# 1 | SUPPORTING INFORMATION

## 1.1 | Details of weighted learning and enhancement in restricted diffusion

The ADC estimate is very dependent on the accuracy of the high-b signal average because it is obtained via the following mono-exponential signal decay model:

$$S(\mathbf{x}, b) = S(\mathbf{x}, 0) exp(-b \times ADC) \tag{S.1}$$

The anomalously low signals due to motion-induced signal loss make the ADC value estimate unreasonably high since the ADC estimate is obtained via least squares fit. Moreover, the variation of signal acquisitions in high b-value signal reduces the contrast between high signals (due to restricted diffusion and possibly cancer) and low signals (due to healthy tissue), which reduces the discrimination between ADC values. The weights $w_k(\mathbf{x}_i)$ are set to take values proportional to the underlying signal intensities:

$$w_k(\mathbf{x}_i) = \exp(\frac{S_k(\mathbf{x}_i, b)}{\tau}), \tag{S.2}$$

where $\tau$ is the "temperature" parameter determining the entropy of the weight distribution. By using these weights in training, the INR converges to the weighted sum of acquisitions, similar to the acquisition-weighting rule of[1]:

$$\begin{aligned} \text{INR}^*(\mathbf{x}, b) &\approx \sum_k \frac{\exp(S_k(\mathbf{x}, b)/\tau)}{\sum_j \exp(S_j(\mathbf{x}, b)/\tau)} S_k(\mathbf{x}, b) \\ &= \sum_k a_k(\mathbf{x}, b) S_k(\mathbf{x}, b) \end{aligned} \tag{S.3}$$

where $a_k(\mathbf{x}, b)$ can be considered as a scalar that represents an estimate of the probability that the given measurement at acquisition $k$ accurately reflects the true signal for the voxel $\mathbf{x}$. The "temperature" parameter $\tau$ is set to vary for the water diffusion behavior for each region. Measurement of the restricted diffusion was employed using the ADC measurement for each voxel. For a voxel with less restricted diffusion of water, i.e. less probability of cancer, the temperature is set to a high value. Consequently, the weight $w_k(\mathbf{x})$ approaches 1 and $a_k$ approaches the uniform distribution. For a voxel with highly restricted diffusion, i.e. higher probability of cancer, the temperature is reduced, and consequently $a_k$ acts as the soft-max function. *For such voxels, if some of the acquisitions suffer from motion-induced signal loss, their contribution is suppressed automatically via $a_k$, even if the number of such acquisitions is high or the signal loss in those acquisitions is drastically high.* The function INR (and the jointly trained PN) minimizes the loss function in equation (6) to yield a derivative of zero at one of the optima in the space of the parameter sets of INR (and PN). If we call this optimum parameter set $\theta^*$ and the function with these parameters INR$^*$, and calculate the derivative of the loss, evaluated for a training point $\mathbf{x}$ we get:

$$\begin{aligned} &\frac{d}{d\theta}\mathcal{L}\Big|_{\theta=\theta^*,\mathbf{x}_i=\mathbf{x}} \approx 0 \\ &\frac{d}{d\theta}\sum_k w_k(\mathbf{x})||S_k(\mathbf{x}) - \mathrm{INR}^*(\mathbf{x}+\boldsymbol{\epsilon}_{x,k})||^2 \approx 0 \end{aligned} \tag{S.4}$$

To present the notation in a simpler form, we denote the output of the perturbation net as $\boldsymbol{\epsilon}_{x,k} = \epsilon_{max}\texttt{tanh}(\mathrm{PN}(\mathbf{x},k))$ and omit the b-direction in $S_k(\mathbf{x},b)$

$$\begin{aligned} &\sum_k w_k(\mathbf{x})(2)(S_k(\mathbf{x}) - \mathrm{INR}^*(\mathbf{x}+\boldsymbol{\epsilon}_{x,k})) \\ &(-1)\frac{d}{d\theta}\mathrm{INR}^*(\mathbf{x}+\boldsymbol{\epsilon}_{x,k}) \approx 0 \end{aligned} \tag{S.5}$$

