Super-Resolution [1]H Magnetic Resonance Spectroscopic Imaging Utilizing Deep Learning

Zohaib Iqbal[1], Dan Nguyen[1], Gilbert Hangel[2], Stanislav Motyka[2], Wolfgang Bogner[2] and Steve Jiang[1]*

[1] Medical Artificial Intelligence and Automation Laboratory, Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX, United States, [2] Christian Doppler Laboratory for Clinical Molecular MR Imaging, Department of Biomedical Imaging and Image-guided Therapy, High Field MR Center, Medical University of Vienna, Vienna, Austria

Magnetic resonance spectroscopic imaging (SI) is a unique imaging technique that provides biochemical information from *in vivo* tissues. The ^{1}H spectra acquired from several spatial regions are quantified to yield metabolite concentrations reflective of tissue metabolism. However, since these metabolites are found in tissues at very low concentrations, SI is often acquired with limited spatial resolution. In this work, we test the hypothesis that deep learning is able to upscale low resolution SI, together with the T1-weighted (T1w) image, to reconstruct high resolution SI. We report on a novel densely connected UNet (D-UNet) architecture capable of producing super-resolution spectroscopic images. The inputs for the D-UNet are the T1w image and the low resolution SI image while the output is the high resolution SI. The results of the D-UNet are compared both qualitatively and quantitatively to simulated and *in vivo* high resolution SI. It is found that this deep learning approach can produce high quality spectroscopic images and reconstruct entire ^{1}H spectra from low resolution acquisitions, which can greatly advance the current SI workflow.

Keywords: super-resolution, magnetic resonance spectroscopic imaging (SI), deep learning (DL), magnetic resonance spectroscopy (^{1}H MRS), artificial intelligence

1. INTRODUCTION

Magnetic resonance imaging (MRI) continues to be a versatile modality capable of providing anatomical, metabolic, and functional information from various regions of the body *in vivo*. In particular, magnetic resonance spectroscopic imaging (SI) (1) is able to yield important data regarding the metabolism of different tissues, and has been especially useful for studying the metabolism of the human brain (2). Some important biochemicals, or metabolites, in the brain include N-acetyl aspartate (NAA), glutamate (Glu), glutamine (Gln), creatine (Cr), choline (Ch), and myo-Inositol (mI) (3). Each metabolite plays an important role in regulating energy consumption in the brain, and some metabolites also play critical functional roles, including roles as neurotransmitters (4). It is well-known that metabolic changes occur in parallel with anatomical changes for a myriad of pathologies (2), and these metabolic changes may even occur before structural changes are detected. While SI has continued to be an active area of research over the past several decades, there are still major roadblocks into standardizing this technique and including it into clinical protocols.

One of the major disadvantages of SI is the long acquisition duration associated with obtaining spectra from several voxels of interest. This is primarily due to the fact that many of the important metabolites are found in the brain at low concentrations; these metabolites are typically present in the brain at 1–12 mM concentrations (3). Therefore, in order to accurately detect these biochemicals, several signal averages have to be obtained or larger voxel volumes have to be acquired to improve the signal to noise ratio (SNR) for the experiment. As a result, spatial resolution tends to be coarse for many SI sequences. This low resolution, coupled with other technical problems such as partial volume effects, hinders the overall diagnostic capabilities of the SI technique.

There have been many advances in the technological implementation of SI that allow for faster acquisition and better spatial resolution. One of the primary acceleration methods is echo planar spectroscopic imaging (EPSI) (5, 6), which collects spectral data from an entire line of k-space in a single repetition time (TR) utilizing an echo planar readout. This spatio-spectral acquisition approach has also been applied in non-cartesian SI methods, such as spiral acquisitions (7), concentric circular acquisitions (8), and rosette acquisitions (9). In addition, parallel imaging (10–12) can also be used to accelerate the collection of SI data. Sensitivity encoding (SENSE) has been applied in combination with EPSI (13) to facilitate even faster acquisition times. Recently, research has also focused on the application of various sampling schemes that allow for reduced scan time (14–18). Some studies (19, 20) have even demonstrated protocols capable of obtaining spectroscopic images at 64x64 or 128x128 resolution in less than 20 min. Although these advances have improved the field significantly, SI is still understandably seen as a low SNR, low resolution technique.

In order to combat the limits of the experimentally acquired resolution, many post-processing methods have been developed for super-resolution SI (21–27). These methods have mainly focused on model-based reconstruction methods and regularized reconstruction approaches. While many super-resolution methods are independent of the acquisition protocols, there are some techniques, such as the spectroscopic imaging by exploiting spatio-spectral correlation (SPICE) method (18), that show reconstruction benefits by employing inter-dependent sequences. Unfortunately, the majority of super-resolution methods either tend to be very complicated to implement, or generally show poor reconstruction results. Since experimental acquisitions have many technical challenges, there is also a large concern over the true gold standard for these super-resolution techniques. Without a true standard of comparison, which is a large problem in the spectroscopic imaging field, many studies qualitatively and quantitatively compare their methods with less ideal standards such as bicubic interpolation.

Deep learning is an advancing field that has shown extraordinary results for image processing (28–30). Convolutional layers and networks are capable of extracting valuable features from images, and can further process these features into labels or other images for classification, segmentation, and other uses. One network that has been extremely beneficial for the field of automated medical imaging segmentation is the UNet (31), which allows for a pixel-wise transformation of an input image into an output image. Essentially, deep learning excels at computing an unknown transformation by using a large example dataset, often referred to as a training set. We hypothesize that UNet, or some other deep neural networks are able to upscale low resolution SI (LRSI), together with the T1-weighted (T1w) image, to produce high resolution SI (HRSI). To test this hypothesis the biggest challenge is that a large, publicly available SI dataset is unavailable and difficult to acquire experimentally. In order to create this data set, HRSI (128x128 pixels) and LRSI (16x16 pixels or some other low resolution) experiments would have to be performed on thousands of diverse patients with different pathologies, which is not feasible. Thus, it is seemingly impossible to perform deep learning for super-resolution SI.

In this paper, we report a novel work on the development of a deep learning technology capable of producing super-resolution spectroscopic images. An SI generator is used to produce LRSI and HRSI data in order to train and test a deep learning model. Using this data, a UNet taking advantage of densely connected layers (D-UNet) is built and trained. The inputs for the D-UNet are the T1w image and the low resolution SI image while the output is the high resolution SI. The results of the D-UNet reconstruction are compared both qualitatively and quantitatively to simulated and *in vivo* high resolution SI data.

2. METHODS

2.1. Spectroscopic Imaging Dataset

Two different MRI data sets were utilized to produce synthetic SI data for developing the deep learning model. The first MRI data set comprised of 27 axial slices from the MATLAB MRI dataset. MR images as well as white matter (WM) and gray matter (GM) masks from the open access series of imaging studies (OASIS) project (32), which contained 416 axial images from subjects ranging in age from 18 to 96 years old, were also used. From these limited data, 102,169 SI datasets were synthesized using an SI generator, the details of which are found below.

2.2. Spectroscopic Imaging Generator

The SI generator was designed to address the lack of T1w images, as well as the lack of paired LRSI and HRSI data. First, the generator created augmented T1w ($aT1w$), white matter (WM), and gray matter (GM) images from an input T1w image. Then, the generator would produce a matched LRSI and HRSI for the $aT1w$ image.

2.2.1. Augmenting T1w Images

An input T1w image is first segmented into WM and GM images via an intensity based approach. First, the maximum WM intensity (WM_{max}), and the minimum GM intensity (GM_{min}) are determined from the image. Then, WM and GM images are made by applying the following:

$$WM = \left(\frac{S - GM_{min}}{WM_{max}} \right) \cdot M \qquad (1)$$

$$GM = \left(1 - WM \right) \cdot M \qquad (2)$$

Above, S is the original signal intensity of the input T1w image, and M is a mask for the brain region only, and is applied through an element-wise multiplication. The above equations ensure that the elements of both the WM and GM images range from zero to one, and are representative of the percentage of WM or GM present in any voxel.

Then, the SI generator modifies the input T1w image to produce an $aT1w$ image. The contrast of the T1w image is altered by the following:

$$aT1w = R\left(S_n^{r_1} + L\right) \qquad (3)$$

Here, S_n is the normalized input T1w signal and r_1 is a random number between 0.5 and 2.5. $R()$ is a rotation and field of view (FOV) truncation transformation that rotates the image randomly in the range of $-15°$ to $15°$ and randomly truncates the image in the range of 0 to 40 pixels in any direction. L is a matrix that represents up to 6 lesions of varying intensity, location, and size. Since this lesion matrix is random, the $aT1w$ image may or may not contain any hyper-intense or hypo-intense regions. The same transformation used in Equation (3) is also applied to the WM and GM images.

2.2.2. Production of Matched LRSI and HRSI Maps

In order to produce data useful for clinical applications, the SI generator operated under an assumption that is biologically valid: WM and GM regions of the brain have metabolism associated with biochemical concentrations (33). With this assumption, a given metabolite could be more concentrated in WM vs. GM, less concentrated in WM vs. GM, or equally concentrated in WM and GM regions.

Working with this biological assumption, a high resolution metabolite map is generated by adding a random ratio of the WM and GM images together:

$$HRSI = r_2 * WM + (1 - r_2) * GM + B + r_3 * L \qquad (4)$$

In Equation (4), r_2 is a random number between 0 and 1. B is a matrix that adds a random signal bias into the metabolite map, which helps to simulate the presence of more metabolite signal from the anterior or posterior, as well as the left or right brain regions. L is the same lesion matrix used in Equation (3), and r_3 is a random number between -1 and 1.

