Generalised Super Resolution for Quantitative MRI Using Self-Supervised Mixture of Experts

Hongxiang Lin[1,2,3⋆], Yukun Zhou[2,4,5⋆], Paddy J. Slator[2,3], and Daniel C. Alexander[2,3]

[1] Research Center for Healthcare Data Science, Zhejiang Lab, Hangzhou, China
harryhxlin@gmail.com
[2] Centre for Medical Image Computing, University College London, London, UK
[3] Department of Computer Science, University College London, London, UK
[4] Department of Medical Physics and Biomedical Engineering, UCL, London, UK
[5] NIHR Biomedical Research Centre at Moorfields Eye Hospital, London, UK

Abstract. Multi-modal and multi-contrast imaging datasets have diverse voxel-wise intensities. For example, quantitative MRI acquisition protocols are designed specifically to yield multiple images with widely-varying contrast that inform models relating MR signals to tissue characteristics. The large variance across images in such data prevents the use of standard normalisation techniques, making super resolution highly challenging. We propose a novel self-supervised mixture-of-experts (SS-MoE) paradigm for deep neural networks, and hence present a method enabling improved super resolution of data where image intensities are diverse and have large variance. Unlike the conventional MoE that automatically aggregates expert results for each input, we explicitly assign an input to the corresponding expert based on the predictive pseudo error labels in a self-supervised fashion. A new gater module is trained to discriminate the error levels of inputs estimated by Multiscale Quantile Segmentation. We show that our new paradigm reduces the error and improves the robustness when super resolving combined diffusion-relaxometry MRI data from the Super MUDI dataset. Our approach is suitable for a wide range of quantitative MRI techniques, and multi-contrast or multi-modal imaging techniques in general. It could be applied to super resolve images with inadequate resolution, or reduce the scanning time needed to acquire images of the required resolution. The source code and the trained models are available at `https://github.com/hongxiangharry/SS-MoE`.

Keywords: Self Supervision · Mixture of Experts · Quantitative MRI · Generalised Super Resolution · Pseudo labels.

1 Introduction

Quantitative Magnetic Resonance Imaging (qMRI) can measure and map chemical, physical, and physiological values that strongly relate to underlying tissue

⋆ HL and YZ contributed equally.

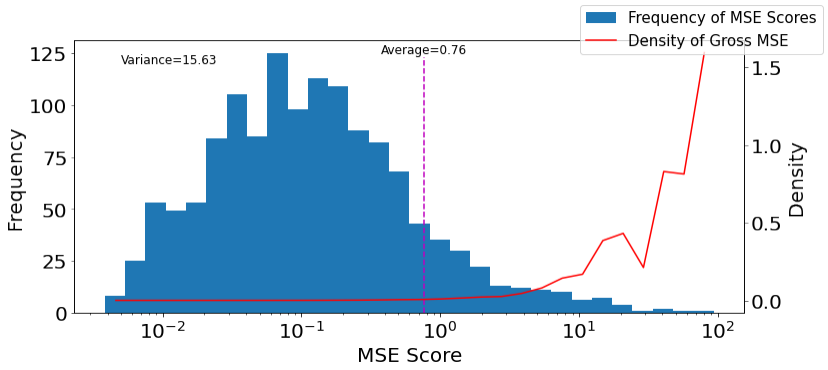

Fig. 1. (Blue block) Histogram of mean-squared-error (MSE) between interpolated LR volumes and HR volumes in Subject cdmri0015 in the first anisotropic Super MUDI Dataset. All MSE scores were clustered into 30 bins within log scale. (Red line) Density of the gross MSE scores over all volumes. Cubic spline method was used to interpolate LR volume.

structure and function. Such measurements have the potential to improve diagnosis, prognosis and monitoring of a wide variety of diseases. However, qMRI has not yet been widely used in the clinic, due to long acquisition times and noise sensitivity. Super-resolution (SR) reconstruction techniques enable images with the same spatial resolution to be acquired within reduced scanning times and with improved signal-to-noise ratios [30]. Improved SR techniques can hence increase the likelihood of clinical adoption of qMRI, as well as similar multi-modal or multi-contrast imaging techniques, such as multi-contrast X-ray [32] and multi-modal functional imaging [29].

