Ultra-low-field paediatric MRI in low- and middle-income countries: super-resolution using a multi-orientation U-Net

Levente Baljer[1], Yiqi Zhang[1], Niall J Bourke[1], Kirsten A Donald[2], Layla E Bradford[2], Jessica E Ringshaw[1,2], Simone R Williams[2], Sean CL Deoni[3], Steven CR Williams[1], Khula SA Study Team[2], František Váša[1,*], Rosalyn J Moran[1,*]

[1] Department of Neuroimaging, Kings College London, UK
[2] Department of Paediatrics and Child Health, Red Cross War Memorial Children's Hospital, University of Cape Town, South Africa
[3] Maternal Newborn and Child Nutrition and Health (MNCNH), Bill & Melinda Gates Foundation, Seattle, Washington, USA

[*]Joint senior authors

Abstract

Super-resolution has been applied to ultra-low-field MRI, however it is yet to be applied specifically to paediatric ultra-low-field imaging. Owing to the high cost of modern MRI systems, their use in clinical care and neurodevelopmental research is limited to hospitals and universities in high income countries. Ultra-low-field systems with significantly lower scanning costs bear the potential for global adoption, however their reduced SNR compared to 1.5 or 3T systems limits their applicability for research and clinical use. In this paper, we describe a deep learning-based super-resolution approach to generate high-resolution isotropic T_2-weighted scans from low-resolution inputs. We train a 'multi-orientation U-Net', which uses multiple low-resolution anisotropic images acquired in orthogonal orientations to construct a super-resolved output. Our approach exhibits improved quality of outputs compared to current state-of-the-art methods for super-resolution of ultra-low-field scans in paediatric populations. Crucially for paediatric development, our approach improves reconstruction on deep brain structures for all measured regions with the greatest improvement in the caudate, where Spearman's correlation coefficient, Md [Q1, Q3], between model outputs and high-field targets increases from 0.75 [0.64, 0.81] (current state-of-the-art) to 0.90, [0.86, 0.93] (U-Net). Our research serves as proof-of-principle of the viability of training deep-learning based super-resolution models for use in neurodevelopmental research and presents the first U-Net trained exclusively on paired ultra-low-field and high-field data from infants.

1 Introduction

Neuroimaging studies on infant and child populations have become increasingly vital in establishing the relationship between neurodevelopment and resultant cognitive functioning. Magnetic resonance imaging (MRI) has proven to be particularly essential in this endeavour, owing to its ability to provide insight into structural, functional, and metabolic brain development, in addition to recognising pathologies pertaining to neurobiological disorders (Nolen-Hoeksema, 2014). Such powerful capabilities are inherently coupled with major financial costs, however: on top of the cost of a scanner itself (roughly $1,000,000 per Tesla (Arnold, 2023)), the high field strength of existing systems requires facilities with sufficient electromagnetic shielding and specialised staffing. Furthermore, since most modern MRI systems use superconducting magnets, they require the use of cryogens, which themselves come with storage, transportation and maintenance costs (Sarracanie et al., 2015). Altogether, these expenses establish strict financial boundaries on neuroimaging studies and clinical work. Even in hospitals and universities within high-income countries (HICs), scanning costs ($500-$1000/hr per research scan) place a limit on the number of participants scanned and the duration of longitudinal research.

More concerningly from a global health perspective, infrastructural costs severely limit MRI accessibility in low- and middle-income countries (LMICs). Evaluating MRI accessibility based on ratio of MRI units per million population, HICs such as Japan and the US boast 51.67 and 38.96 units/million, respectively (Ogbole et al., 2018). Comparing these values with those of Nigeria and Ghana (which own 0.30 and 0.48 units/million, respectively) reflects a hundred-fold difference between HICs and LMICs (Jalloul et al., 2023). Consequently, our current understanding of neurodevelopment in such regions, where adversities pertaining to nutrition, sanitation and higher rates of infectious risk play a crucial role, is primarily based on psychometric measures or low-cost, functional imaging methods such as EEG or fNIRS (Perdue et al., 2019, Jensen et al., 2021).

Ultra-low-field (ULF) imaging systems with magnetic field strengths ranging from 50 to 100mT (e.g., the 64mT Hyperfine Swoop) present a potential solution to issues of access inequality (Sarracanie et al., 2015). Such systems rely on the use of large, permanent magnets instead of superconducting electromagnets, and as such have significantly reduced component

prices, reduced room requirements, and reduced costs for power, cooling and maintenance (Campbell-Washburn et al., 2019). Unfortunately, these benefits are paired with a significantly diminished signal-to-noise ratio (SNR) (Klein, 2020), rendering image outputs sub-optimal for visual reading by radiologists or for a large portion of automated processing methods in currently available neuroimaging toolkits. Such processing methods usually require high-resolution MRI scans (1mm isotropic), whereas product sequences on the Hyperfine system yield a default spatial resolution of 1.5×1.5×5 mm. Higher resolution images may be acquired with increased acquisition times (∼12–15 min for a single 2×2×2mm T_2-weighted image), however the former approach is considered more clinically feasible in sensitive populations - such as infants or elderly participants - as acquisition time is reduced to approximately 3-6 minutes per image.

One technique that has found success in enhancing the SNR of ULF outputs is super-resolution (SR) reconstruction of a single image from multiple lower-resolution images, acquired in three orthogonal orientations (i.e., axial, sagittal, and coronal) (Deoni et al., 2022). This is carried out through repeated multi-resolution registration (MRR) of the scans using the Advanced Normalization Tools (ANTs) multivariate template construction tool, outputting a final image with effective dimensions of 1.5×1.5×1.5mm (Figure 1).