For the sake of deriving this calculation we can safely assume that the term $\frac{d}{d\theta}\mathrm{INR}^*(\mathbf{x}+\boldsymbol{\epsilon}_{x,k})$ is non-zero, except for some occasional voxels, if any. Therefore, at an optimum:

$$\sum_k w_k(\mathbf{x})\mathrm{INR}^*(\mathbf{x}+\boldsymbol{\epsilon}_{x,k}) \approx \sum_k w_k(\mathbf{x})S_k(\mathbf{x}) \tag{S.6}$$

We can now conduct an ablation analysis and interpret equation (S.6) in three scenarios of training cost we have introduced in this paper:

1. If the input perturbation net is not in effect and there is no acquisition weighting, i.e equation (3) is used, then $\boldsymbol{\epsilon}_{x,k} = 0$ and $w_k = 1$ and Equation (S.6) becomes:

$$\begin{aligned} \sum_k \mathrm{INR}^*(\mathbf{x}) &\approx \sum_k S_k(\mathbf{x}) \\ \mathrm{INR}^*(\mathbf{x}) &\approx \frac{1}{K}\sum_k S_k(\mathbf{x}), \end{aligned} \tag{S.7}$$

   which corresponds to the mean DWI, for the values in the training set.

2. If the input perturbation net is not in effect and the acquisition weighting is used as proposed in equation (7):

$$\begin{aligned} \sum_k w_k(\mathbf{x})\mathrm{INR}^*(\mathbf{x}) &\approx \sum_k w_k(\mathbf{x})S_k(\mathbf{x}) \\ \mathrm{INR}^*(\mathbf{x}) &\approx \sum_k \frac{w_k(\mathbf{x})}{\sum_j w_j(\mathbf{x})}S_k(\mathbf{x}) \\ &= \sum_k a_k(\mathbf{x})S_k(\mathbf{x}) \end{aligned} \tag{S.8}$$

   which yields the equation (8) that yields the DW image enhancement technique that emphasizes regions with high restriction of diffusion as proposed in [1].

3. In order to be able to interpret the functionality of the perturbation network, for the sake of simplicity, let us first assume that there is no acquisition weighting and the input perturbation is locally transitional, i.e. $\mathrm{PN}(\mathbf{x},k) \approx \mathrm{PN}(\mathbf{x}+\mathrm{PN}(\mathbf{x}),k)$. This is a fair assumption in very small local regions (i.e. less than a voxel size), even if the overall image warping becomes non-linear. It should be noted that by minimizing the loss function in equation (6), INR is never explicitly trained to

estimate the signal values for $\mathbf{x}$ that exist in the training set. Instead, they are trained with $S_k(\mathbf{x})$ to infer the signal values of $\mathbf{x} + \boldsymbol{\epsilon}_{x,k}$. If we rearrange equation (S.6) for $w_k(\mathbf{x}) = 1 \quad \forall k$

$$\sum_k \left( S_k(\mathbf{x}) - \text{INR}^*(\mathbf{x} + \boldsymbol{\epsilon}_{x,k}) \right) \approx 0 \tag{S.9}$$

since, both $S_k(\mathbf{x})$ and $\text{INR}^*(\mathbf{x} + \boldsymbol{\epsilon}_{x,k})$ are non-negative quantities, then (S.9) entails that

$$S_k(\mathbf{x}) - \text{INR}^*(\mathbf{x} + \boldsymbol{\epsilon}_{x,k}) \approx 0 \quad \forall k, \tag{S.10}$$

At this optimum of INR, the cost is also expected to be minimized for other voxel locations that don't exist in the training set. If we take $\mathbf{y} = \mathbf{x} - \boldsymbol{\epsilon}_{x,k}$ as the input location,

$$S_k(\mathbf{y}) - \text{INR}^*(\mathbf{y} + \boldsymbol{\epsilon}_{y,k}) \approx 0 \quad \forall k, \tag{S.11}$$

can be rewritten as

$$S_k(\mathbf{x} - \boldsymbol{\epsilon}_{\mathbf{x},\mathbf{k}}) - \text{INR}^*(\mathbf{x} - \boldsymbol{\epsilon}_{\mathbf{x},\mathbf{k}} + \boldsymbol{\epsilon}_{y,k}) \approx 0 \quad \forall k, \tag{S.12}$$