Deep learning based SR for medical imaging has demonstrated significant improvements over existing techniques [4, 5, 18, 19, 33, 34]. However, the data normalisation required for deep learning SR hinders its application to multi-modal or multi-contrast techniques such as qMRI, as such imaging datasets have diverse voxel-wise intensities leading to large variances that prevent the use of standard normalisation techniques. In conventional MRI SR, intensity normalisation is performed within single images using a method such as Z-score, fussy C-mean, or Gaussian mixture model [22]. These approaches can be applied to individual qMRI images sequentially to normalise the intensity scale, but this affects the relationship between voxelwise intensities and MR sequence parameters, biasing downstream analyses that interpret these relationships to estimate underlying tissue properties. An alternative, used by most state-of-the-art deep learning architectures in computer vision and medical imaging, is batch normalisation [13]. However, similarly to intensity normalisation, the reconstruction accuracy degrades rapidly when the training batches have a large variance [23].

To generalise SR to data with large underlying variance, such as qMRI, we propose a self-supervised mixture-of-experts (SS-MoE) paradigm that can aug-

ment any encoder-decoder network backbone. The conventional mixture of experts automatically aggregates expert results for each input; see [14] and its recent extensions in [8, 24, 35]. Unlike this, our proposed SS-MoE discriminates an input data by predicting the error class in a self-supervised fashion, to make hard assignments to the corresponding expert decoder network [8]. This divide-and-conquer strategy predictively clusters input data belonging to the same error level, thereby reducing the variance in each cluster. As such, the resulting outputs of the multi-hand networks formulate a predictive distribution with multiple peaks with respect to error clusters, rather than a single peak; see a similar analysis developed in [31]. In this paper, we select a U-Net variant as an encoder-decoder network backbone [11, 17], which was employed in the Image Quality Transfer framework, a patch-based machine learning approach used to enhance the resolution and/or contrast in diffusion MRI and structural MRI [2, 3, 17, 26, 27]. We apply our method to Super MUDI challenge [21] combined diffusion-relaxometry MRI data; this is a challenging qMRI dataset for SR as the simultaneous inversion recovery, multi-echo gradient echo, and diffusion sequences all combine to yield large intensity variance in voxels across volumes.

Our main contributions are: 1) To reduce population variance, we separately assign the inputs into multiple expert networks based on the predictive pseudo error labels. 2) We train a new gater module to predict the pseudo error labels from the extracted high-level perceptual features from the baseline network; pseudo error labels are typically estimated by unsupervised or heuristic ways, and are used to train the gater. Our overall paradigm is non-end-to-end so is potentially extendable to most other encoder-decoder network architectures.

2 Method

2.1 Data Description

We perform SR on the publicly available Super MUDI dataset [1, 21], which comprises combined diffusion-relaxometry brain scans on 5 healthy subjects using the ZEBRA technique [12]. Each subject comprises: original isotropic data with $2.5 \times 2.5 \times 2.5$ mm high-resolution (HR) voxels, corresponding $2\times$-downsampled anisotropic data with $2.5 \times 2.5 \times 5$ mm low-resolution (LR) voxels, and $2\times$-downsampled isotropic data with $5 \times 5 \times 5$ mm LR voxels. Thus, two super resolution tasks can be defined: $2\times$ through-plane SR and $2\times$ isotropic SR. We split each kind of data to have 6720 3D volumes to enable an error analysis across the volumes on a subject. We observed that the majority of reconstruction errors are concentrated in regions with small errors, whilst the largest errors contribute most to the overall error; see Fig. 1.