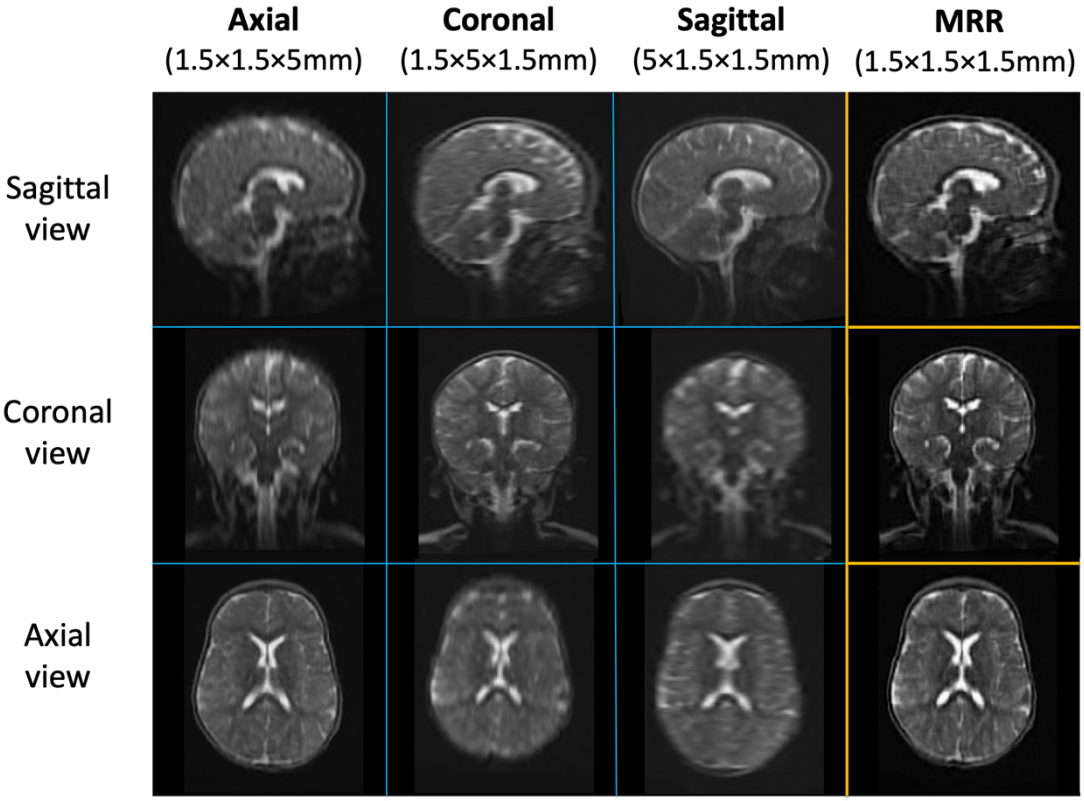

Figure 1) Multi-Resolution Registration (MRR) - Columns 1-3: non-isotropic Hyperfine scans of a 6-month-old subject (1.5 x 1.5 x 5 mm, with high-resolution plane along each of 3 orthogonal orientations); Column 4: isotropic output of MRR (1.5 x 1.5 x 1.5 mm)

An alternative approach involves the use of modern image synthesis methods based on convolutional neural networks (CNNs), which learn a transformation between a source image and a target image. A decade of research has gone into the use of CNNs for super-resolution, with the earliest models using as few as three-layers for patch extraction, nonlinear mapping, and image reconstruction (Dong et al., 2014). In the years following, architectures gradually increased in complexity as researchers introduced deeper models (Kim et al., 2015) with added functionality, such as sub-pixel convolution (Shi et al., 2016). Although initially developed for single-image super-resolution (SISR), such networks quickly found success in the context of super-resolution of 3D MRI data (Oktay, 2016, Pham et al., 2019).

In contrast, the U-Net was originally developed for semantic segmentation of biomedical image data (Ronneberger, 2015) and since its inception in 2015 it has remained one of the most popular models for use in this field. The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization of biological structures. Through the contracting path, spatial information is normally lost,

however skip connections between the two paths ensure feature reusability. Owing to these design choices, use of the U-Net has become ubiquitous in SR tasks. They have been applied to super-resolution of MRI images in their standard form using synthetic training data (Iglesias et al., 2021, 2023), or with an attached discriminator to learn via a Generative Adversarial Network (GAN)-style training scheme (Zhou, 2022). The latter approach has been investigated with several modifications, such as the integration of a VGG-based perceptual loss (Zhang, 2022) or the exchange of U-Net downsampling and upsampling blocks with ResNet counterparts (Islam, 2023). More recently, within the realm of SISR, U-Nets have also been supplemented with vision transformers (Zhu, 2023) and diffusion processes (Wu, 2023). Despite all these research efforts, deep learning-based SR has only been investigated in adult populations, where in many cases models are being trained on partially or even entirely synthetic datasets. As such, here we propose the first U-Net trained on real, paired ULF-HF dataset from a paediatric population, and demonstrate its ability to surpass MRR in quality of SR outputs.

2 Methods

2.1 Imaging data

MRI data used in this paper stems from a study investigating the neurodevelopment of executive function over the first 3 years of life in participants based in South Africa and Malawi (Zieff, 2024). Images used here included subjects from South Africa exclusively, at ages of either 3- or 6-months, with no known neurological abnormalities. A total of 82 subjects attended two scanning sessions (ULF and HF); all subjects had HF T_2-weighted scans acquired using a Siemens 3T scanner and had ULF T_2-weighted scans acquired using a Hyperfine Swoop, with high in-plane resolution along three orthogonal planes (axial, sagittal, coronal). Of these subjects, 63 successfully completed all four scanning protocols (one HF and three ULF scans). Pre-processing involved rigid registration between all HF scans and a custom age-specific HF template, and each subject's ULF with the corresponding pre-registered HF scan (Ashburner, 2007). Following this, we skull-stripped HF and ULF scans separately using HD-BET, a deep-learning tool for MRI brain extraction (Isensee et al., 2019). As seven subjects had poor image outputs following pre-processing steps, our final sample size consisted of N=56 participants (26 Male, 30 Female), of which 16 were scanned at 3-months, and 40 at 6-months. Each subject had three ULF (64mT; 1.5×1.5×5 mm) and one HF (3T; 1×1×1×mm) T_2-weighted scans. To maximise the data available from a single subject, we built a network that used all

three ULF inputs (Hyperfine scans acquired in three orthogonal orientations), thus each matched 'pair' of ULF-HF scans included three orthogonal ULF scans and one HF scan. The full pre-processing pipeline and exclusion of images is shown in Figure 2.

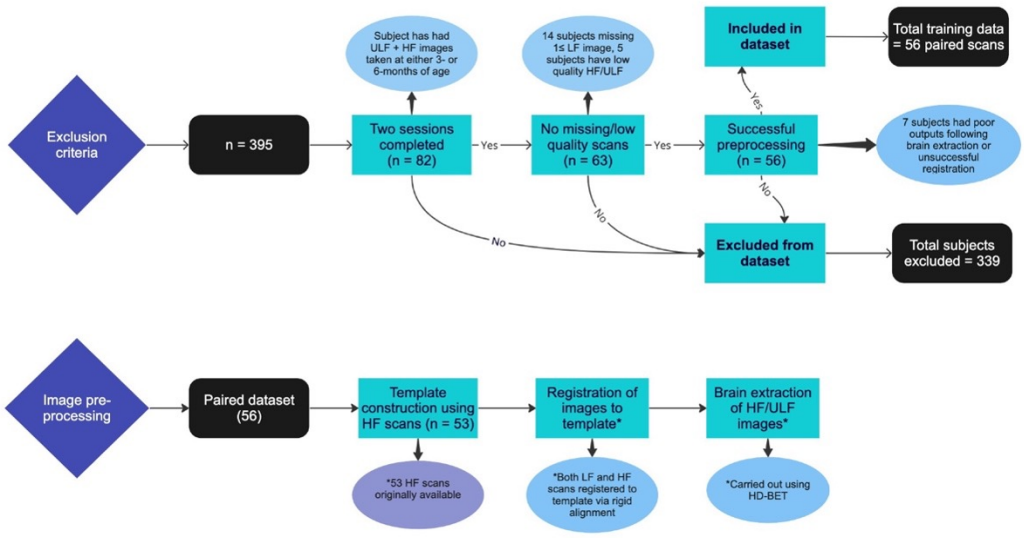

Figure 2) Top row: Participant exclusion; Bottom row: Pre-processing of images prior to being fed into the model.

2.2 CNN architecture

Our CNN (Figure 3) builds on the architecture of the 3D U-net (Çiçek et al., 2016). It has three input channels which lead into three encoding levels, each consisting of a convolution with a 3×3×3 kernel (allowing the model to capture fine-grained details while reducing parameter requirements), group normalisation (speeding up convergence by preventing internal covariate shift), a rectified-linear unit (ReLU) activation and maxpooling. The first layer has 64 features, with the number of features doubling after each maxpooling and halving after each upsampling. The final layer utilises a linear activation to produce the final image output. The model is implemented on a GPU using PyTorch.