Finally, the HRSI is downsampled to the desired low resolution via k-space truncation. Random noise is also added to this low resolution k-space data before a Fourier transformation is used to bring this data back to the spatial domain. Next, the low resolution image is upscaled to the same resolution as the HRSI using nearest-neighbor interpolation to yield the final low resolution SI.

It is important to note that because of the variables r_1, r_2, r_3, and L, it is possible to produce several different matched $aT1w$ images, HRSI, and LRSI from the same input T1w image. In addition, the same $aT1w$ image can give rise to a large number of matched HRSI and LRSI, and thus this transformation is a one to many transformation. Therefore, a single input T1w image can produce hundreds of unique datasets for training a deep learning model.

2.3. Densely Connected UNet (D-UNet) Architecture and Training

The UNet architecture (31) is typically implemented for segmentation purposes, however it primarily operates by performing pixel-wise transformations on input images, which is applicable to the SI super-resolution problem. Using standard convolutional and max pooling layers, the UNet first continuously convolves and pools the input image until the image reaches a small size, which aids in extracting valuable global features. Next the image is scaled up through a combination of up-pooling, transpose convolutions, and feature concatenations. This second process helps to identify vital local features so that the UNet can refine the image at a finer resolution. However, due to the number of features necessary for this process, the classical UNet suffers from extremely long training times, overfitting issues, and potential inefficiencies when tuning the weights. Therefore, this study utilized densely connected convolutional layers (34) to develop the novel densely connected UNet (D-UNet) architecture, and the workflow for training is shown in **Figure 1**. Densely connected networks carry over features from layer to layer, allowing for all previous information to be used for determining important features. The general architecture of the D-UNet used in this study is shown in **Figure 2**. The D-UNet utilized 32 feature maps at every max pooling layer. In addition, all convolutional layers made use of the ReLU activation function (30) and used a dropout (35) of 0.1. Certain features, shown in green and orange in **Figure 2**, were copied over to the following layers, and were also concatenated later on in the network. In total, three max pooling layers were used for the D-UNet. Since low resolution SI experiments can have diverse resolutions, three identical D-UNets were made to upscale low resolution spectroscopic images for acquisitions with 16x16, 24x24, and 32x32 spatial points.

The D-UNet required two inputs: a rescaled (128x128 points) T1w image and the corresponding LRSI image (16x16, 24x24, or 32x32 points) upscaled using nearest-neighbor interpolation (128x128 points). The predicted output of the D-UNet was a denoised HRSI image (128x128 points). For training, $aT1w$, HRSI, and LRSI were created from the SI generator, as described above. The Adam optimizer (36) was used with a learning rate set to 1×10^{-3}, and mean squared error (MSE) was used as the cost function, which determined the difference between the D-UNet output and the desired output:

$$MSE = \sum \sum \frac{(O - HRSI)^2}{m^2} \qquad (5)$$

Above, O is the output of the D-UNet, $HRSI$ is the true simulated high resolution SI, and m is the output dimension of the network, which in this case is 128. The summations are performed over both dimensions to yield a single value. The network was trained on an 8GB Quadro K5200 graphical processing unit (GPU) using the Keras (37) and Tensorflow (38) packages in Python 3.6.

Two datasets were made for the development and evaluation of the three D-UNets: a training dataset and a testing dataset. The training dataset comprised of 102,000 data from the SI generator using 135 axial images. The testing dataset used 169 different
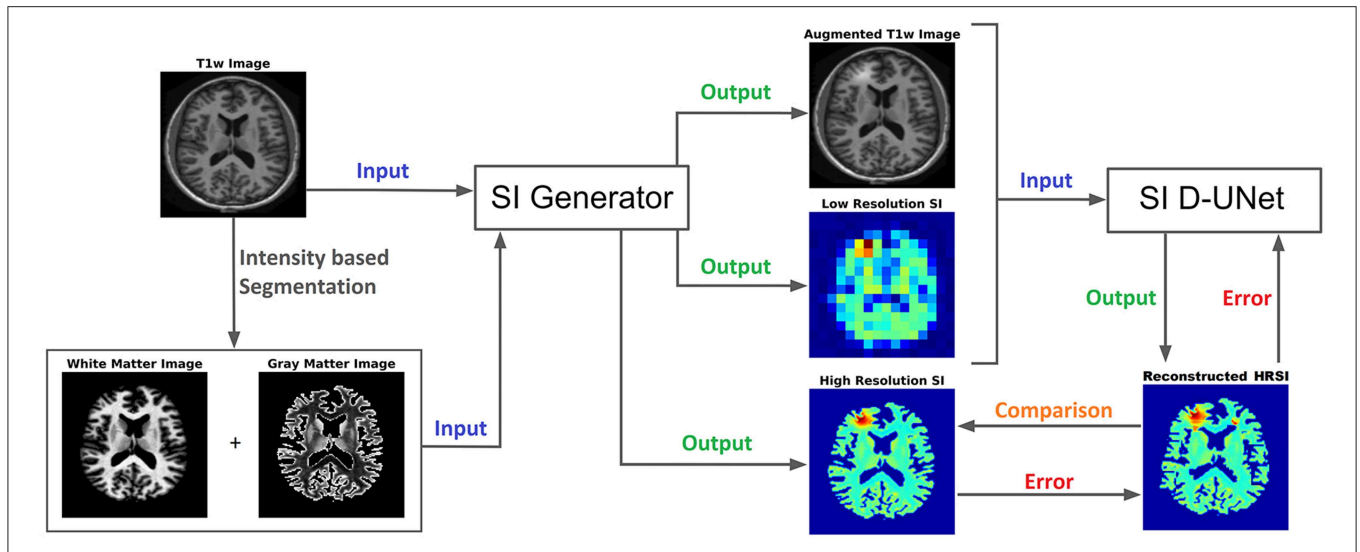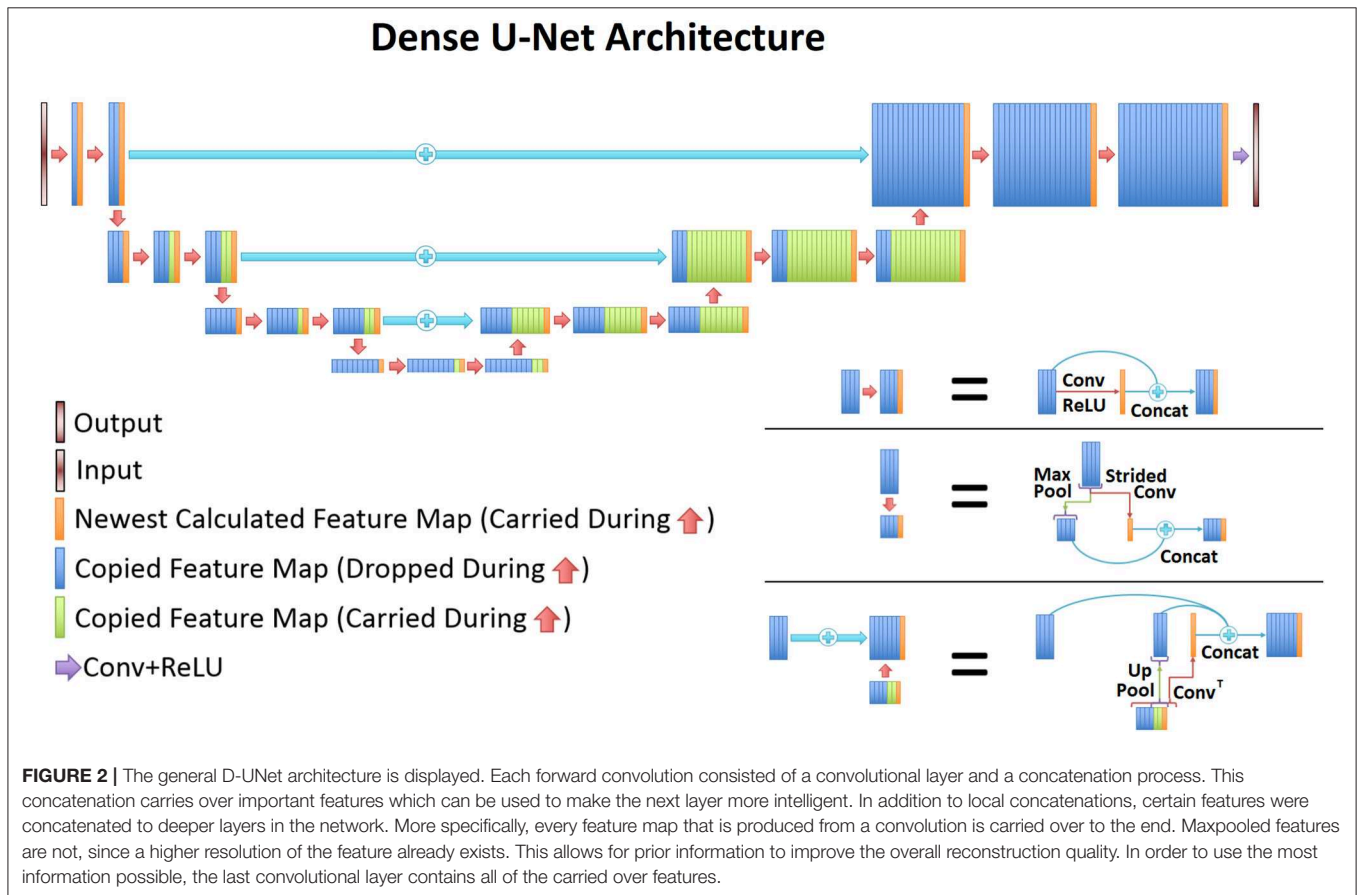
FIGURE 1 | The workflow for training the D-UNet model is shown. The SI generator provides a dataset consisting of an augmented T1w image, a low resolution spectroscopic image, and a ground truth high resolution spectroscopic image. The spectroscopic images already show the distribution of a particular metabolite (or the distribution of a particular spectral point), such as choline, and therefore do not contain a spectral dimension. Then, the network transforms the $aT1w$ (128x128 pixels) and LRSI (128x128 pixels after nearest-neighbor interpolation) into an initial HRSI reconstruction (128x128 pixels). In the example above, the LRSI and HRSI reconstruction have in-plane spatial resolutions of 1.4×1.4 cm^2 and 1.7×1.7 mm^2, respectively. This reconstruction is compared to the ground truth, and the mean squared error is calculated. Utilizing this error, the model changes the weighting parameters for the features, and continues training by using a different dataset. After training on 102,000 datasets, the model weights are refined and the reconstruction errors are minimized.