Since we assume $\text{PN}(\mathbf{x}, k) \approx \text{PN}(\mathbf{x} + \text{PN}(\mathbf{x}), k)$, $\boldsymbol{\epsilon}_{x,k} \approx \boldsymbol{\epsilon}_{y,k}$, therefore (S.12) becomes

$$S_k(\mathbf{x} - \boldsymbol{\epsilon}_{\mathbf{x},\mathbf{k}}) - \text{INR}^*(\mathbf{x}) \approx 0 \quad \forall k, \tag{S.13}$$

Obviously, $S_k(\mathbf{x} - \boldsymbol{\epsilon}_{\mathbf{x},\mathbf{k}})$ does not physically exist, as $S_k$ is discrete and $\boldsymbol{\epsilon}_{x,k}$ is upper bounded to be less than a voxel size. Nevertheless, we adopt this notation to present a different view to the geometries of the proposed method. If we plug the (S.13) to the global cost in (S.9), the compromise value of $\text{INR}^*(\mathbf{x})$ becomes

$$\text{INR}^*(\mathbf{x}) \approx \frac{1}{K} \sum_k S_k(\mathbf{x} - \boldsymbol{\epsilon}_{\mathbf{x},\mathbf{k}}), \tag{S.14}$$

which is equivalent to the multi-frame super-resolution technique similar to the works in[2,3] and[4], where several acquisitions interspersed with sub-pixel shifts are combined. Except, of course, our model provides non-linear registration to each acquisition, and this registration is jointly learned by the parameters of the PN.

## 1.2 | Design of the Temperature Parameter

As it was shown in[1], inter-acquisition variation gets larger and more left-skewed in the areas with more restricted diffusion, i.e. cancer. Therefore, we propose adjusting the $\tau$ parameter as a function of the apparent diffusion coefficient of the underlying voxel such that $a_k(\mathbf{x})$ tends to suppress acquisitions with lossy signals more in the regions which are highly likely to exhibit restricted diffusion of water, than other regions. Therefore, we use the following monotonically increasing function to decide $\tau$:

$$\tau = a(1 + \tanh(b(ADC - c))), \tag{S.15}$$

The interpretation that goes into how this function was used to design the $\tau$ parameter was demonstrated in Figure S1 . The parameter $a$ is chosen as a big enough number to avoid floating point errors that could be caused by the exponentiation in equation (6). $b$ and $c$ are particularly chosen such that the yellow zone in Figure S1 corresponds to the range of ADC values evaluated for prostate cancer, i.e. $0.86 \pm 0.18 \times 10^{-3} mm^2/s$ as reported in [5].

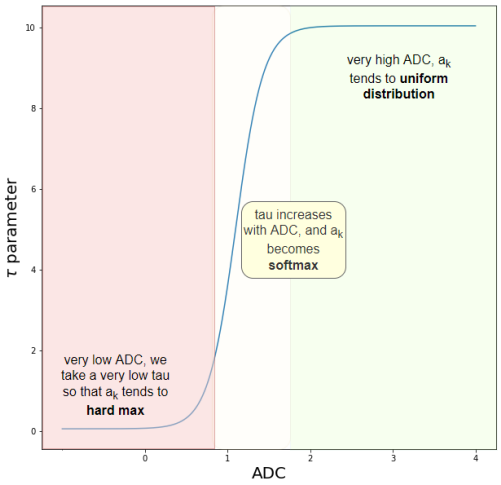



**FIGURE S1** Design of the $\tau$ parameter.

## 1.3 | Similarity Comparisons

In order to evaluate the improvement in the high-frequency components that SR provides we applied the high pass filter whose frequency response is given in Figure S2 to all SR, bi-cubic and GT images, to remove the low spatial frequency components. In addition to Table I, the superiority of SR in all metrics can be observed better in the scatter plots in Figure S3 .

## 1.4 | More In-vivo Examples

More through-plane examples can be seen in Figures S4 , S5 , and S6 . More in-plane examples can be seen in Figures S7 , S8 , S9 and S10 .

## AVAILABILITY OF MATERIALS AND DATA

The source code of this work is publicly available and accessible via this GitHub repository. The data supporting the findings of this study (the motion-simulated, static and very high resolution phantom scans as well as one de-identified prostate diffusion image) are also made accessible in this Box folder.