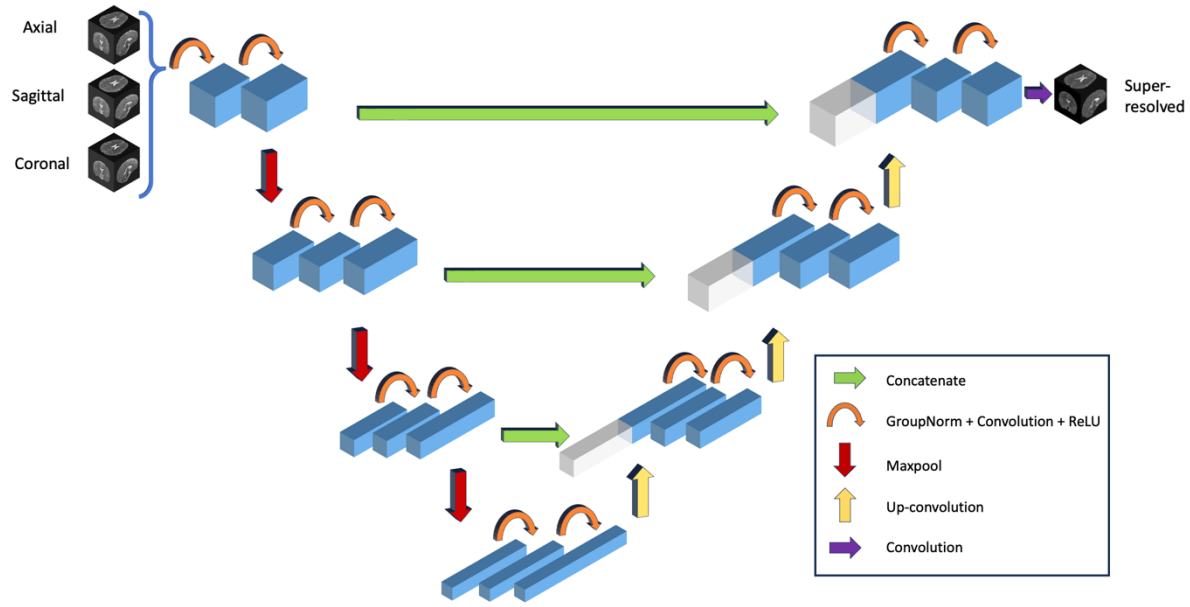

Figure 3) Model architecture: 3D-U-NET with three input channels and one output channel, allowing a transformation from three ULF scans to one SR image. Using this architecture, we aim to maximise the data available from a single subject.

2.3 Learning and inference

The training of our network minimises the peak-signal-to-noise ratio with respect to the neural network parameters θ:

$$\mathcal{L}_{PSNR}(\theta; \hat{Y}, Y) = 10 * \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{1}$$

where $\boldsymbol{MAX_I}$ denotes the maximum image intensity (a pre-defined input parameter) and $\boldsymbol{MSE}$ is the mean-squared error, or:

$$\mathcal{L}_{MSE}(\theta; \hat{Y}, Y) = \sum_{i=1}^{n} (y - \hat{y})^2 \tag{2}$$

where $\hat{\boldsymbol{y}}$ denotes the model output and $\boldsymbol{y}$ denotes the target image.

To assess the dependence of network outputs on the size of the training dataset (i.e. number of matched pairs of HF and ULF scans), we trained 4 models on 10, 20, 30, and 40 paired images, respectively, with the remaining 16 pairs being used to test model outputs. Each U-Net was trained using the Adam optimizer (Kingma, 2014) with a learning rate of 10^{-4} for 1000 epochs (~9 hours for the 10-scan model, ~30 hours for the 40-scan model using an Nvidia RTX 3090 GPU).

ULF images were normalised and scaled to 1-mm isotropic resolution, ensuring that input and output resolutions match. Augmentation of images was carried out on the fly, with random affine transform or random elastic deformation applied to both ULF and HF data with pre-set probability ($p = 0.5$). After training, weights of the model were frozen and inference was carried out (~10 seconds on an RTX 3090 GPU and ~1 minute on a modern CPU).

2.4 Evaluation of model outputs

The quality of model outputs from each of the 4 models relative to "ground-truth" HF images was evaluated via Dice overlap of segmented brain regions within participants and tissue volume correlations across participants, with both segmentation outputs and volume estimations obtained using SynthSeg (Billot, 2020). In addition to segmenting cortical grey matter (GM), cortical white matter (WM) and cerebrospinal fluid (CSF), we segmented the following deep brain structures: accumbens, amygdala, pallidum, hippocampus, caudate, putamen, thalamus, and lateral ventricles. All segmentations were visually QC'ed prior to running analyses. Considering that one of the primary applications of successful SR techniques would be use in ULF-based neurodevelopmental research in LMIC settings, selecting regions that play an important role in neurodevelopmental outcomes is particularly beneficial to underscore the value of our technique. In the case of conditions such as deep nuclear grey matter injury in neonatal hypoxic-ischemic encephalopathy (HIE) (Hassett et al., 2022) or perinatal stroke (Ilves et al., 2022), all regions described are affected in a manner such that damage is associated with worse neurodevelopmental outcomes. To compare how our model fares in capturing these regions compared to current state-of-the-art methods for super-resolution of paediatric ULF scans, we additionally computed the same metrics on images obtained using MRR.

Using our pre-trained models, we additionally assessed performance with varying numbers of input scans. More specifically, we assessed model outputs when the U-Nets only received an axial scan or axial and sagittal scans (as opposed to all three orientations, with which it was originally trained). Since our models require three input volumes, inference with missing scans was achieved by cloning the axial scan either once or twice, depending on whether one or two distinct inputs were provided. As such, we ran inference with each subject a total of three times, with the following combination of scans: 1) axial, axial, axial; 2) axial, axial, sagittal; 3) axial, sagittal, coronal. The choice of prioritising scans in the order of axial >

sagittal > coronal was done based on the scanning protocol used, which involved scanning participants in this exact order.

2.5 Statistical methods

Given the small test sample (n=16) we opted for nonparametric statistics to compare Dice overlap by using the Wilcoxon signed-rank tests to compare the medians (Wilcoxon, 1945). Tissue volume correlations across participants were assessed using Pearson's (Pearson, 1896) and Spearman's (Spearman, 1904) correlation coefficients

3 Results

Outputs from our U-Nets yielded visually superior images compared to original ULF scans. Model outputs appeared less noisy, but also smoother than raw ULF scans or MRR outputs (Figure 4). Importantly, a smoother image does not necessarily indicate that tissue volume is not appropriately captured, as we investigate below.