Next, to establish the training paired patches for the two tasks, we first randomly cropped N patches of the shape $(16, 16, 16)$ whose non-background voxels account for over 50% of patch volume from the original resolution data to serve as the ground-truth HR patch y_i, where $i = 1, \cdots, N$. To form the corresponding LR half x_i, we cropped the same field of view, respectively from the two downsampled data.

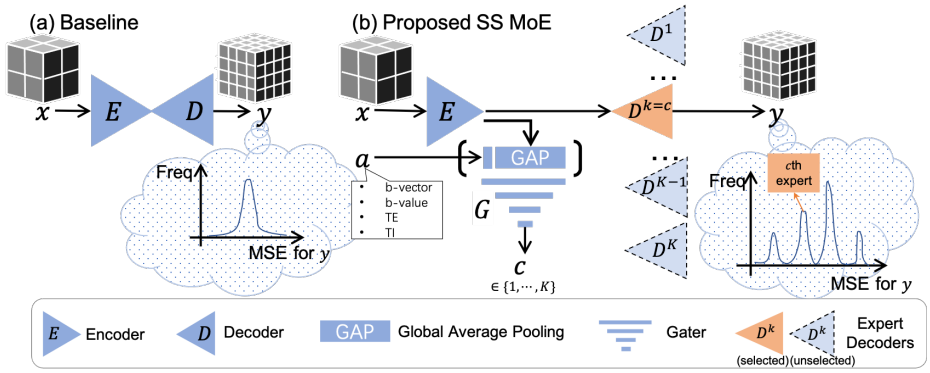

Fig. 2. Conceptual comparison of (a) the baseline network and (b) the proposed self-supervised mixture-of-expert (SS-MoE) network. The two networks are commonly built on an encoder-decoder (E-D) architecture. The gater G infers the pseudo error label c by inputting a combination of the acquisition parameter and GAP-encoded features from the input LR patch x. (a) The HR patch y is predicted from a single network that is trained on all data with diverse intensity levels. The predicted HR patch will be subject to a distribution with single peak centred at the average MSE score. (b) The network performs in two stages: we first infer the pseudo error label c for the input LR patch x. Then the HR patch outputs via the particular expert decoder D^c identified by the error class c. Under this framework, the output HR patches demonstrate multi-peak distribution which satisfies the need of quantitative super resolution with diverse intensities.

2.2 Backbone Network Architecture

Figure 2 (a) and (b) show two digraphs of the proposed backbone network, where its nodes represent the block of neural network layers and the edges are directed. Let adjacency matrices E, D, and G be the encoder, decoder, gater branches, respectively, and let $E^O := O \circ E$ output the last activation in the encoder where O denotes an adjacency matrix used to operate Global Average Pooling (GAP) to the last activation node of the encoder. The nodes in E and D consist of regular convolutional neural networks with the down- or up-sampling operations, whereas the nodes in G comprise a feedforward network with a softmax activation at the end. Given the input LR patch x a combination of E and D outputs the HR patches y:

$$y = D \circ E(x). \tag{1}$$

Moreover, we can incorporate the additional condition of the MR acquisition parameter a into the encoder features, making the error class estimation more robust in terms of the scanning process. A combination of E^O and G outputs a predictive pseudo error label c:

$$c = G \circ [E^O(x), a], \tag{2}$$

where $[\cdot]$ is a concatenation operation. This implies that x should be assigned to the cth expert network. Note that the weights of E are shared in Eqs. 1 and 2,

which enables the gater to identify the error class and then assign the identical high-level features to the corresponding expert decoder.

Here, we adopt a variant of SR U-Net in [11, 17] as the exampled encoder-decoder architecture. It is comprised of two common modules, bottleneck block right before concatenation and residual block in each level. The bottleneck block has a similar design as FSRCNN [6]. The residual block includes a skip connection over a number of "Conv+ReLU+BN" operations [10]. The detailed specifications of the backbone network are given in Fig. S1.