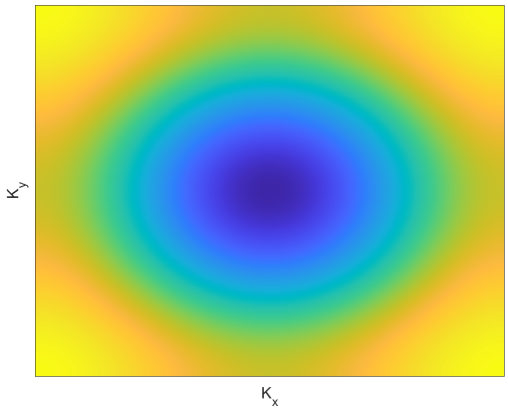



**FIGURE S2** Frequency response of the high pass filter applied to suppress low resolution components and focus on high resolution components

## References

1. Gundogdu B, Pittman JM, Chatterjee A, et al. Directional and inter-acquisition variability in diffusion-weighted imaging and editing for restricted diffusion. *Magnetic Resonance in Medicine* 2022.

2. Kennedy JA, Israel O, Frenkel A, Bar-Shalom R, Azhari H. Improved image fusion in PET/CT using hybrid image reconstruction and super-resolution. *International journal of biomedical imaging* 2007; 2007.

3. Vis G, Nilsson M, Westin CF, Szczepankiewicz F. Accuracy and precision in super-resolution MRI: Enabling spherical tensor diffusion encoding at ultra-high b-values and high resolution. *NeuroImage* 2021; 245: 118673.

4. Mahmoudzadeh AP, Kashou NH. Interpolation-based super-resolution reconstruction: effects of slice thickness. *Journal of Medical Imaging* 2014; 1(3): 034007.

5. Chatterjee A, Bourne RM, Wang S, et al. Diagnosis of prostate cancer with noninvasive estimation of prostate tissue composition by using hybrid multidimensional MR imaging: a feasibility study. *Radiology* 2018; 287(3): 864.

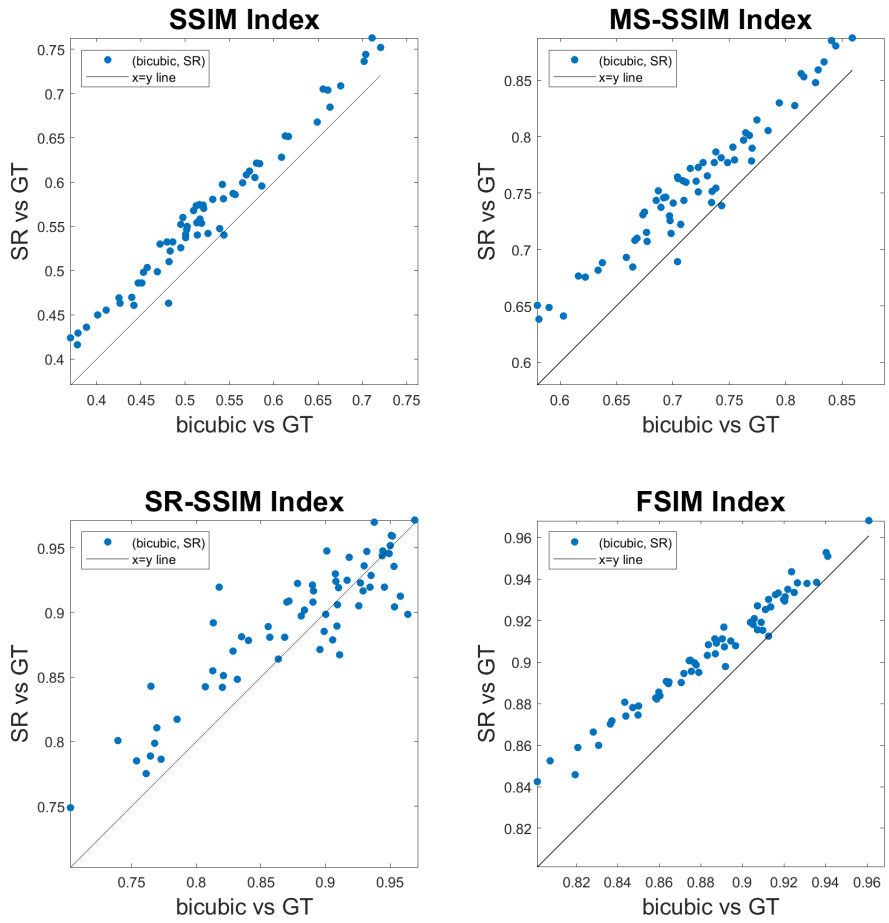
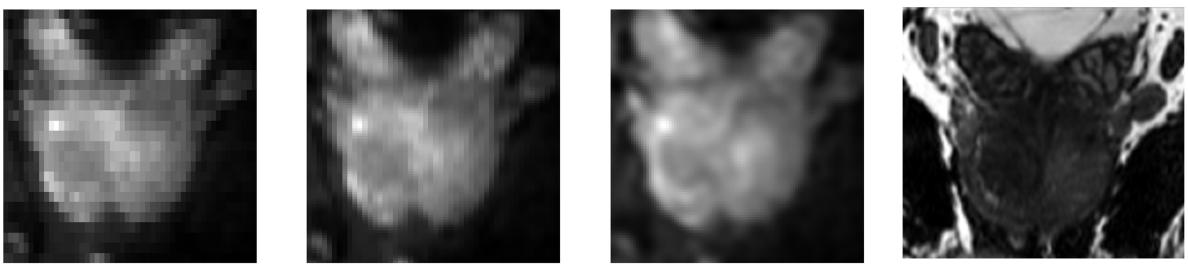