FIGURE 2 | The general D-UNet architecture is displayed. Each forward convolution consisted of a convolutional layer and a concatenation process. This concatenation carries over important features which can be used to make the next layer more intelligent. In addition to local concatenations, certain features were concatenated to deeper layers in the network. More specifically, every feature map that is produced from a convolution is carried over to the end. Maxpooled features are not, since a higher resolution of the feature already exists. This allows for prior information to improve the overall reconstruction quality. In order to use the most information possible, the last convolutional layer contains all of the carried over features.

axial images (independent from the training set) from the OASIS project, and 169 matched $aT1w$, HRSI, and LRSI images were produced via the SI generator. Each of the three D-UNets were trained for a total of 102 epochs. For this study, an epoch was defined as a new set of 1,000 matched HRSI and LRSI data. The first two epochs were trained using a batch size of one to ensure that the network would not fall into a local minimum. The remaining 100 epochs were trained with a batch size of 10. Varying batch size in this manner has been shown to help reduce the number of epochs necessary for training, while also reducing the need for hyper-parameter tuning (39).

2.4. D-UNet Evaluation and Comparison Metrics

2.4.1. Testing Set Evaluation

The three D-UNets evaluated all 169 matched images ($aT1w$ and LRSI) to produce reconstructed high resolution spectroscopic images ($Recon_{16x16}$, $Recon_{24x24}$, and $Recon_{32x32}$). These reconstructed images were compared to the ground truth HRSI using mean squared error. This process was repeated with varying noise levels inserted into the input LRSI in order to determine the role of noise on the reconstruction process. Example low resolution spectroscopic images can be seen in **Figure 3**. The reconstructed images were also compared to zero-filling and bicubic interpolation to assess the improvement of the D-UNet results over standard methods. For this comparison, both zero-filling and bicubic interpolation were applied to an LRSI of 32x32 points to generate the 128x128 interpolated images.

2.4.2. Spectral Reconstruction Evaluation

In addition, the three D-UNets were used to reconstruct magnitude spectra point-by-point from low spatial resolution to high spatial resolution. Magnitude spectra were used because the model was not trained for evaluating real and imaginary numbers simultaneously. From the test set, a single subject was used to generate high resolution chemical maps of the major metabolites, including NAA, Glu, Gln, Cr, Ch, and mI. GAMMA simulation (40) was used to simulate the spectra for these metabolites using an echo time (TE) = 30 ms, spectral bandwidth of 2,000 Hz, and time points = 512 for a magnetic field strength (B_0) of 3T. Also, the spectra were exponentially line broadened to roughly 8 Hz. These spectra were then distributed spatially based on their respective high resolution maps, and were transformed to produce LRSI. The T1w image and LRSI were input into the three D-UNets to produce $Recon_{16x16}$, $Recon_{24x24}$, and $Recon_{32x32}$ spectral data. Two example spectra were extracted from these reconstructed images and compared to the simulated ground truth using mean squared error.

2.4.3. *In vivo* Evaluation

Finally, high resolution spectroscopic images were acquired on a 7T whole-body MR scanner (Magnetom, Siemens Healthcare, Erlangen, Germany) using a previously published protocol (20). The Institutional Review Board (IRB) at the Medical University of Vienna approved the study and ten healthy volunteers (mean age = 31.7 years old) signed written and informed consent

forms. All experiments were performed in accordance with relevant guidelines and regulations. The protocol utilized free induction decay based MR spectroscopic imaging (41) with TR = 200 ms for a total scan time of 21 min. After acquisition, residual lipids were removed using ℓ_2 regularization (42) and the spectra were quantified using the LCModel (43) package to yield concentrations for several metabolites. Therefore, high resolution (128x128 pixels, 1.7×1.7 mm^2) metabolite maps for NAA, Cr, Ch, Glu, Gln, and mI were obtained. These metabolite maps were down-sampled to 32x32 resolution images and were input into the 32x32 D-UNet along with corresponding T1w images to yield $Recon_{32x32}$ for all datasets. These reconstructed images were then compared to the experimentally acquired HRSI using mean squared error as described in Equation (5). In addition, Glu/Cr and Ch/Cr ratios for both the reconstructed and experimentally acquired images were measured over all ten subjects. These ratios were investigated as a function of T1w intensity, which directly corresponds to the ratio of WM and GM in the brain. Finally, correlations between the reconstructed and experimental results were performed to yield the correlation coefficients (r) for the Glu/Cr and Ch/Cr ratios.
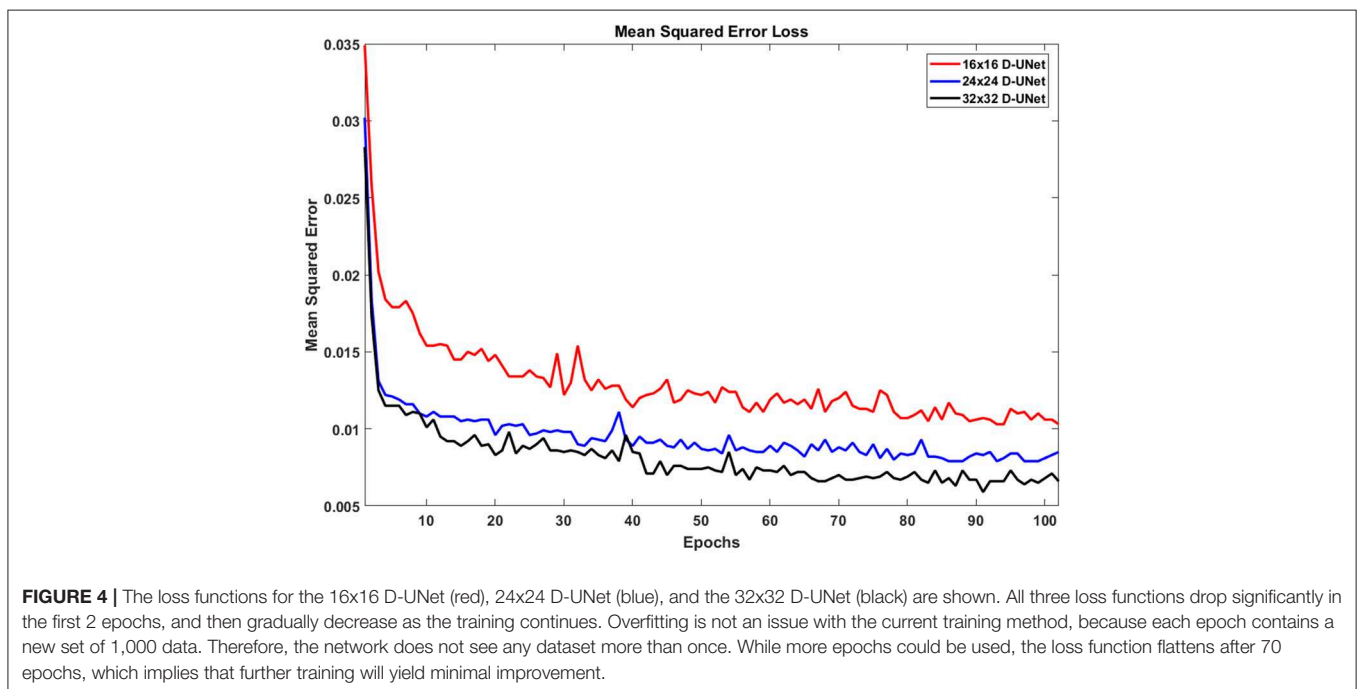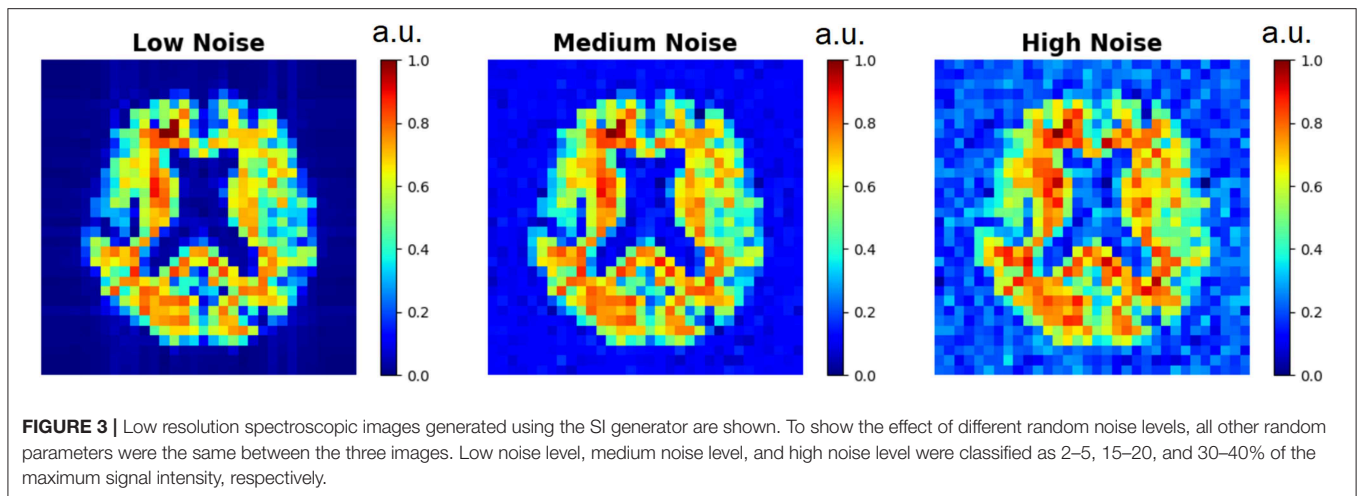
3. RESULTS