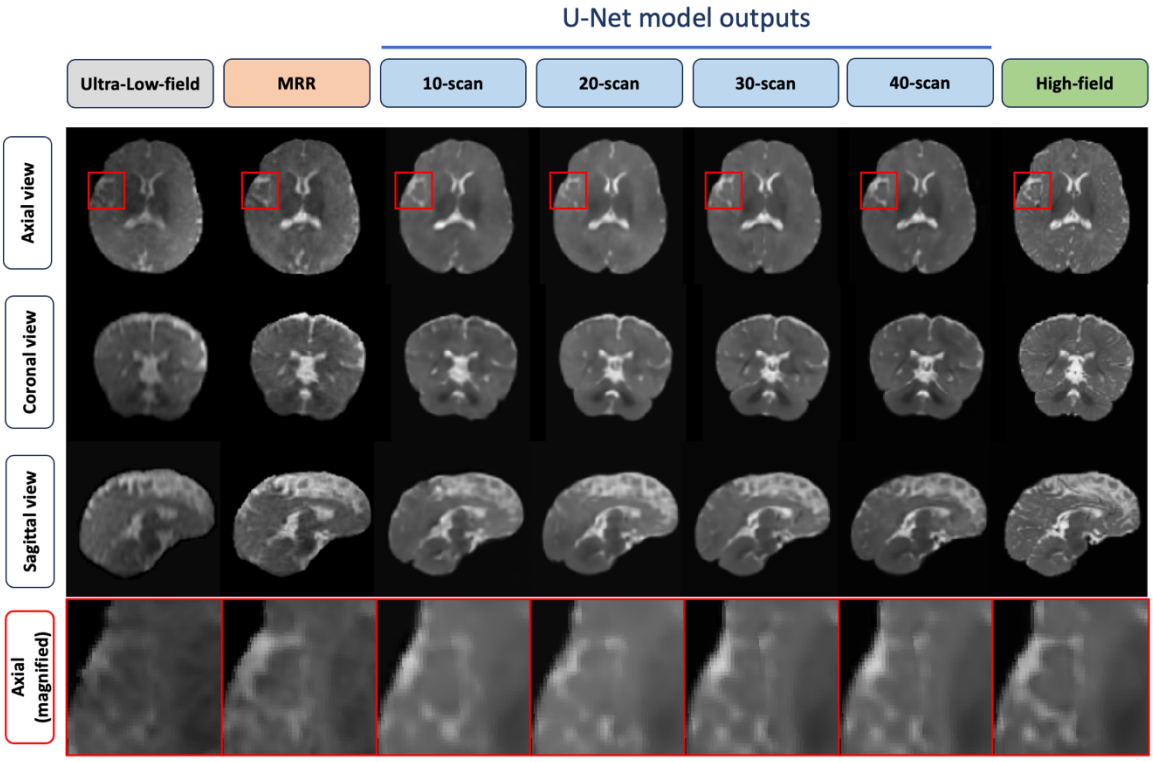

Figure 4) Model outputs from a single out-of-sample subject. Left to right: raw Axial ULF scan, output of multi-resolution registration, outputs of U-Nets as a function of increasing training set size, and the "ground-truth" high-field scan.

We observed a progressive increase in Dice overlap of segmented model predictions with segmented HF scans as the training set increased in size, across segmentation labels (Figure S1). Comparing Dice overlap of model predictions and HF scans to that of multi-resolution

registration outputs and HF scans, we began seeing an improvement relative to MRR starting with the 20-scan model, where median Dice scores increased for all deep brain structures apart from the hippocampus and lateral ventricles (Figure 5). A universal increase in the median Dice scores was observed with the 40-scan model, with the greatest difference seen in the caudate, putamen and pallidum (Δ Dice = 0.07) and the lowest in the lateral ventricles (Δ Dice = 0.02).

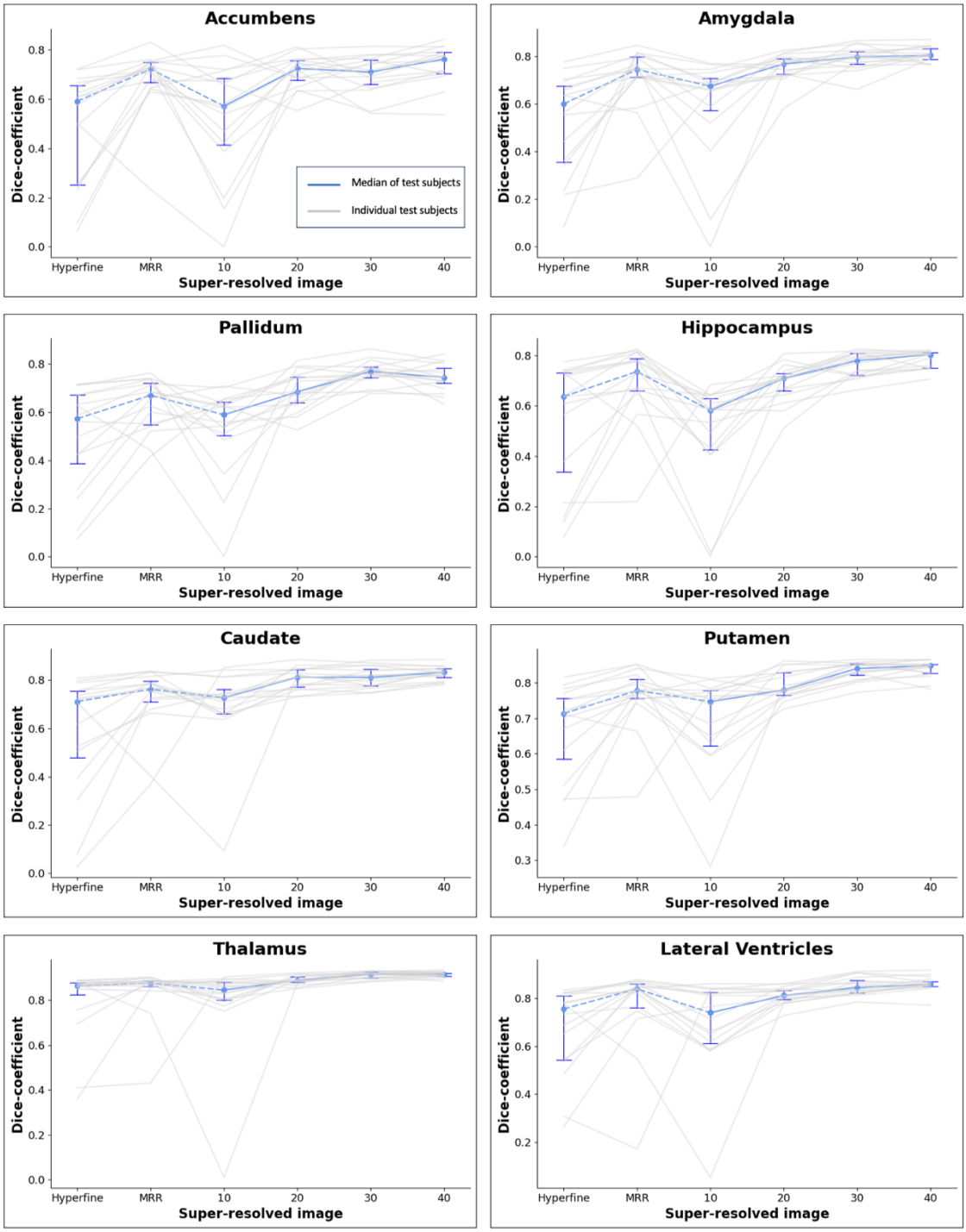

Figure 5) Region-specific Dice coefficients between HF scan and the original Hyperfine scan, MRR output, and outputs of U-Nets as a function of increasing training set. Dice score of original Hyperfine scan is obtained via averaging Dice of axial, coronal, and sagittal scans; IQR is displayed in dark blue

A similar pattern was observed when quantifying super-resolution quality by comparing regional volumes between super-resolved outputs and HF scans, across participants (Figure S2). The average Spearman's correlation coefficient (r_s) of deep brain structures across both hemispheres rose from 0.71 in MRR outputs to 0.77 in predictions of the 40-scan model, reflecting an 8.5% increase. Peak r_s, Md [Q_1, Q_3] = 0.96 [0.94, 0.97] was seen in the thalamus and lateral ventricles (Figures 6 & 7, Table S4).