2.3 Self-supervised Mixture of Experts

Training Phase One: Estimate Pseudo Error Labels via Baseline Model. We first train a baseline model based on Eq.1. Given N training LR-HR patch pairs $\{(x_i, y_i)\}_{i=1}^{N}$, we optimise the weights θ, φ in terms of the baseline encoder and decoder by minimising the mean-squared-error (MSE) loss function $\mathcal{L}$:

$$\theta^*, \varphi^* = \arg\min_{\theta,\varphi} \sum_{i=1}^{N} \mathcal{L}(y_i, D_\varphi \circ E_\theta(x_i)). \tag{3}$$

The trained encoder and decoder can be used to calculate the baseline MSE score: $e_i = \mathcal{L}(y_i, D_{\varphi^*} \circ E_{\theta^*}(x_i))$.

Next, we estimate K clusters from the obtained error scores. Multiscale quantile segmentation (MQS) is a way to partition the error scores into K clusters at $K-1$ quantiles [15]. Here, we adopted MQS, denoted by $\mathcal{J}$, to presumably identify pseudo error labels $c_i \in \{1, \cdots, K\}$ mapped from e_i, i.e.

$$c_i = \mathcal{J}(e_i) = \mathcal{J}\left(\mathcal{L}(y_i, D_{\varphi^*} \circ E_{\theta^*}(x_i))\right). \tag{4}$$

The estimated c_i will be used as the ground-truth labels when training the gater G in Phase Two. We also tested on other alternative segmenters such as the empirical rule [6] or K-means clustering, but observed that the overall approach performed best using MQS.

Training Phase Two: Train Gater to Classify Input Patch for Expert Network. We adopt a supervised way to train the gater G. The detailed architecture of G is specified in the supplementary material. Given the trained encoder E_{θ^*}, the input LR patch x_i, the pseudo error class c_i, and the acquisition parameter a_i, we optimise the weights ψ of the gater by minimising the cross-entropy loss $\mathcal{L}_{CE}$:

$$\psi^* = \arg\min_{\psi} \sum_{i=1}^{N} \mathcal{L}_{CE}(c_i, G_\psi \circ [E^O_{\theta^*}(x_i), a_i]). \tag{5}$$

[6] The rule empirically selects an equispaced grid along a power-law distribution as class boundaries

Then we calculate the predictive pseudo error labels $\tilde{c}_i$ by Eq. 2, that is, $\tilde{c}_i = G_{\psi^*} \circ [E^O_{\theta^*}(x_i), a_i]$.

We finally group up LR-HR patch pairs into K subsets according to their error class indices. We denote the kth subset as $S^{(k)}$, where $k = 1, \cdots, K$, and rename the indices of LR-HR patch pairs into the kth subset as $S^{(k)} = \{(x_i^{(k)}, y_i^{(k)}) : i = 1, \cdots, N^{(k)}\}$. Each subset will be used to train an expert network in Phase Three.

Training Phase Three: Train Multiple Expert Networks with Assigned Training Patches. We freeze the encoder section and train multiple decoders with respect to the aforementioned split training subsets. Given any $S^{(k)}$ for $k = 1, \cdots, K$, we optimise the weights $\varphi^{(k)}$ of the kth decoder $D^{(k)}$ in a way similar to Eq.3:

$$\varphi^{(k)*} = \arg\min_{\varphi^{(k)}} \sum_{i=1}^{N^{(k)}} \mathcal{L}(y_i^{(k)}, D^{(k)}_{\varphi^{(k)}} \circ E_{\theta^*}(x_i^{(k)})). \tag{6}$$

Usually, we can train the decoder $D^{(k)}$ from scratch. However, when one subset has a relatively small number of data, we choose to initialise the decoder weights with the pre-trained baseline decoder weights φ^*, and then continuously train on the subset $S^{(k)}$.