**FIGURE S3** Scatter plot that compares SR against bicubic interpolation in several structural similarity metrics. It can be seen that the superiority of SR is more emphasized especially for the images with lower LR-HR similarity scores.

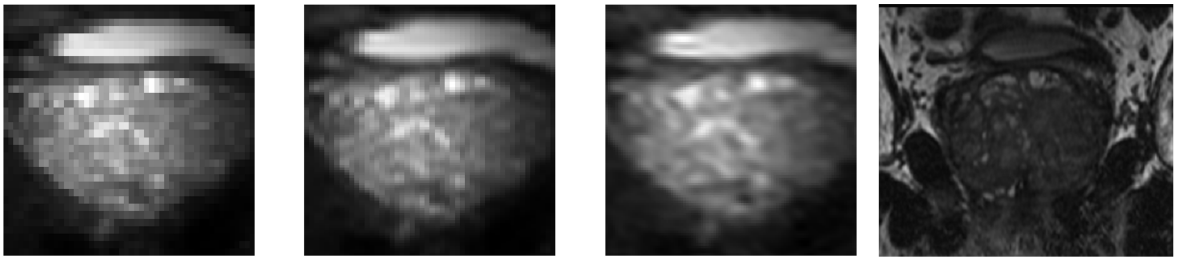


(a) DWI coronal view (b) Interpolation along z-axis (c) INR output (d) T2W Coronal image

**FIGURE S4** An example of through-plane super-resolution of DWI (coronal view)

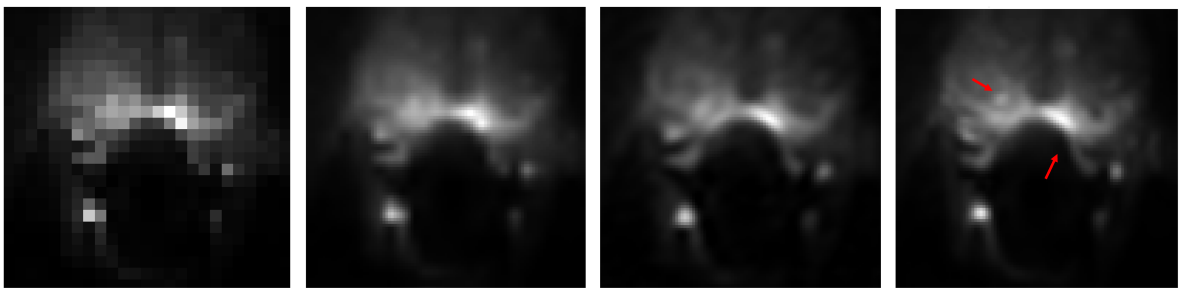
(a) DWI coronal view (b) Interpolation along z-axis (c) INR output (d) T2W Coronal image