3.1. Training Results

Due to the novel D-UNet architecture, the mean squared error loss rapidly converged close to a reasonable value after only 2 epochs for all three networks, and the loss functions are shown in **Figure 4**. The loss continued to decrease with more epochs when a larger batch size was used for the remaining 100 epochs. From **Figure 4**, it is clear that the final loss was better for the 32x32 D-UNet than the 24x24 or 16x16 D-UNets. This is theoretically expected because higher initial resolution should aid in the estimation of unknown points, and this is true for conventional resolution enhancement techniques as well. While a low dropout was used in the architecture, overfitting was not a primary concern for the D-UNet training framework because of the reduced number of weighting parameters in the model. The results from the testing dataset also highlight the fact that the D-UNet training was generalized and applicable to never before seen data.

3.2. Test Set Results

Figure 5 displays the results from the three different D-UNet reconstructions, as well as the results of the standard zero-filling and bicubic interpolation methods. In order to provide a more stringent comparison, both zero-filling and bicubic interpolation were applied to the 32x32 low resolution metabolite maps instead of the lower resolution 16x16 or 24x24 metabolite maps. All of the D-UNet reconstructions are able to determine the abnormally high signal from the lesion shown in the T1w image. While zero-filling outperforms both bicubic interpolation and the 16x16 D-UNet, both the 24x24 and 32x32 D-UNets yield better results than zero-filling.

To demonstrate the capability of the SI generator, **Figure 6** shows a sample of the possible images produced from the same $aT1w$ image. The $Recon_{32x32}$ images are also shown, as well as
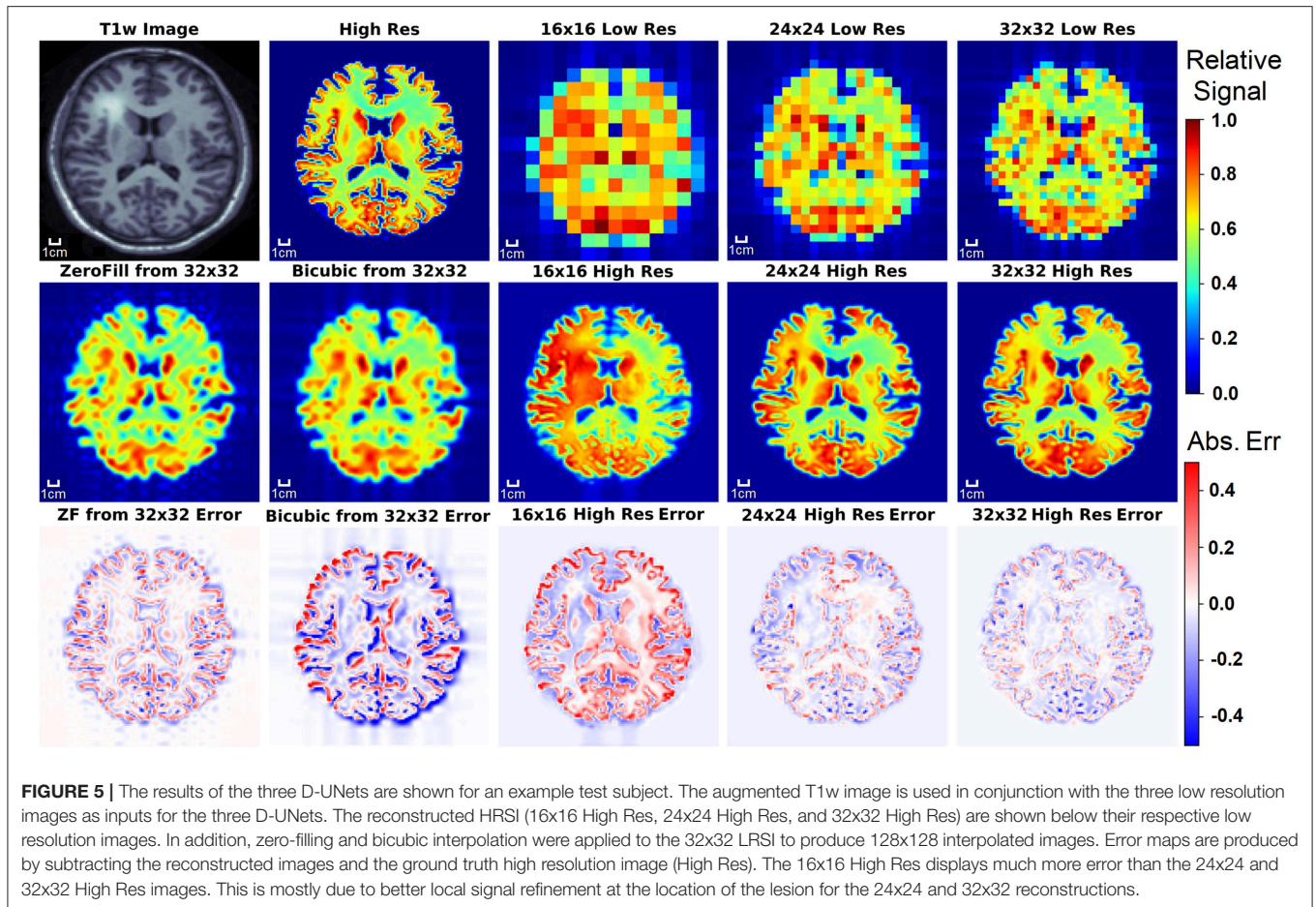
FIGURE 3 | Low resolution spectroscopic images generated using the SI generator are shown. To show the effect of different random noise levels, all other random parameters were the same between the three images. Low noise level, medium noise level, and high noise level were classified as 2–5, 15–20, and 30–40% of the maximum signal intensity, respectively.

FIGURE 4 | The loss functions for the 16x16 D-UNet (red), 24x24 D-UNet (blue), and the 32x32 D-UNet (black) are shown. All three loss functions drop significantly in the first 2 epochs, and then gradually decrease as the training continues. Overfitting is not an issue with the current training method, because each epoch contains a new set of 1,000 data. Therefore, the network does not see any dataset more than once. While more epochs could be used, the loss function flattens after 70 epochs, which implies that further training will yield minimal improvement.

difference maps between the HRSI and $Recon_{32x32}$. It is clear that the SI generator is capable of producing a wide variety of *SI* images that mimic biochemicals that are more prominent in GM, more prominent in WM, or equally prominent in both tissue types.

In addition, a quantitative comparison between these methods is shown in **Table 1**. Noise level was varied to determine the effect of noise on the super-resolution methods. Low noise level, medium noise level, and high noise level were classified as 2–5, 15–20, and 30–40% of the maximum signal intensity, respectively. From **Table 1**, the 32x32 D-UNet demonstrated the best performance at every noise level. At medium noise levels, the 24x24 D-UNet outperformed zero-filling, and at high noise levels both the 16x16 D-UNet and 24x24 D-UNet outperformed both zero-filling and bicubic interpolation.

3.3. Spectral Reconstruction Results

The ability of the D-UNets to reconstruct spectra at high spatial resolutions are highlighted in **Figure 7**. The 32x32 D-UNet reconstructs the lesion and contra-lateral white matter spectra reliably. In contrast, the 16x16 D-UNet underestimates the white matter spectrum. The 24x24 D-UNet performs very similarly to the 32x32 D-UNet, however it overestimates the Ch and mI signals in the lesion spectrum by roughly 20%. Overall, the mean squared error for the healthy white matter spectrum was 0.34, 0.030, and 0.0085 for the 16x16 D-UNet, 24x24 D-UNet, and 32x32 D-UNet, respectively. For the lesion spectrum, the mean squared error was 0.051, 0.36, and 0.13 for the 16x16 D-UNet, 24x24 D-UNet, and 32x32 D-UNet, respectively. From a quantitative standpoint, all three D-UNets would be able to

FIGURE 5 | The results of the three D-UNets are shown for an example test subject. The augmented T1w image is used in conjunction with the three low resolution images as inputs for the three D-UNets. The reconstructed HRSI (16x16 High Res, 24x24 High Res, and 32x32 High Res) are shown below their respective low resolution images. In addition, zero-filling and bicubic interpolation were applied to the 32x32 LRSI to produce 128x128 interpolated images. Error maps are produced by subtracting the reconstructed images and the ground truth high resolution image (High Res). The 16x16 High Res displays much more error than the 24x24 and 32x32 High Res images. This is mostly due to better local signal refinement at the location of the lesion for the 24x24 and 32x32 reconstructions.

determine the abnormally elevated Ch, as demonstrated from the metabolite maps.