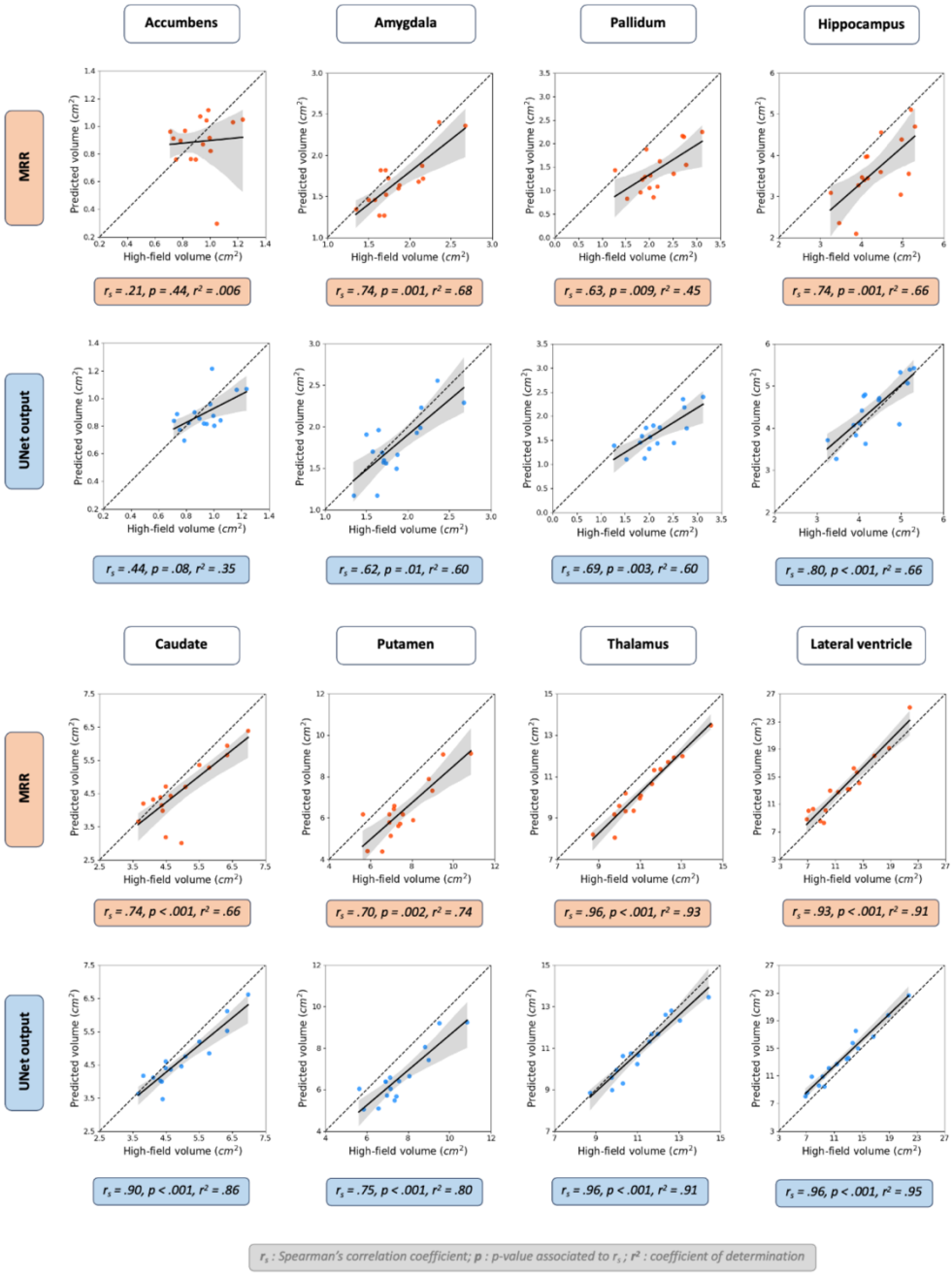

Figure 6) Region-specific volume correlations between MRR outputs and ground-truth HF scans (orange box) vs volume correlations between 40-scan U-Net output and ground-truth HF scan (blue box)

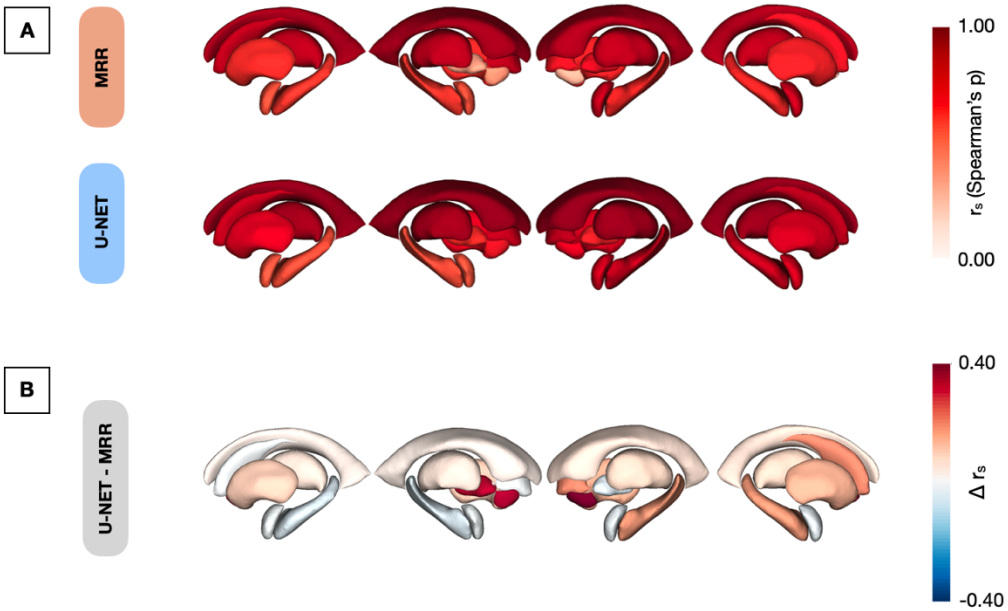

Figure 7) 3D-plot of deep brain structures (left & right accumbens, amygdala, caudate, hippocampus, pallidum, putamen, thalamus and lateral ventricle). A) Volume correlations with HF scans; Top row = MRR outputs, Bottom row = U-Ne outputs. B) Difference in volume correlation values between U-Net outputs and MRR outputs (U-Net - MRR). Images constructed using the ENIGMA toolbox (Larivière et al, 2022)

Significance testing using the Wilcoxon signed-rank test was applied to Dice scores obtained for all sub-cortical regions described above, and additionally to Dice scores for cerebrospinal fluid (CSF), cortical grey matter (GM) and cortical white matter (WM). Given a family-wise error rate (FWER) of 0.0045, significant differences in all subcortical regions were observed when comparing Dice scores from U-Net outputs to those of ULF scans. In contrast, six out of eight regions displayed a significant difference in Dice scores between MRR outputs and ULF scans. Furthermore, when comparing the two SR techniques (our proposed U-Net vs MRR), a greater Dice score was observed in all eight subcortical regions when using the U-Net, with 5 of these regions exhibiting a statistically significant increase. With regards to GM, WM and CSF, Dice scores obtained from MRR outputs did not exhibit any significant difference to those obtained from ULF scans. In contrast, all three of these tissues showed significant improvements in Dice scores following SR our U-Net (Table 1).