Test Phase. At the test phase, we need to first predict the pseudo error label by the test LR patch $\hat{x}$ and its acquisition parameter $\hat{a}$, and then assign $\hat{x}$ to the corresponding expert network to predict the output HR patch $\hat{y}$. Specifically, we predict the pseudo error class label $\hat{c}$ by Eq.2:

$$\hat{c} = G_{\psi^*} \circ [E^O_{\theta^*}(\hat{x}), \hat{a}], \tag{7}$$

and then predict the HR output by Eq.1:

$$\hat{y} = D^{\hat{c}}_{\varphi^{\hat{c}}} \circ E_{\theta^*}(\hat{x}). \tag{8}$$

3 Experiments

3.1 Implementation details

The overall method was implemented by Tensorflow 2.0. Our program is required to run on an Nvidia GPU with at least 12 gigabyte memory. For training, we used ADAM [16] as the optimiser with a starting learning rate of 10^{-3} and a decay of 10^{-6}. We set the batch size as 64. We initialised the network weights with Glorot normal initialiser [7]. All networks required 100/20/20 training epochs respectively from Phase One to Three. All networks at Phases One and Two were trained on uniformly sampled patch pairs of around 270k [7], while at Phase

[7] Uniformly crop patches by the function `extract_patches` in scikit-learn 0.22.

Three, the expert networks fine-tuned through randomly sampled 100k pairs for speedup. MSE was used as both the loss function and the evaluation metric. We employed 5-fold cross validation to evaluate the proposed method on the Super MUDI datasets. Specifically, one of the subjects containing 1344 volumes was used for validation on a fold, and we randomly sampled 100k patch pairs out of the remaining 5376 volumes for training. We employed the statistics, such as average, variance, and quartiles, to summarise of the distribution of MSE scores over all volumes. We used a two-tailed Wilcoxon signed-rank test to determine statistical significance of the performance between any two compared methods.

3.2 Results

Since the results comprised 1344 MSE scores we used descriptive statistics to characterise the distribution of MSE scores. We conducted the comparative study over cubic spline interpolation, SR U-Net [17] as a backbone, Hard MoE [8] as a baseline model, and the proposed SS-MoE. All the neural networks had comparable model capacity. In Table 1, we observe that SS-MoE had the best performance over the others measured by average, variance, and median, and significantly reduced maximal MSE score; The full MSE score distributions are shown in Fig. S2. With Table S1, we further confirmed that SS-MoE boosted the performance of the SR U-Net backbone, outperformed over nearly all the rest methods with statistical significance ($p < 0.001$), and reduced the distribution variance. To visualise SR performance in an individual volume map, Fig. 3 compares the coronal views of different reconstructed results on the 119th volume of the last subject for $2\times$ through-plane SR task. We observe that our proposed method enhanced resolution and showed lower error score in a zoomed region.

We further analysed the effect of hyper-parameters by increasing the number of pseudo error classes as shown in Table 2. We observe a large improvement over all statistics when increasing from 2 to 4, and a modest improvement going from 4 to 8. The computational cost grew several times with more pseudo error classes since the number of network weights increased. Considering the cost-performance ratio, we recommend choosing 4 pseudo error classes for SS-MoE in this context. We infer from Fig. S3 that the mis-classified labels were mostly concentrated around the hard boundary, which implied that clusters of predictive error labels may be overlapping but may not largely degrade the performance of SS-MoE.

4 Discussion and Conclusion

We propose a novel SS-MoE paradigm for SR of multi-modal or multi-contrast imaging datasets that have diverse intensities and large variance, such as [9, 28, 29, 32]. Our SS-MoE approach can append to any baseline encoder-decoder network, allowing incorporation of state-of-the-art SR networks; in this paper we utilised the best deep neural network in the leaderboard of the Super MUDI challenge. We demonstrate that our approach reduces both errors and variances when super resolving combined diffusion-relaxometry qMRI data. The proposed

Table 1. 5-fold cross-validation results for cubic spline interpolation, SR U-Net, Hard MoE and the proposed SS-MoE both with the SR U-Net backbone. Two SR tasks were performed on the anisotropic-voxel (Aniso.) and the isotropic-voxel (Iso.) Super MUDI datasets. For each fold, we evaluated MSE scores on 1344 volumes and used their statistics (Stats.) to characterise the distribution. The mean and std of the average statistics over the 5 cross-validation folds were computed.