3.4. *In vivo* Results

The ability of the 32x32 D-UNet to reconstruct the LRSI of Cr, NAA, Glu, Gln, Ch, and mI for the *in vivo* data is shown in **Figure 8**. This figure shows the reconstructed images, experimental HRSI, and difference maps between the two for each metabolite for one healthy volunteer. All reconstructed images retain the metabolite signals from the low resolution maps, and also show regional changes similar to the HRSI. For example, Glu is more concentrated in the GM and less concentrated in the WM, which is a well-known regional difference in the brain (33). Another well-known regional difference is that Ch is more concentrated in WM regions, which is apparent in both the reconstructed and experimental images. **Figure 9** shows reconstructions with low, average, and large MSE values. In general, lower SNR metabolites appeared to have a larger MSE value compared to higher SNR metabolites. From a quantitative standpoint, the average MSE values over the ten volunteers for Cr, NAA, Glu, Gln, Ch, and mI were 0.0048, 0.0042, 0.0060, 0.0079, 0.0059, and 0.0056 respectively. These errors are displayed in **Figure 10D** and plotted against the average MSE values obtained for the testing set using different

noise levels (low, medium, high). It is clear that the MSE values are in most cases comparable to simulated test images with 2–20% noise, with the exception of Gln which is most comparable to test images with 35% noise.

Figure 10 also shows the Glu/Cr and Ch/Cr ratios as a function of the T1w intensity averaged over the ten volunteers. The ratios are taken after normalization of the metabolites as part of the super-resolution reconstruction, which is why Ch/Cr appears larger than Glu/Cr in the figure. The trend shows that with higher WM content, Glu/Cr decreases while Ch/Cr increases. The correlation between the experimental HRSI and Recon results are shown in **Figure 10C**. Quantitatively, both Glu/Cr and Ch/Cr ratios have high squared correlation coefficients, $r^2 > 0.99$. This highlights the fact that important biological relationships are preserved in the reconstructed images.

4. DISCUSSION

Although SI provides invaluable information regarding the biomolecular processes of tissues *in vivo*, experimental limitations have greatly hindered the integration of this method into standard clinical protocols. This study demonstrates a technique capable of overcoming one of the greatest challenges
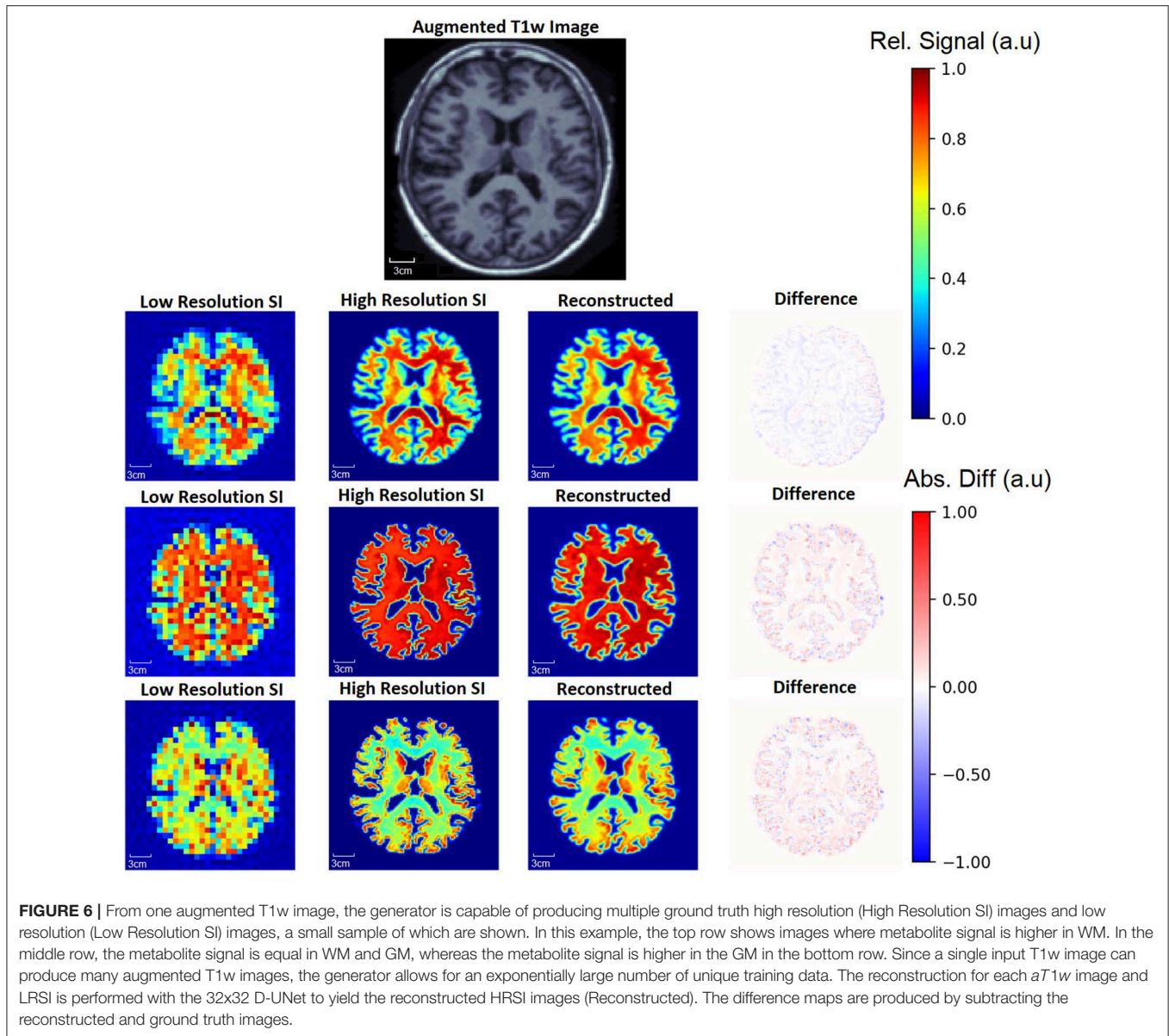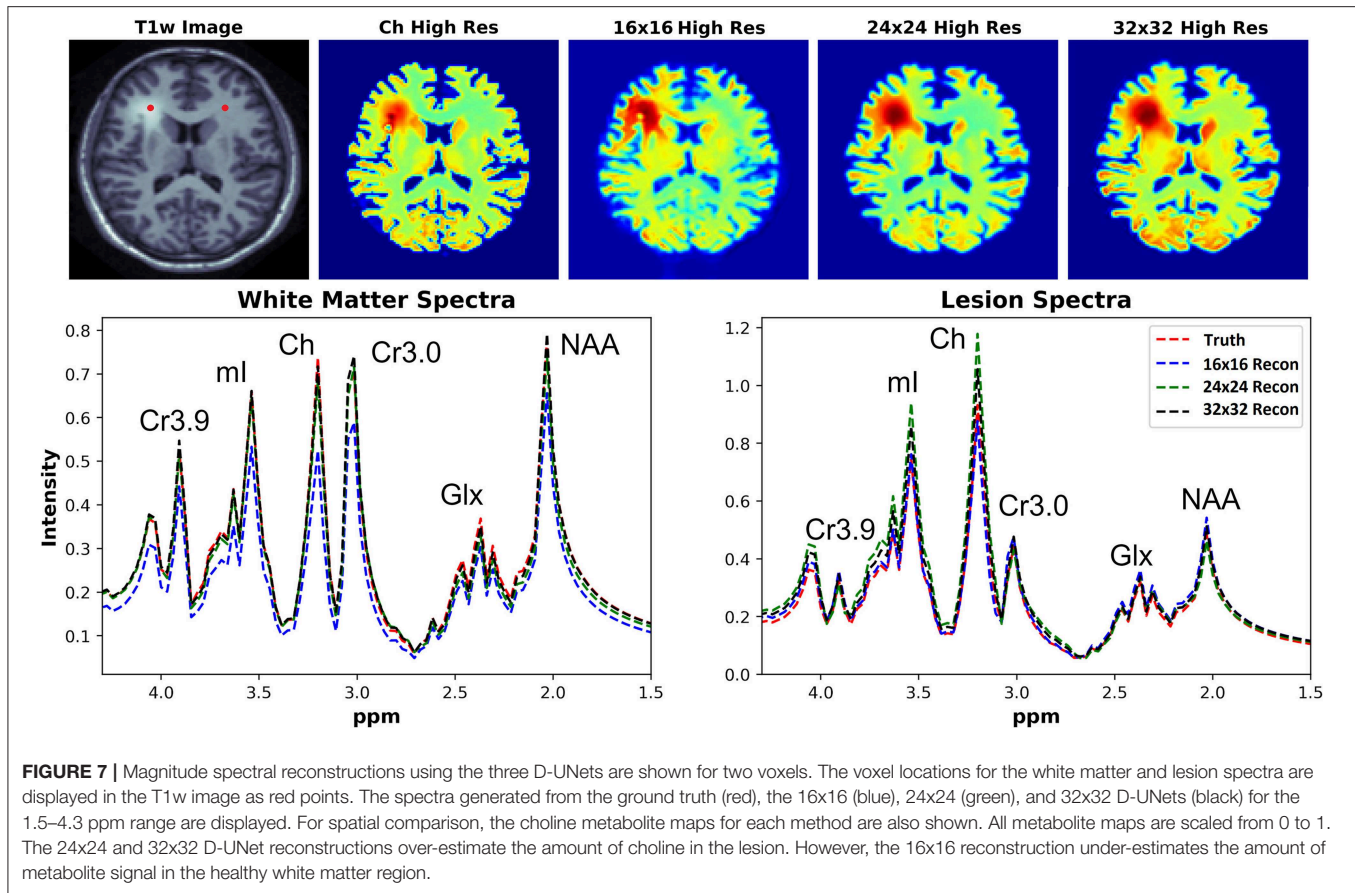
FIGURE 6 | From one augmented T1w image, the generator is capable of producing multiple ground truth high resolution (High Resolution SI) images and low resolution (Low Resolution SI) images, a small sample of which are shown. In this example, the top row shows images where metabolite signal is higher in WM. In the middle row, the metabolite signal is equal in WM and GM, whereas the metabolite signal is higher in the GM in the bottom row. Since a single input T1w image can produce many augmented T1w images, the generator allows for an exponentially large number of unique training data. The reconstruction for each $aT1w$ image and LRSI is performed with the 32x32 D-UNet to yield the reconstructed HRSI images (Reconstructed). The difference maps are produced by subtracting the reconstructed and ground truth images.