**FIGURE S5** An example of through-plane super-resolution of DWI (coronal view)



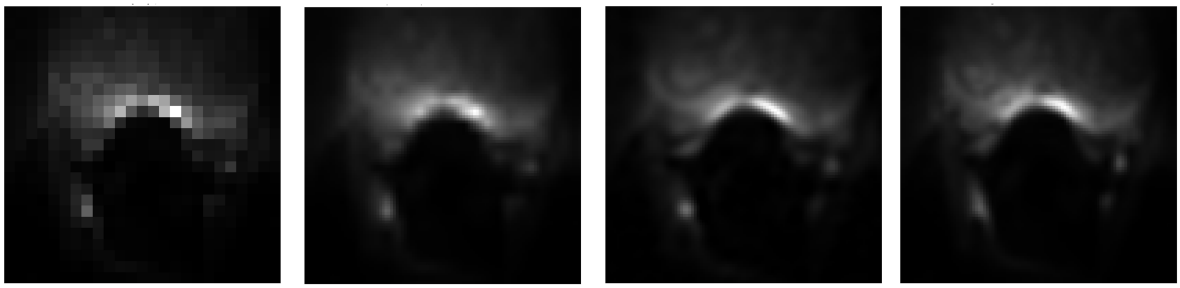
(a) DWI coronal view (b) Interpolation along z-axis (c) INR output (d) T2W Coronal image

**FIGURE S6** An example of through-plane super-resolution of DWI (coronal view)



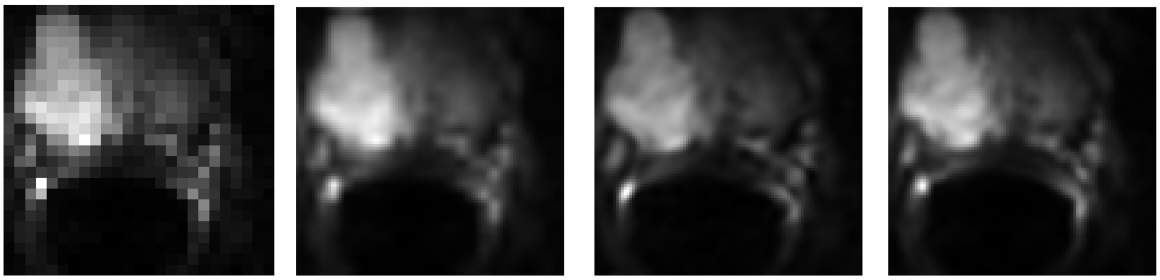
(a) LR (b) bi-cubic interpolation (c) INR output (d) HR ground truth

**FIGURE S7** An example of in-plane super-resolution of DWI (axial view)



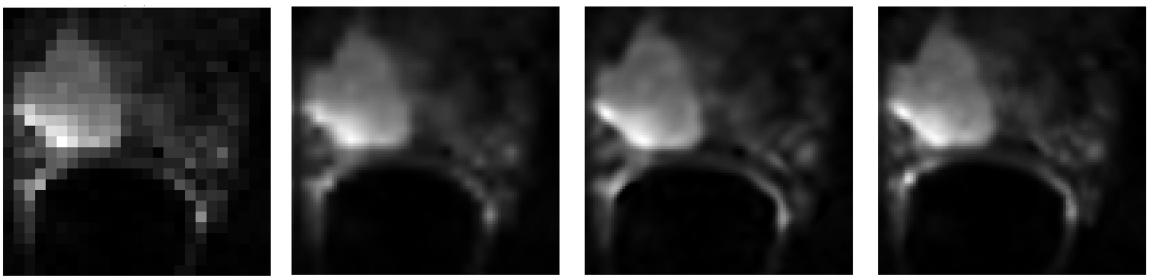
(a) LR (b) bi-cubic interpolation (c) INR output (d) HR ground truth

**FIGURE S8** An example of in-plane super-resolution of DWI (axial view)



(a) LR (b) bi-cubic interpolation (c) INR output (d) HR ground truth

**FIGURE S9** An example of in-plane super-resolution of DWI (axial view)



(a) LR (b) bi-cubic interpolation (c) INR output (d) HR ground truth

**FIGURE S10** An example of in-plane super-resolution of DWI (axial view)