Dataset	Stats.	Cubic Spline Interpolation	SR U-Net [17] (Baseline)	Hard MoE [8] (MoE baseline)	SS-MoE (Proposed)
Aniso.	Average	0.744 ± 0.010	0.363 ± 0.049	0.383 ± 0.042	$\mathbf{0.305} \pm 0.048$
	Variance	14.899 ± 1.496	3.607 ± 1.105	4.230 ± 0.857	$\mathbf{2.765} \pm 0.941$
	Median	0.104 ± 0.002	0.048 ± 0.006	0.045 ± 0.008	$\mathbf{0.041} \pm 0.006$
	Max	97.28 ± 12.02	50.45 ± 6.77	55.24 ± 4.64	$\mathbf{43.59} \pm 5.94.$
Iso.	Average	1.583 ± 0.014	0.658 ± 0.075	0.717 ± 0.088	$\mathbf{0.648} \pm 0.036$
	Variance	62.905 ± 4.2393	13.927 ± 3.893	17.289 ± 5.353	$\mathbf{13.829} \pm 4.064$
	Median	0.2351 ± 0.0029	0.0781 ± 0.010	0.083 ± 0.009	$\mathbf{0.075} \pm 0.010$
	Max	194.89 ± 18.67	96.64 ± 11.12	108.91 ± 14.27	$\mathbf{96.09} \pm 11.01$

Table 2. Statistics of MSE-score distribution on the isotropic-voxel Super MUDI dataset v.s. the number of pseudo error classes (#Classes) in SS-MoE. The number of network Weights (#Weights) are given. All the experiments were validated on the setup of the first cross-validation fold that was used to train the SS-MoE model and predict on the same 1344 volumes.

#Classes	#Weights	Average	Variance	Min	Q1	Median	Q3	Max
2	4.42×10^6	0.5593	9.8079	0.0032	0.0230	0.0664	0.2297	85.0411
4	8.76×10^6	0.5537	9.5210	0.0032	0.0228	0.0653	0.2294	83.8385
8	1.74×10^7	0.5527	9.4552	0.0031	0.0228	0.0649	0.2282	83.3759

SS-MoE performed better than MoE due to convex loss function enabling robust training and memory footprint independent to the number of experts [20].

The SS-MoE paradigm also provides a way for future improvement and application. First, the pseudo error labels are estimated and then predicted in a self-supervised way, and hence the gater used to predict them may highly depend on the segmenters, such as MQS, and how good the baseline model is. This may limit the entire performance of SS-MoE, especially for the generalisablility of the gater; see the supplementary material. Automatically discriminating the inputs without using the baseline model like the idea in [25] will be valuable to explore. On the other hand, our method has the potential to super resolve a variety of qMRI data types, ultimately accelerating the acquisition process and increasing clinical viability. In future work, we will investigate if super resolved images offer better visualisation of pathologies, such as lesions or tumours.

Acknowledgements

This work was supported by EPSRC grants EP/M020533/1, EP/R014019/1, and EP/V034537/1 as well as the NIHR UCLH Biomedical Research Centre.

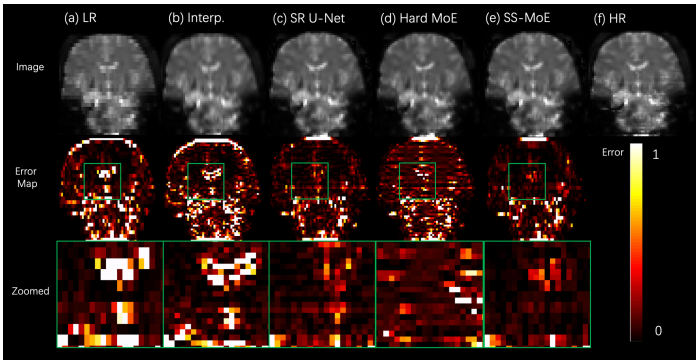

Fig. 3. Visual comparison of the coronal views by (a) LR, (b) Interpolation (Interp.), (c) the variant SR U-Net backbone, (d) the Hard MoE baseline, (e) the proposed SS-MoE, and (f) HR images on the 119th volume of the last subject for 2× through-plane SR task. The error maps are normalised square difference between the reconstructed volumes and the HR volumes for each voxel. Zoomed regions of the error maps are highlighted.