TABLE 1 | The mean squared error between the high resolution ground truth (HRSI) and several methods are tabulated.

Method	Noise Level		
	Low	**Medium**	**High**
Zero-Fill from 32x32	1.109	1.652	4.505
Bicubic from 32x32	2.794	3.129	3.820
16x16 D-UNet	1.863	2.420	2.761
24x24 D-UNet	1.139	1.316	1.745
32x32 D-UNet	**0.7460**	**0.9722**	**1.599**

These values are the total sum of the mean squared error over 169 test subjects. The 32x32 D-UNet reconstruction outperforms all of the other methods. With higher random noise present in the LRSI, the 16x16 and 24x24 D-UNets outperform both zero-filling and bicubic interpolation. It is important to note that this is true even though the zero-filling and bicubic interpolation methods are applied to a 32x32 image. Bold values indicate the method with the lowest mean squared error for each comparison.

in SI, which is poor spatial resolution. By utilizing a deep learning framework, it is shown in **Figures 5–9** that high resolution spectroscopic images can be produced from the combination of low resolution spectroscopic images and T1w images. In addition, as seen in **Figure 7**, it is possible to reconstruct spectra at higher spatial resolutions. The reconstruction method also preserves important regional metabolic differences and shows low errors for *in vivo* reconstructions, as shown in **Figure 10**. This deep learning super-resolution method was compared to both zero-filling and bicubic interpolation, and proved to be better than these methods for all noise levels.

Deep learning requires large datasets, which are not readily available for SI. Unfortunately, there is also a lack of ground truth for high resolution spectroscopic imaging due to the fact that experimental results may contain chemical shift displacement

FIGURE 7 | Magnitude spectral reconstructions using the three D-UNets are shown for two voxels. The voxel locations for the white matter and lesion spectra are displayed in the T1w image as red points. The spectra generated from the ground truth (red), the 16x16 (blue), 24x24 (green), and 32x32 D-UNets (black) for the 1.5–4.3 ppm range are displayed. For spatial comparison, the choline metabolite maps for each method are also shown. All metabolite maps are scaled from 0 to 1. The 24x24 and 32x32 D-UNet reconstructions over-estimate the amount of choline in the lesion. However, the 16x16 reconstruction under-estimates the amount of metabolite signal in the healthy white matter region.

artifacts, B_0 inhomogeneity issues, partial volume effects, low signal to noise ratios, water contamination, or other forms of signal contamination. It is also prohibitively long to scan at high resolution (128x128) without using several acceleration methods, making a ground truth impossible to obtain from the human brain with current technology. Therefore, an SI generator was developed to simulate training and testing data from a publicly available dataset. By including various probabilistic transformations, such as contrast variations, metabolic signal changes, and FOV variations, the SI generator was capable of providing a diverse and large dataset for the training of the three D-UNets. These data may not be entirely realistic, and this generator must be validated more rigorously in the future. For this study, the dataset does seem to be representative of real acquisitions, as seen from the *in vivo* results.

The $Recon_{32x32}$ and HRSI experimental images are very similar, as seen from **Figures 8–10**. The reconstructed images show better resemblance to the anatomical T1w images, including cerebral spinal fluid localization. However, both the $Recon_{32x32}$ and HRSI experimental images provide similar quantitative results, as seen in **Figure 10**. Theoretically, the $Recon_{32x32}$ images would require $\frac{1}{16}th$ to $\frac{1}{4}th$ the time to acquire, depending on the acceleration methods implemented. Therefore, it is important to note that aside from super-resolution, the D-UNet may also be used as a means to accelerate a spectroscopic

imaging protocol in the future. Additionally, the reconstructed *in vivo* images are denoised while retaining essential metabolic information for different tissues of the brain, which may be desirable for certain applications. While the simulated and *in vivo* data demonstrate that the reconstruction method is accurate, one of the main disadvantages of this work is that it has not been validated *in vitro*. This is due to the fact that a high resolution SI phantom similar to the human brain is not available. Since the D-UNet model is trained using *in vivo* anatomy, it is not capable of reconstructing high resolution images from unrealistic geometries. Therefore, future work will focus on the development of a realistic, high resolution SI phantom for validation.

Even though the D-UNets outperformed zero-filling and bicubic interpolation, these models may not be perfect for HRSI reconstruction primarily due to experimental imperfections. As seen from **Table 1**, error increases as a function of noise. Intuitively, chemicals that are found in the body at lower concentration may have larger reconstruction errors than chemicals with higher SNR, which is also supported by the *in vivo* results shown in **Figure 9** where the Gln reconstructed images have higher error than the other metabolite images. Therefore, prediction accuracy is limited by the quality of the original LRSI. Also, while the *in vivo* results have low mean-squared errors, it is important to note that down-sampling from a high resolution acquisition decreases potential acquisition problems
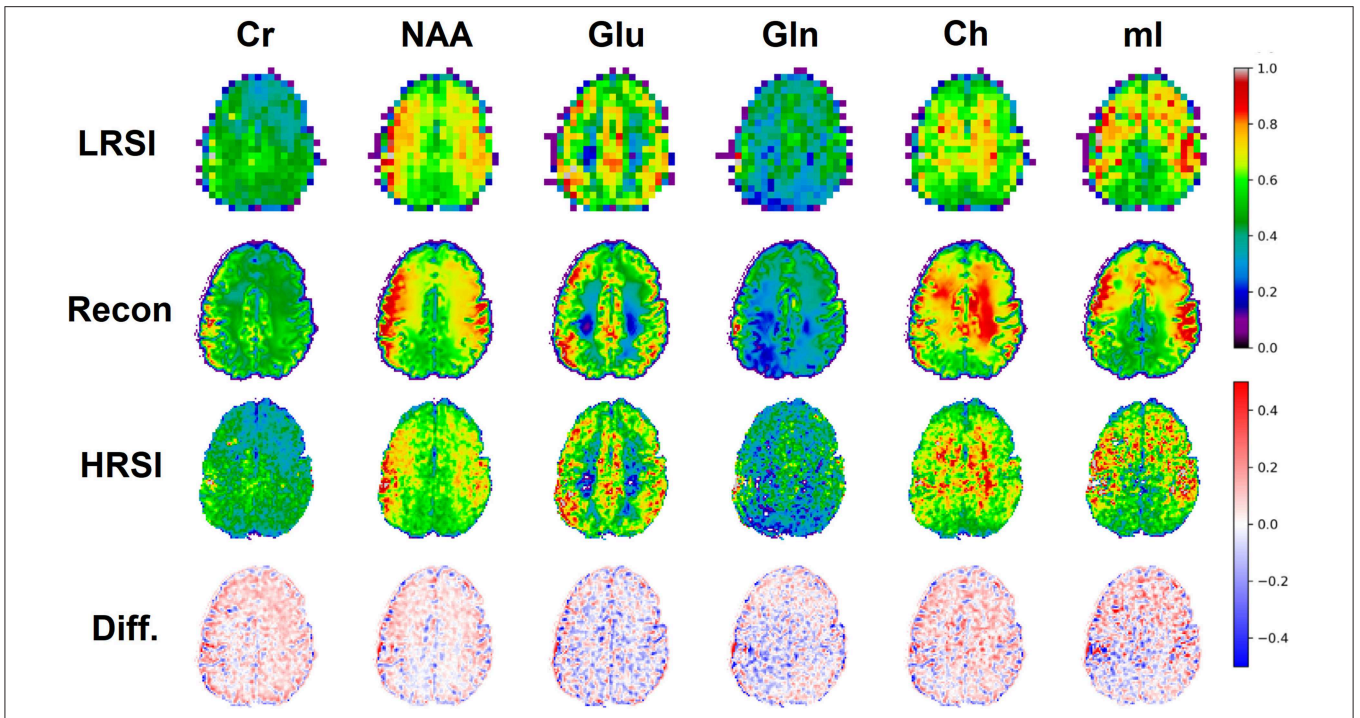
FIGURE 8 | An *in vivo* example of a healthy volunteer is used to demonstrate the potential application for the D-UNet. The experimental high resolution SI (HRSI) data was acquired at 128x128 resolution using an accelerated acquisition protocol (20). This data was then down-sampled to produce 32x32 low resolution SI (LRSI) metabolite maps for Cr, NAA, Glu, Gln, Ch, and mI. Together with the T1w image, the low resolution metabolite images were used to reconstruct high resolution spectroscopic images (Recon) using the 32x32 D-UNet model. The difference maps between the Recon and HRSI images (Diff) are also shown.
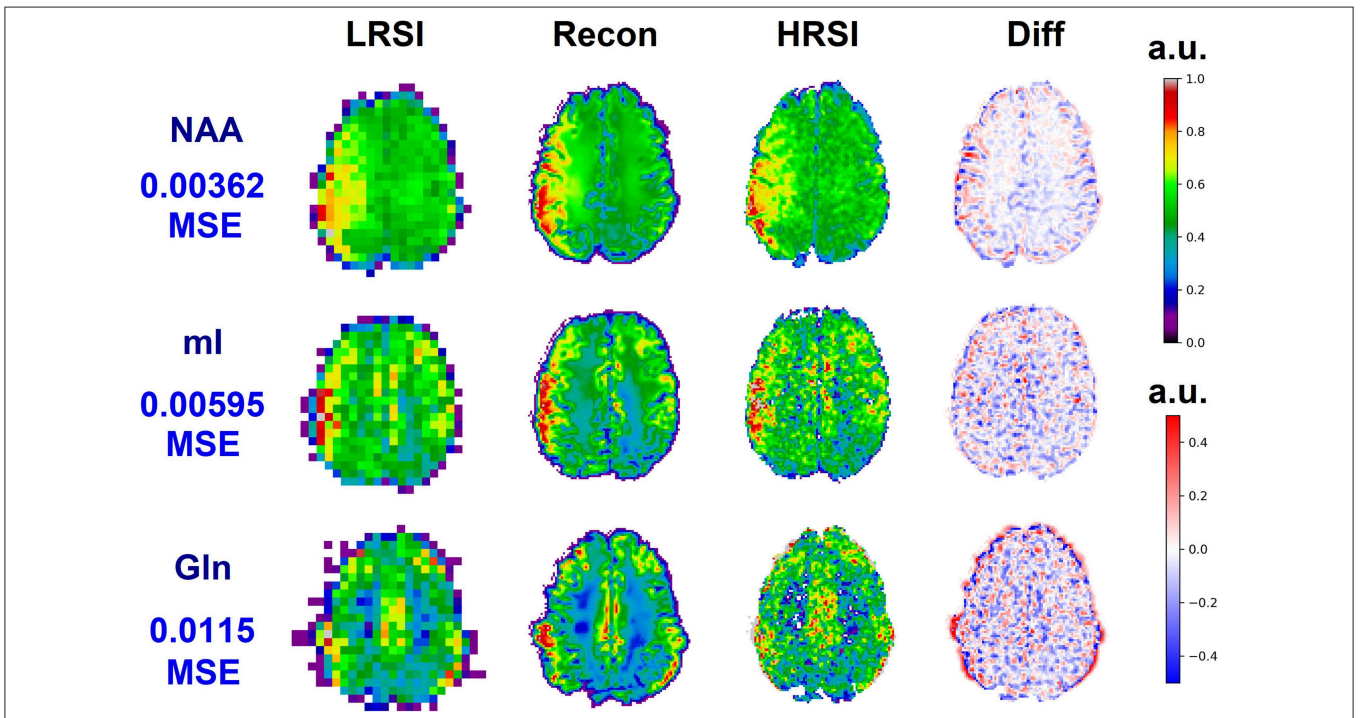
FIGURE 9 | Three reconstructions showing low MSE (top), average MSE (middle), and large MSE (bottom) are shown for three different volunteers. In general, lower MSE were observed for metabolites with higher signal-to-noise ratios such as NAA, whereas larger MSE values were calculated for low SNR metabolites such as Gln. The different MSE values are highlighted by the errors seen in the difference maps.