Table 1) Wilcoxon signed-rank test applied to Dice scores of 16 test subjects, comparing outputs of MRR to Hyperfine, UNet to Hyperfine, and UNet to MRR. For each comparison, the rank biserial correlation (RBC) and significance are displayed.

Region	**MRR > Hyperfine**		**UNet > Hyperfine**		**UNet > MRR**	
Subcortical	*RBC*	*Sig.*	*RBC*	*Sig.*	*RBC*	*Sig.*
Accumbens	0.83	0.00084*	0.91	0.00021*	0.34	0.13
Amygdala	0.91	0.00021*	0.99	0.000031*	0.82	0.0011*
Pallidum	0.82	0.0011*	1.0	0.000015*	0.87	0.00050*
Hippocampus	0.82	0.0011*	1.0	0.000015*	0.72	0.0046
Caudate	0.82	0.0011*	1.0	0.000015*	0.94	0.00011*
Putamen	1.0	0.00015*	1.0	0.000015*	1.0	0.000015*
Thalamus	0.60	0.017	1.0	0.000015*	0.99	0.000031*
Lateral ventricles	0.68	0.0078	1.0	0.000015*	0.86	0.00066*
Cortical	*RBC*	*Sig.*	*RBC*	*Sig.*	*RBC*	*Sig.*
Grey Matter	0.54	0.029	0.99	0.00031*	0.22	0.23
White Matter	0.54	0.029	1.0	0.000015*	0.46	0.06
CSF	0.54	0.029	0.91	0.000021*	0.044	0.45

When assessing how our models performed using a reduced number of inputs, we observed minimal difference in Dice scores given axial only or axial + sagittal inputs compared to all three orientations in all models trained on >10 paired sets ULF-HF scans. In instances when only one or two input scans were given, a training-set of 30 paired scans was sufficient for the U-Net to supersede MRR in Dice scores for all subcortical and cortical regions (Table 2; Figure S3).

Table 2) Region-specific Dice scores between segmentations obtained from MRR vs. model outputs as a function of increasing training set size. Size of training data is further stratified according to how many unique inputs each model received (A = Axial, AS = Axial + Sagittal, ACS = Axial + Sagittal + Coronal). Grey colouring is used to indicate points at which U-Net outputs exceed MRR outputs in terms of Dice scores. Abbreviations: Subcort. = subcortical, Accumb. = Accumbens, Hip. = hippocampus, Lat. ven. = lateral ventricles.

Region		**10-scan model**			**20-scan model**			**30-scan model**			**40-scan model**		
Subcort.	*MRR*	*A*	*AS*	*ACS*	*A*	*AS*	*ACS*	*A*	*AS*	*ACS*	*A*	*AS*	*ACS*
Accumb.	.69	.37	.34	.52	.67	.71	.71	.69	.69	.70	.71	.72	.74
Amygdala	.71	.34	.34	.59	.76	.76	.75	.78	.77	.79	.78	.80	.81
Pallidum	.63	.44	.45	.53	.66	.67	.69	.76	.77	.77	.72	.73	.74
Hip.	.69	.34	.35	.49	.69	.67	.69	.74	.75	.76	.77	.77	.78
Caudate	.72	.55	.50	.69	.79	.79	.80	.82	.81	.81	.82	.82	.83
Putamen	.76	.57	.56	.68	.79	.80	.79	.84	.84	.83	.82	.83	.84
Thalamus	.84	.60	.67	.79	.89	.89	.89	.91	.91	.91	.91	.91	.91
Lat. ven.	.77	.52	.49	.69	.81	.81	.81	.86	.85	.85	.86	.85	.86
Cortical	*MRR*	*A*	*AS*	*ACS*	*A*	*AS*	*ACS*	*A*	*AS*	*ACS*	*A*	*AS*	*ACS*
GM	.70	.61	.60	.66	.71	.71	.70	.74	.74	.74	.74	.75	.74
WM	.74	.63	.63	.69	.74	.74	.74	.77	.77	.77	.77	.77	.77
CSF	.47	.33	.33	.41	.48	.49	.48	.55	.56	.53	.54	.55	.53

4 Discussion

Obtaining accurate, high-resolution MRI scans in paediatric populations is essential for studies of neurodevelopment and for diagnosis of neurological conditions. In LMICs where MRI accessibility is significantly reduced, this may be achieved via the use of ULF scanners, supplemented with techniques to improve scan quality. Here we demonstrate improved results compared to previous efforts (Deoni et al., 2022) using a U-Net trained for SR of ULF paediatric images. More specifically, we present greater Dice overlap measures of segmented deep-brain structures and improved volume correlations in these regions. We additionally demonstrate the feasibility of training an SR model with a limited number of paired scans, as with 20-40 paired ULF-HF scans we show success in surpassing MRR on unseen subjects from the same sample population.

The design of our model assumes that feeding in three separate ULF scans maximises the amount of anatomical information available, allowing the U-Net to more accurately reconstruct a given subject's brain. When assessing Dice scores of deep brain structures from model outputs with varying numbers of unique input scans, however, we see marginal improvement in reconstructions from one scan vs. three unique scans. Tolerance to long scanning times is low in infants (Barkovich et al., 2019), thus successfully completing three separate 4-minute-scanning sessions is often infeasible. As such, having a model that can reconstruct an image from just one ULF scan has major practical benefits. An alternative approach would involve modifying our architecture to have one full encoding branch for each input image, where these separate encodings are then combined before they enter the expanding path (Lau et al 2022). This would encourage distinct, image-specific encodings, which may further maximise the anatomical information extracted from each scan, but would impede the applicability of the model to scenarios with a reduced number of scans.

There are several limitations of this study worth noting. Firstly, we only assess model performance on unseen subjects of the same age from the same scanning site. To fully assess the limits of our model, we would run inference on ULF scans from subjects of varying ages and alternative sites, investigating the magnitude of deviation from the training set that results in a sharp decline in model performance. Additionally, including test subjects with pre-specified neurobiological conditions/lesions would allow us to investigate whether using deep learning-based SR runs the risk of failing to capture pathologies. This would provide insight

as to whether a model trained on normative images can be applied to subjects with potential pathologies, or if disease-specific models need to be trained to ensure reliable SR outputs.

An additional limitation stems from the use of SynthSeg for image segmentation (Billot et al., 2023). As SynthSeg was trained on adult data, its application to paediatric images runs the risk of producing erroneous volume estimations. Owing to differing water and fat content in the developing brain compared to adults, alterations in signal intensities in newborns and infants are seen with both T_1- and T_2-weighted anatomical MRI. In particular, the first 6 months of life are characterised by a reversal of the normal adult contrast (T_1-weighted: lower WM intensity than GM; T_2-weighted: higher WM intensity than GM) (Dubois et al., 2021). These neurodevelopmental changes amplify the drastic difference already present in T_1- and T_2-weighted contrast between HF and ULF MRI, and as such a more accurate approach to testing model performance would involve the use of a segmentation model trained specifically on paediatric scans. Given such a reliable segmentation scheme, we could further expand upon our performance metrics with additional boundary-based measures (such as the Hausdorff distance or average symmetric surface distance (Reineke et al., 2024)).

Finally, although the U-Net has proven to be the gold-standard for many deep-learning applications to medical imaging, the use of novel architectures such as diffusion models or visual transformers may yield further improvements in quality of model outputs. These methods have the potential to address critical limitations of previous techniques, enhancing overall realism in super-resolved images. Within the domain of SISR of medical modalities, both diffusion models and visual transformers have been shown to outperform CNN and GAN-based models in terms of Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR) (Kazerouni et al., 2023, Zhu et al., 2023). As such, future work should explore the use of 3D counterparts of such models.

