and testing (f) can be seen.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 2.

An overall schematic of the DeepResolve experiment can be seen here. High-resolution ground-truth slices (a) are used to simulate acquisition of slices with a higher section thickness (b) for varying downsampling factors (DSFs). These slices are subsequently tricubically interpolated to the ground-truth slice locations (c) and the DeepResolve training identifies a residual function (d) that ideally can be added to the low-resolution thick-slices to produce the high-resolution thin-slices (e). During inference, a residual image is created for the testing input using the learned residual model. Consequently, the residual can be added to the low-resolution input in order to output a super-resolution image.

Figure 3.

Coronal DeepResolve images with varying downsampling factors (DSFs) (a-f) are more closely analyzed here along with their corresponding 2D structural similarity (SSIM) maps (g-l), compared to the ground-truth slice. While images with a higher DSF appear over-smoothed, the directionality of the blurring is also more apparent. In the femoral bone marrow, pixels with imperfect fat-saturation appear more blurred in the left-right direction, than in the inferior-superior direction (dotted arrow). The medial collateral ligament is an ideal tissue to illustrate image fidelity in the coronal reformation since it is very thin in the left-right direction (solid arrow). The SSIM maps clearly illustrate that the MCL is not reproduced with high diagnostic quality for DSFs of 6 and 8.

Figure 4.

An example coronal ground-truth image (a) and the resolution-enhanced images with a downsampling factor (DSF) of 3x at the same slice location using for (b) DeepResolve, (c) Fourier interpolation (FI), (d) tricubic interpolation (TCI), and (e) sparse-coding super-resolution (ScSR). The DeepResolve-generated residual map (f) as well difference images (scaled by a factor of 5 to accentuate subtle differences) between the comparison methods and the ground-truth image are shown (g-j). The corresponding 2D structural similarity (SSIM) maps for the resolution-enhanced images compared to the ground-truth are also shown (k-n). DeepResolve shows the best resemblance to the ground-truth image. Features such as the medial collateral ligament (solid arrow), a small osteophyte on the lateral tibial plateau (dashed arrow), and inflammation with sharp features (dotted arrow) can be easily visualized on the MDSR image, however, visualization is far more challenging on the TCI, FI, and ScSR images. These same regions have a lower structural similarity with TCI, FI, and ScSR compared to DeepResolve. Abbreviations: Diff. Maps = Difference Maps.

Figure 5.

Example coronal ground-truth, DeepResolve, and TCI images are shown here (a-c) in addition to the generated DeepResolve residual map (d) in comparison to the actual residual between the ground-truth and TCI (e). The corresponding k-space representations in the superior-inferior (SI) and left-right (LR) directions of the three ground-truth, DeepResolve, and TCI images are also shown (f-h). Performing a simple average of the k-space energies in the SI direction (i) demonstrates that DeepResolve was able to recover additional higher-resolution information than TCI images. The TCI coronal image had a DSF of 3x, which corresponds to approximately one-third of the full-width half-max of the k-space averaging as compared to the ground-truth.

Figure 6.

All four resolution enhancement methods (DeepResolve, TCI, FI, and ScSR) were compared to the original ground-truth images for quantitative image similarity metrics using (a) structural similarity (SSIM), (b) peak signal-to-noise ratio (pSNR), and (c) root-mean-square error (RMSE). Mann-Whitney U tests compared whether the DeepResolve metrics were different than those for TCI, FI, and ScSR. 'X' indicates the outliers. Statistical significance is indicated on the x-axes with ** $P < 0.0005$ for all comparisons and * $P < 0.001$ for all comparisons, except ScSR for a downsampling factor of 4 ($P = 0.08$) and 8 ($P = 0.11$).

Figure 8.

Two radiologists qualitatively assessed the diagnostic quality of the ground-truth, DeepResolve, and TCI images (blinded and randomized to scan type) for categories of contrast, sharpness, signal-to-noise ratio, artifacts, and overall image quality on a 1-5 scale (1=non-diagnostic, 2=limited, 3=diagnostic, 4=good, 5=excellent). All images had a DSF of 3x. Mann-Whitney U tests assessed whether the ground-truth images scores were different than the DeepResolve and TCI scores, and whether DeepResolve scores was different than TCI scores. All average DeepResolve image metric scores were of diagnostic quality (dotted horizontal line) or better. * $P < 0.05$.

Figure 9.

An example of a horizontal tear in the body of the lateral meniscus can be identified with the hyperintense DESS signal. The tear (arrow) can be seen relatively similarly in the sagittal ground-truth, DeepResolve, and TCI images (a-c). However, there is some blurring evident in the posterior femoral cartilage which overestimated the cartilage thickness. The coronal reformation also indicates the same meniscal tear (d-f). A zoomed section of the coronal image (yellow bracket) shows that the DeepResolve image appears more smoothed and less noisy than the ground-truth image (g-i). Comparing the TCI image to both the ground-truth and DeepResolve images shows considerably more blurring in the TCI image. The border of the inflammation (green arrow) in the TCI image does not have the same contours as the ground-truth while the central cartilage has a stair-like jagged appearance (dotted arrow) instead of smooth edges.

Figure 10.

An example of a subtle case of grade 2A chondromalacia of the lateral patellar cartilage was identified with a hypointense DESS signal. The cartilage (indicated with arrows) appears similar on all three sagittal images. However, the axial reformations demonstrate the slightly higher fidelity of the DeepResolve images compared to the TCI images (a-c). The joint effusion (dotted arrows) on the axial images around the lateral femoral condyle appears with a very jagged appearance on the TCI images compared to the ground-truth and DeepResolve (d-f). The zoomed section of the axial reformations shows contours of the patellar cartilage and the heterogeneous signal maintained better in the DeepResolve images than TCI images, but not as well as the ground-truth images (g-i).