FIGURE 10 | The Glu/Cr **(A)** and Ch/Cr **(B)** signals averaged over ten subjects are shown for the experimentally acquired images (HRSI) and the images reconstructed from low resolution 32x32 images (Recon). The signals are shown as a function of the T1w intensity, which is representative of the gray and white matter content of the voxel. The red dotted line represents the point at which the gray matter content equals the white matter content in a voxel. The correlation between the HRSI and Recon values are plotted **(C)** with linear fits. For both Glu/Cr and Ch/Cr, the r^2 values of the fits are above 0.99. Finally, the mean squared error for the ten subjects calculated between the HRSI and Recon for each metabolite map **(D)** is displayed. The dotted black lines reflect the MSE from the testing set for different noise values (low, medium, and high).

such as lipid contamination and partial volume effects. Therefore, it is expected that a prospectively acquired low resolution data set will yield higher errors when reconstructed using the D-UNet. This must be evaluated in a more rigorous study where both low resolution and high resolution experimental SI data are acquired.

Of course, the original resolution of the experimental SI plays a large role in the reconstruction process. While 24x24 and 32x32 matrices provide relatively accurate high resolution reconstructions, the 16x16 resolution does not perform as well. This suggests that there is a lower bound necessary to accurately upscale high resolution SI. This might be true for other super-resolution techniques (21), so a more thorough comparison between this deep learning method and other methods may aid in identifying this lower bound. Furthermore, results may be biased by the quantitative methods implemented to produce the LRSI before the super-resolution process is performed. This bias could be removed in the future by developing a deep learning based approach to metabolite quantitation (44). However, it may be worthwhile to explore the differences between common one dimensional spectral quantitation

programs, such as LCModel (43) or TARQUIN (45), on the upscaling process.

From the spectral reconstruction results shown in **Figure 7**, it is apparent that some metabolites are over- and under- estimated during the reconstruction process. Therefore, clinical diagnosis based on the D-UNet reconstruction must be made with caution, as results from this method could lead to false positives or false negatives. Before basing diagnosis on the D-UNet reconstruction, the process should be evaluated *in vivo* in a well-known brain cancer pathology to assess the rates of false positives or false negatives detected by experienced radiologists in the field.

The deep learning method presented in this study may be useful for other super-resolution transformations in the field of medical imaging. This is especially true for spectroscopic imaging of other nuclei, such as ^{13}C and ^{31}P, where lower SNR results in low spatial resolution acquisitions. Recently, accelerated hyper-polarized ^{13}C spectroscopic imaging has shown to be promising for imaging prostate cancer (46, 47), and this technique could benefit by using the D-UNet model. In addition, ^{31}P spectroscopic imaging has also been used to image

cancer (48, 49). The main drawback, again, is the lack of SNR to adequately acquire high spatial resolution data. High resolution acquisition schemes have been proposed for ^{31}P spectroscopic imaging (50), and the D-UNet model could provide an alternative for improving spatial resolution. The same SI generation process could be used for training for these other nuclei, however different anatomical sites must be included (breast, prostate, etc.) to yield accurate results depending on the desired application.

The same principles discussed in this work may also apply to positron emission tomography (PET) (51). It is well-known that the radioactive tracer is more prominent in certain tissues and lesions, and positrons from this tracer travel some distance before annihilating to produce the PET signal. The distance between the source and the annihilation can be thought of as a partial volume effect. This model can potentially be used to learn how to remove this partial volume effect artifact, and this would be applicable for CT-PET or MR-PET acquisitions. Ultimately, this deep learning model allows for the acquisition of high quality images without increasing the scan time or improving the hardware of the imaging system.

5. CONCLUSION

The D-UNet model presented in this study allows for the reconstruction of accurate super-resolution magnetic resonance spectroscopic images from the human brain. Utilizing this method, we demonstrate that a simulated, low resolution chemical map can be transformed together with the T1w image to produce a high resolution chemical map. This method demonstrates better accuracy than typical zero-filling and bicubic interpolation methods. Furthermore, we demonstrate that the accuracy of this model holds when evaluating our method on retrospective *in vivo* data. This model still needs to be validated on prospective *in vivo* data in the future. After further *in vitro* and *in vivo* validation, this method may be utilized for denoising, scan acceleration, and improved tissue delineation.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The Institutional Review Board (IRB) at the Medical University of Vienna approved the study and ten healthy, adult volunteers signed written and informed consent forms prior to imaging studies. All experiments were performed in accordance with relevant guidelines and regulations of the IRB.

AUTHOR CONTRIBUTIONS

ZI and SJ conceived the experiments. DN designed the deep learning architecture. ZI and DN conducted the deep learning experiments. GH, SM, and WB acquired and processed the *in vivo* data. ZI and SJ analyzed the results. All authors reviewed the manuscript.

FUNDING

REFERENCES

1. Brown T, Kincaid B, Ugurbil K. NMR chemical shift imaging in three dimensions. *Proc Natl Acad Sci USA.* (1982) 79:3523–6. doi: 10.1073/pnas.79.11.3523
2. Soares D, Law M. Magnetic resonance spectroscopy of the brain: review of metabolites and clinical applications. *Clin Radiol.* (2009) 64:12–21. doi: 10.1016/j.crad.2008.07.002
3. Govindaraju V, Young K, Maudsley AA. Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR Biomed.* (2000) 13:129–53. doi: 10.1002/1099-1492(200005)13:3<129::AID-NBM619>3.0.CO;2-V
4. Ramadan S, Lin A, Stanwell P. Glutamate and glutamine: a review of *in vivo* mrs in the human brain. *NMR Biomed.* (2013) 26:1630–46. doi: 10.1002/nbm.3045
5. Mansfield P. Spatial mapping of the chemical shift in NMR. *Magn Reson Med.* (1984) 1:370–86. doi: 10.1002/mrm.1910010308
6. Posse S, Tedeschi G, Risinger R, Ogg R, Bihan DL. High speed ^{1}H spectroscopic imaging in human brain by echo planar spatial-spectral encoding. *Magn Reson Med.* (1995) 33:34–40. doi: 10.1002/mrm.1910330106
7. Adalsteinsson E, Irarrazabal P, Topp S, Meyer C, Macovski A, Spielman DM. Volumetric spectroscopic imaging with spiral-based k-space trajectories. *Magn Reson Med.*(1998) 39:889–98.
8. Furuyama JK, Wilson NE, Thomas MA. Spectroscopic imaging using concentrically circular echo-planar trajectories *in vivo*. *Magn Reson Med.* (2012) 67:1515–22. doi: 10.1002/mrm.23184