5 Conclusion

Ultra-low-field (ULF) imaging presents a potential paradigm-shift in neuroimaging, however its ingress into widespread research and clinical use is impeded by limitations on resolution and the range of available contrasts. Here we demonstrate how the use of deep learning can aid in deriving high-resolution T_2-weighted scans of healthy paediatric subjects from ULF scanners. More specifically, we demonstrate the feasibility of training U-Nets with a limited

number of paired ULF-HF scans, such that their outputs outperform current non-deep-learning based state-of-the-art methods for super-resolving paediatric scans.

Funding information

Artificial Intelligence Methods for Low Field MRI Enhancement, Bill and Melinda Gates Foundation (INV- 032788). Wellcome Leap 1kD programme (The First 1000 Days) [222076/Z/20/Z].

Conflict of interest

The authors report no significant financial conflicts of interest with respect to the subject matter of this manuscript.

Code availability statement

All model training code is available at https://github.com/levente-1/Multi-orientation-3D-U-Net.

References

Arnold, T.C., Freeman, C.W., Litt, B. and Stein, J.M. (2023), Low-field MRI: Clinical promise and challenges. *J Magn Reson Imaging, 57*: 25-44. https://doi.org/10.1002/jmri.28408

Ashburner J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage* 38, 95–113 10.1016/j.neuroimage.2007.07.007

Barkovich MJ, Li Y, Desikan RS, Barkovich AJ, Xu D. Challenges in pediatric neuroimaging. *Neuroimage*. 2019 Jan 15;185:793-801. doi: 10.1016/j.neuroimage.2018.04.044.

Billot, B., et al. (2020). A Learning Strategy for Contrast-agnostic MRI Segmentation. Proceedings of the Third Conference on Medical Imaging with Deep Learning. *Proceedings of Machine Learning Research, PMLR*. 121: 75--93

Campbell-Washburn, A. E., et al. (2019). "Opportunities in Interventional and Diagnostic Imaging by Using High-Performance Low-Field-Strength MRI." *Radiology* 293(2): 384-393.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O. (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. *Springer, Cham*. https://doi.org/10.1007/978-3-319-46723-8_49

Deoni SCL, O'Muircheartaigh J, Ljungberg E, Huentelman M, Williams SCR. Simultaneous high-resolution T2-weighted imaging and quantitative T2 mapping at low magnetic field strengths using a multiple TE and multi-orientation acquisition approach. *Magn Reson Med.* 2022; 88(3): 1273-1281. doi:10.1002/mrm.29273

Dong, C., Loy, C.C., He, K., Tang, X. (2014). Learning a Deep Convolutional Network for Image Super-Resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8692. *Springer, Cham*. https://doi.org/10.1007/978-3-319-10593-2_13

Dubois, J., Alison, M., Counsell, S.J., Hertz-Pannier, L., Hüppi, P.S. and Benders, M.J.N.L. (2021), MRI of the Neonatal Brain: A Review of Methodological Challenges and Neuroscientific Advances. *J Magn Reson Imaging*, 53: 1318-1343. https://doi.org/10.1002/jmri.27192

Hassett, J., Carlson, H., Babwani, A., & Kirton, A. (2022). Bihemispheric developmental alterations in basal ganglia volumes following unilateral perinatal stroke. *NeuroImage: Clinical, 35*, 103143. https://doi.org/10.1016/j.nicl.2022.103143

Iglesias, J. E., Billot, B., Balbastre, Y., Tabari, A., Conklin, J., Gilberto González, R., . . . Fischl, B. (2021). Joint super-resolution and synthesis of 1 mm isotropic MP-RAGE volumes from clinical MRI exams with scans of different orientation, resolution and contrast. *NeuroImage*, 237, 118206. https://doi.org/10.1016/j.neuroimage.2021.118206

Iglesias, J. E., et al. (2023). "Quantitative Brain Morphometry of Portable Low-Field-Strength MRI Using Super-Resolution Machine Learning." *Radiology* 306(3): e220522.

Ilves, N., Männamaa, M., Laugesaar, R., Ilves, N., Loorits, D., Vaher, U., . . . Ilves, P. (2022). Language lateralization and outcome in perinatal stroke patients with different vascular types. *Brain Lang, 228,* 105108. doi:10.1016/j.bandl.2022.105108

Isensee F, Schell M, Tursunova I, Brugnara G, Bonekamp D, Neuberger U, Wick A, Schlemmer HP, Heiland S, Wick W, Bendszus M, Maier-Hein KH, Kickingereder P. Automated brain extraction of multi-sequence MRI using artificial neural networks. *Hum Brain Mapp*. 2019; 1–13. https://doi.org/10.1002/hbm.24750

Islam, K. T., et al. (2023). "Improving portable low-field MRI image quality through image-to-image translation using paired low- and high-field images." *Sci Rep* 13(1): 21183.

Jalloul, M., et al. (2023). "MRI scarcity in low- and middle-income countries." *NMR Biomed* 36(12): e5022.

Jensen, S. K. G., Xie, W., Kumar, S., Haque, R., Petri, W. A., & Nelson, C. A., 3rd. (2021). Associations of socioeconomic and other environmental factors with early brain development in Bangladeshi infants and children. *Dev Cogn Neurosci, 50,* 100981. doi:10.1016/j.dcn.2021.100981

Kazerouni, A., et al. (2023). "Diffusion models in medical imaging: A comprehensive survey." *Medical Image Analysis* 88: 102846.

Kim, J., Lee J. K., Lee, K. M. "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Las Vegas, NV, USA, 2016, pp. 1646-1654, doi: 10.1109/CVPR.2016.182.

Kingma, D.P., Ba, J., Adam: A method for stochastic optimization. *arXiv*:1412.6980 [cs.LG] (22 December 2014).

Klein, H. M. (2020). "Low-Field Magnetic Resonance Imaging." *Rofo* 192(6): 537-548.

Larivière, S., Paquola, C., Park, By. et al. The ENIGMA Toolbox: multiscale neural contextualization of multisite neuroimaging datasets. *Nat Methods* 18, 698–700 (2021). https://doi.org/10.1038/s41592-021-01186-4

Lau V, Xiao L, Zhao Y, et al. Pushing the limits of low-cost ultra-low-field MRI by dual-acquisition deep learning 3D superresolution. *Magn Reson Med*. 2023; 90: 400-416. doi: 10.1002/mrm.29642

Nolen-Hoeksema, S. (2014). *Abnormal psychology*. New York, NY, McGraw-Hill Education.