References

1. Cdmri supermudi challenge 2020, `https://www.developingbrain.co.uk/data/`
2. Alexander, D.C., Zikic, D., Ghosh, A., et al.: Image quality transfer and applications in diffusion mri. NeuroImage **152**, 283–298 (2017)
3. Blumberg, S.B., Tanno, R., Kokkinos, I., Alexander, D.C.: Deeper image quality transfer: Training low-memory neural networks for 3D images. In: MICCAI. pp. 118–125 (2018)
4. Chen, G., Dong, B., Zhang, Y., Lin, W., Shen, D., Yap, P.T.: XQ-SR: Joint x-q space super-resolution with application to infant diffusion MRI. Medical Image Analysis **57**, 44–55 (2019)
5. Chen, Y., Shi, F., Christodoulou, A.G., Xie, Y., Zhou, Z., Li, D.: Efficient and Accurate MRI Super-Resolution Using a Generative Adversarial Network and 3D Multi-level Densely Connected Network. In: MICCAI, vol. 11070 LNCS, pp. 91–99 (2018)
6. Dong, C., Loy, C.C., Tang, X.: Accelerating the super-resolution convolutional neural network. In: ECCV. pp. 391–407. Springer International Publishing, Cham (2016)
7. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp. 249–256 (2010)
8. Gross, S., Ranzato, M., Szlam, A.: Hard Mixtures of Experts for Large Scale Weakly Supervised Vision. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). vol. 2017-Janua, pp. 5085–5093. IEEE (2017)
9. Grussu, F., Battiston, M., Veraart, J., Schneider, T., Cohen-Adad, J., Shepherd, T.M., Alexander, D.C., Fieremans, E., Novikov, D.S., Gandini Wheeler-Kingshott, C.A.: Multi-parametric quantitative in vivo spinal cord MRI with unified signal readout and image denoising. NeuroImage **217**(March), 116884 (2020)

10. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778. IEEE (2016)
11. Heinrich, L., Bogovic, J.A., Saalfeld, S.: Deep learning for isotropic super-resolution from non-isotropic 3d electron microscopy. In: MICCAI. pp. 135–143 (2017)
12. Hutter, J., Slator, P.J., Christiaens, D., et al.: Integrated and efficient diffusion-relaxometry using ZEBRA. Scientific Reports **8**(1), 1–13 (2018)
13. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 37, pp. 448–456. PMLR, Lille, France (2015)
14. Jacobs, R.A., Jordan, M.I., Nowlan, S.E., Hinton, G.E.: Adaptive mixture of experts (1991)
15. Jula Vanegas, L., Behr, M., Munk, A.: Multiscale Quantile Segmentation. Journal of the American Statistical Association pp. 1–14 (jan 2021)
16. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
17. Lin, H., Figini, M., Tanno, R., et al.: Deep learning for low-field to high-field mr: Image quality transfer with probabilistic decimation simulator. In: International Workshop on Machine Learning for Medical Image Reconstruction. pp. 58–70 (2019)
18. Lyu, Q., Shan, H., Steber, C., Helis, C., Whitlow, C., Chan, M., Wang, G.: Multi-Contrast Super-Resolution MRI Through a Progressive Network. IEEE transactions on medical imaging **39**(9), 2738–2749 (2020)
19. Ma, J., Yu, J., Liu, S., et al.: PathSRGAN: Multi-Supervised Super-Resolution for Cytopathological Images Using Generative Adversarial Network. IEEE Transactions on Medical Imaging **39**(9), 2920–2930 (sep 2020)
20. Makkuva, A., Viswanath, P., Kannan, S., Oh, S.: Breaking the gridlock in mixture-of-experts: Consistent and efficient algorithms. In: Proceedings of the 36th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 97, pp. 4304–4313. PMLR (2019)
21. Pizzolato, M., Palombo, M., Bonet-Carne, E., et al.: Acquiring and Predicting Multidimensional Diffusion (MUDI) Data: An Open Challenge. In: Computational Diffusion MRI. pp. 195–208. Springer International Publishing, Cham (2020)
22. Reinhold, J.C., Dewey, B.E., Carass, A., Prince, J.L.: Evaluating the impact of intensity normalization on MR image synthesis. In: Medical Imaging 2019: Image Processing. p. 126. SPIE (2019)
23. Shen, S., Yao, Z., Gholami, A., Mahoney, M., Keutzer, K.: PowerNorm: Rethinking batch normalization in transformers. In: Proceedings of the 37th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 119, pp. 8741–8751. PMLR (2020)
24. Shi, Y., Siddharth, N., Paige, B., Torr, P.H.S.: Variational Mixture-of-Experts Autoencoders for Multi-Modal Deep Generative Models (NeurIPS) (2019)
25. Tanno, R., Arulkumaran, K., Alexander, D., Criminisi, A., Nori, A.: Adaptive neural trees. In: Proceedings of the 36th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 97, pp. 6166–6175. PMLR (2019)
26. Tanno, R., Worrall, D.E., Ghosh, A., et al.: Bayesian image quality transfer with CNNs: Exploring uncertainty in dMRI super-resolution. In: MICCAI. pp. 611–619 (2017)

27. Tanno, R., Worrall, D.E., Kaden, E., et al.: Uncertainty modelling in deep learning for safer neuroimage enhancement: Demonstration in diffusion mri. NeuroImage **225**, 117366 (2020)
28. Tong, Q., He, H., Gong, T., Li, C., Liang, P., Qian, T., Sun, Y., Ding, Q., Li, K., Zhong, J.: Multicenter dataset of multi-shell diffusion MRI in healthy traveling adults with identical settings. Scientific Data **7**(1), 1–7 (2020)
29. Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K.: The wu-minn human connectome project: An overview. NeuroImage **80**, 62–79 (2013)
30. Van Steenkiste, G., Poot, D.H., Jeurissen, B., et al.: Super-resolution T1 estimation: Quantitative high resolution T1 mapping from a set of low resolution T1-weighted images with different slice orientations. Magnetic Resonance in Medicine **77**(5), 1818–1830 (2017)
31. Wilson, A.G., Izmailov, P.: Bayesian deep learning and a probabilistic perspective of generalization. In: Advances in Neural Information Processing Systems. vol. 33, pp. 4697–4708. Curran Associates, Inc. (2020)
32. Zhang, R., Garrett, J., Ge, Y., Ji, X., Chen, G.H., Li, K.: Design, construction, and initial results of a prototype multi-contrast x-ray breast imaging system. In: Medical Imaging 2018: Physics of Medical Imaging. vol. 176, p. 31. SPIE (2018)
33. Zhang, Y., Yap, P.T., Chen, G., Lin, W., Wang, L., Shen, D.: Super-resolution reconstruction of neonatal brain magnetic resonance images via residual structured sparse representation. Medical Image Analysis **55**, 76–87 (2019)
34. Zhao, C., Shao, M., Carass, A., et al.: Applications of a deep learning method for anti-aliasing and super-resolution in mri. Magnetic resonance imaging **64**, 132–141 (2019)
35. Zheng, Z., Yuan, C., Zhu, X., Lin, Z., Cheng, Y., Shi, C., Ye, J.: Self-Supervised Mixture-of-Experts by Uncertainty Estimation. Proceedings of the AAAI Conference on Artificial Intelligence **33**, 5933–5940 (2019)