9. Schirda CV, Tanase C, Boada FE. Rosette spectroscopic imaging: optimal parameters for alias-free, high sensitivity spectroscopic imaging. *J Magn Reson Imaging.* (2009) 29:1375–85. doi: 10.1002/jmri.21760
10. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. Sense: sensitivity encoding for fast MRI. *Magn Reson Med.* (1999) 42:952–62. doi: 10.1002/(SICI)1522-2594(199911)42:5<952::AID-MRM16>3.0.CO;2-S
11. Dydak U, Weiger M, Pruessmann KP, Meier D, Boesiger P. Sensitivity-encoded spectroscopic imaging. *Magn Reson Med.* (2001) 46:713–22. doi: 10.1002/mrm.1250
12. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med.* (2002) 47:1202–10. doi: 10.1002/mrm.10171
13. Otazo R, Tsai SY, Lin FH, Posse S. Accelerated short-te 3d proton echo-planar spectroscopic imaging using 2d-sense with a 32-channel array coil. *Magn Reson Med.* (2007) 58:1107–16. doi: 10.1002/mrm.21426
14. Strasser B, Považan M, Hangel G, Hingerl L, Chmelik M, Gruber S, et al. (2+1) d-caipirinha accelerated MR spectroscopic imaging of the brain at 7t. *Magn Reson Med.* (2017) 78:429–40. doi: 10.1002/mrm.26386
15. Wilson NE, Iqbal Z, Burns BL, Keller M, Thomas MA. Accelerated five-dimensional echo planar J-resolved spectroscopic imaging: implementation and pilot validation in human brain. *Magn Reson Med.* (2016) 75:42–51. doi: 10.1002/mrm.25605
16. Iqbal Z, Wilson NE, Thomas MA. 3d spatially encoded and accelerated te-averaged echo planar spectroscopic imaging in healthy human brain. *NMR Biomed.* (2016) 29:329–39. doi: 10.1002/nbm.3469

17. Posse S, Otazo R, Dager SR, Alger J. MR spectroscopic imaging: principles and recent advances. *J. Magn. Reson. Imaging.* (2013) 37:1301–25. doi: 10.1002/jmri.23945

18. Lam F, Liang ZP. A subspace approach to high-resolution spectroscopic imaging. *Magn Reson Med.* (2014) 71:1349–57. doi: 10.1002/mrm.25168

19. Hingerl L, Bogner W, Moser P, Považan M, Hangel G, Heckova E, et al. Density-weighted concentric circle trajectories for high resolution brain magnetic resonance spectroscopic imaging at 7t. *Magn Reson Med.* (2017) 79:2874–85. doi: 10.1002/mrm.26987

20. Hangel G, Strasser B, Považan M, Heckova E, Hingerl L, Boubela R, et al. Ultra-high resolution brain metabolite mapping at 7 t by short-TR hadamard-encoded FID-MRSI. *Neuroimage.* (2016) 168:199–210. doi: 10.1016/j.neuroimage.2016.10.043

21. Jain S, Sima DM, Sanaei Nezhad F, Hangel G, Bogner W, Williams S, et al. Patch-based super-resolution of MR spectroscopic images: application to multiple sclerosis. *Front Neurosci.* (2017) 11:13. doi: 10.3389/fnins.2017.00013

22. Haldar JP, Hernando D, Song SK, Liang ZP. Anatomically constrained reconstruction from noisy data. *Magn Reson Med.* (2008) 59:810–8. doi: 10.1002/mrm.21536

23. Hu X, Levin DN, Lauterbur PC, Spraggins T. Slim: spectral localization by imaging. *Magn Reson Med.* (1988) 8:314–22. doi: 10.1002/mrm.1910080308

24. Cengiz S, Valdes-Hernandez MdC, Ozturk-Isik E. Super resolution convolutional neural networks for increasing spatial resolution of ^{1} h magnetic resonance spectroscopic imaging. In: *Annual Conference on Medical Image Understanding and Analysis.* Cham: Springer (2017). p. 641–50.

25. Jacob M, Zhu X, Ebel A, Schuff N, Liang ZP. Improved model-based magnetic resonance spectroscopic imaging. *IEEE Trans Med Imaging.* (2007) 26:1305–18. doi: 10.1109/TMI.2007.898583

26. Liang ZP, Lauterbur PC. A generalized series approach to MR spectroscopic imaging. *IEEE Trans Med Imaging.* (1991) 10:132–7. doi: 10.1109/42.79470

27. Kasten J, Klauser A, Lazeyras F, Van De Ville D. Magnetic resonance spectroscopic imaging at superresolution: overview and perspectives. *J Magn Reson.* (2016) 263:193–208. doi: 10.1016/j.jmr.2015.11.003

28. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* (1989) 1:541–51. doi: 10.1162/neco.1989.1.4.541

29. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* (2015) 521:436. doi: 10.1038/nature14539

30. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems.* (2012). p. 1097–105.

31. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Cham: Springer (2015). p. 234–41.

32. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *J Cogn Neurosci.* (2007) 19:1498–507.doi: 10.1162/jocn.2007.19.9.1498

33. Pouwels PJ, Frahm J. Regional metabolite concentrations in human brain as determined by quantitative localized proton MRS. *Magn Reson Med.* (1998) 39:53–60. doi: 10.1002/mrm.1910390110

34. Huang G, Liu Z, Weinberger KQ, van der Maaten L. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Vol. 1.* (2017). p. 3.

35. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res.* (2014) 15:1929–58.

36. Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv:1412.6980* (2014).

37. Chollet F. *Keras.* (2015). Available online at: https://github.com/fchollet/keras

38. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. Tensorflow: large-scale machine learning on heterogeneous distributed systems. *arXiv:1603.04467* (2016).

39. Smith SL, Kindermans PJ, Ying C, Le QV. Don't decay the learning rate, increase the batch size. *arXiv:1711.00489* (2017).

40. Smith S, Levante T, Meier BH, Ernst RR. Computer simulations in magnetic resonance. An object-oriented programming approach. *J Magn Reson Ser A.* (1994) 106:75–105. doi: 10.1006/jmra.1994.1008

41. Bogner W, Gruber S, Trattnig S, Chmelik M. High-resolution mapping of human brain metabolites by free induction decay ^{1}H MRSI at 7 T. *NMR Biomed.* (2012) 25:873–82. doi: 10.1002/nbm.1805

42. Bilgic B, Chatnuntawech I, Fan AP, Setsompop K, Cauley SF, Wald LL, et al. Fast image reconstruction with l2-regularization. *J Magn Reson Imaging.* (2014) 40:181–91. doi: 10.1002/jmri.24365

43. Provencher SW. Estimation of metabolite concentrations from localized *in vivo* proton nmr spectra. *Magn Reson Med.* (1993) 30:672–9. doi: 10.1002/mrm.1910300604

44. Hatami N, Sdika M, Ratiney H. Magnetic resonance spectroscopy quantification using deep learning. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Cham: Springer (2018). p. 467–75.

45. Wilson M, Reynolds G, Kauppinen RA, Arvanitis TN, Peet AC. A constrained least-squares approach to the automated quantitation of *in vivo* ^{1}H magnetic resonance spectroscopy data. *Magn Reson Med.* (2011) 65:1–12. doi: 10.1002/mrm.22579

46. Hu S, Lustig M, Chen AP, Crane J, Kerr A, Kelley DA, et al. Compressed sensing for resolution enhancement of hyperpolarized ^{13}C flyback 3D-MRSI. *J Magn Reson.* (2008) 192:258–64. doi: 10.1016/j.jmr.2008.03.003

47. Nelson SJ, Kurhanewicz J, Vigneron DB, Larson PE, Harzstark AL, Ferrone M, et al. Metabolic imaging of patients with prostate cancer using hyperpolarized [1-13c] pyruvate. *Sci Transl Med.* (2013) 5:198ra108. doi: 10.1126/scitranslmed.3006070

48. Hugg JW, Matson GB, Twieg DB, Maudsley AA, Sappey-Marinier D, Weiner MW. Phosphorus-31 mr spectroscopic imaging (MRSI) of normal and pathological human brains. *Magn Reson Imaging.* (1992) 10:227–43. doi: 10.1016/0730-725X(92)90483-G

49. Klomp DW, van de Bank BL, Raaijmakers A, Korteweg MA, Possanzini C, Boer VO, et al. 31p MRSI and ^{1}H MRS at 7 T: initial results in human breast cancer. *NMR Biomed.* (2011) 24:1337–42. doi: 10.1002/nbm.1696

50. Ma C, Clifford B, Liu Y, Gu Y, Lam F, Yu X, et al. High-resolution dynamic 31p-MRSI using a low-rank tensor model. *Magn Reson Med.* (2017) 78:419–28. doi: 10.1002/mrm.26762

51. Kennedy JA, Israel O, Frenkel A, Bar-Shalom R, Azhari H. Super-resolution in pet imaging. *IEEE Trans Med Imaging.* (2006) 25:137–47. doi: 10.1109/TMI.2005.861705

52. Iqbal Z, Nguyen D, Hangel G, Motyka S, Bogner W, Jiang S. Super-resolution ^{1}H magnetic resonance spectroscopic imaging utilizing deep learning. *arXiv:1802.07909* (2018).