Ogbole GI, Adeyomoye AO, Badu-Peprah A, Mensah Y, Nzeh DA. Survey of magnetic resonance imaging availability in West Africa. *Pan Afr Med J*. 2018 Jul 31;30:240. doi: 10.11604/pamj.2018.30.240.14000

Oktay, O. et al. (2016). Multi-input Cardiac Image Super-Resolution Using Convolutional Neural Networks. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016, *Springer, Cham*. https://doi.org/10.1007/978-3-319-46726-9_29

Perdue, K. L., Jensen, S. K. G., Kumar, S., Richards, J. E., Kakon, S. H., Haque, R., . . . Nelson, C. A. (2019). Using functional near-infrared spectroscopy to assess social information processing in poor urban Bangladeshi infants and toddlers. *Dev Sci*, 22(5), e12839. doi:10.1111/desc.12839

Pham, C.-H., Tor-Díez, C., Meunier, H., Bednarek, N., Fablet, R., Passat, N., & Rousseau, F. (2019). Multiscale brain MRI super-resolution using deep 3D convolutional networks. *Computerized Medical Imaging and Graphics, 77,* 101647. https://doi.org/10.1016/j.compmedimag.2019.101647

Reinke, A., et al. (2024). "Understanding metric-related pitfalls in image analysis validation." Nature Methods 21(2): 182-194

Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, *Springer, Cham*. https://doi.org/10.1007/978-3-319-24574-4_28

Sarracanie, M., LaPierre, C. D., Salameh, N., Waddington, D. E. J., Witzel, T., & Rosen, M. S. (2015). Low-Cost High-Performance MRI. *Scientific Reports*, 5(1), 15177. doi:10.1038/srep15177

Shi, W. et al., "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Las Vegas, NV, USA, 2016 pp. 1874-1883. doi: 10.1109/CVPR.2016.207

Wilcoxon, Frank (Dec 1945). "Individual comparisons by ranking methods". *Biometrics Bulletin. 1* (6): 80–83. doi:10.2307/3001968

Wu, Z., et al. (2023). "Super-resolution of brain MRI images based on denoising diffusion probabilistic model." Biomedical Signal Processing and Control 85: 104901

Zieff MR, Miles M, Mbale E et al. Characterizing developing executive functions in the first 1000 days in South Africa and Malawi: The Khula Study [version 1; peer review: awaiting peer review]. Wellcome Open Res 2024, 9:157 (https://doi.org/10.12688/wellcomeopenres.19638.1)

Zhang, K., et al. (2022). "SOUP-GAN: Super-Resolution MRI Using Generative Adversarial Networks." Tomography 8(2): 905-919

Zhou, Z., et al. (2022). "Super-resolution of brain tumor MRI images based on deep learning." J Appl Clin Med Phys 23(11): e13758

Zhu, J., et al. (2023) A residual dense vision transformer for medical image super-resolution with segmentation-based perceptual loss fine-tuning. *arXiv*:2302.11184 DOI: 10.48550/arXiv.2302.11184

Supplementary Information

S1 & S2 – Global dice scores and volume correlations

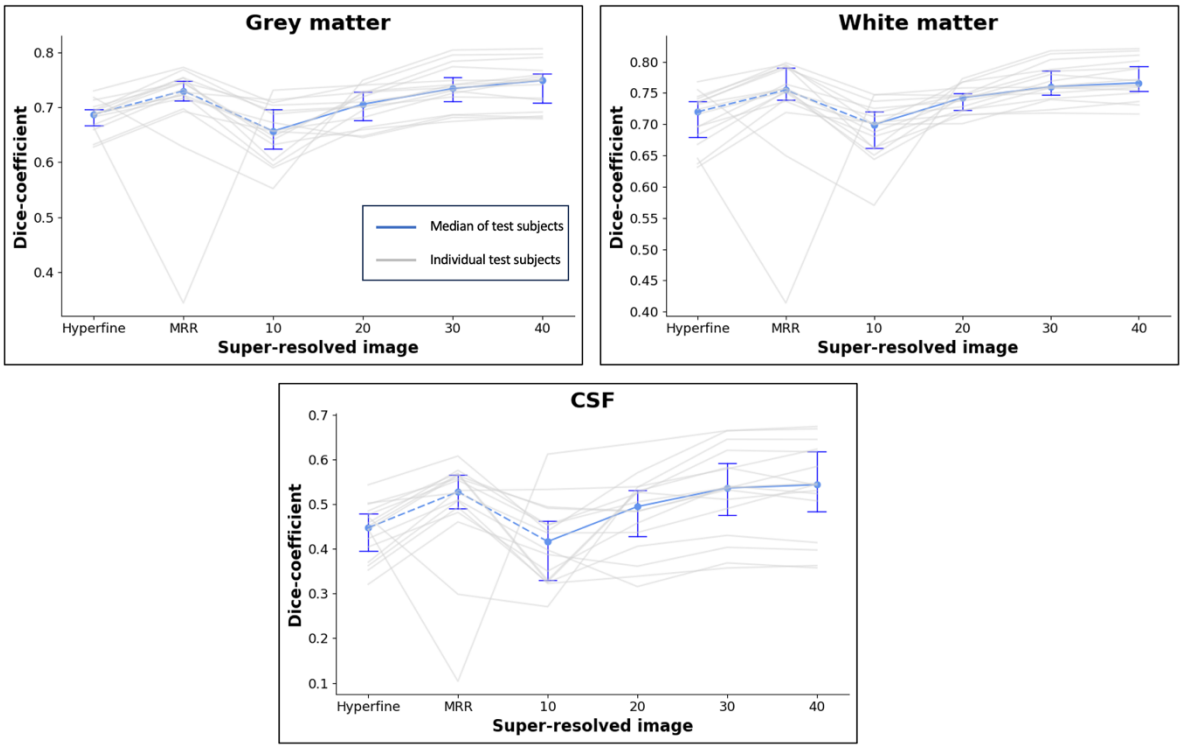

Figure S1) Region-specific Dice coefficients between HF scan and the original Hyperfine scan, MRR output, and outputs of U-Nets as a function of increasing training set. Regions depicted: Grey Matter, White Matter and Cerebrospinal fluid (CSF)

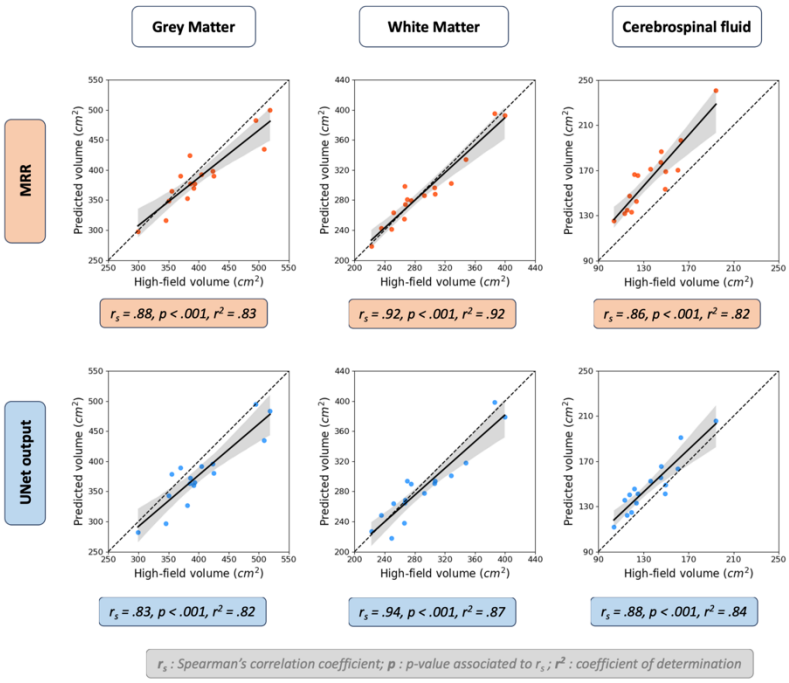

Figure S2) Region-specific volume correlations between MRR outputs and ground-truth HF scans (orange box) vs volume correlations between 40-scan U-Net output and ground-truth HF scan (blue box). Regions depicted: Grey Matter, White Matter, Cerebrospinal fluid

S3 – Dice scores for varying number of input scans

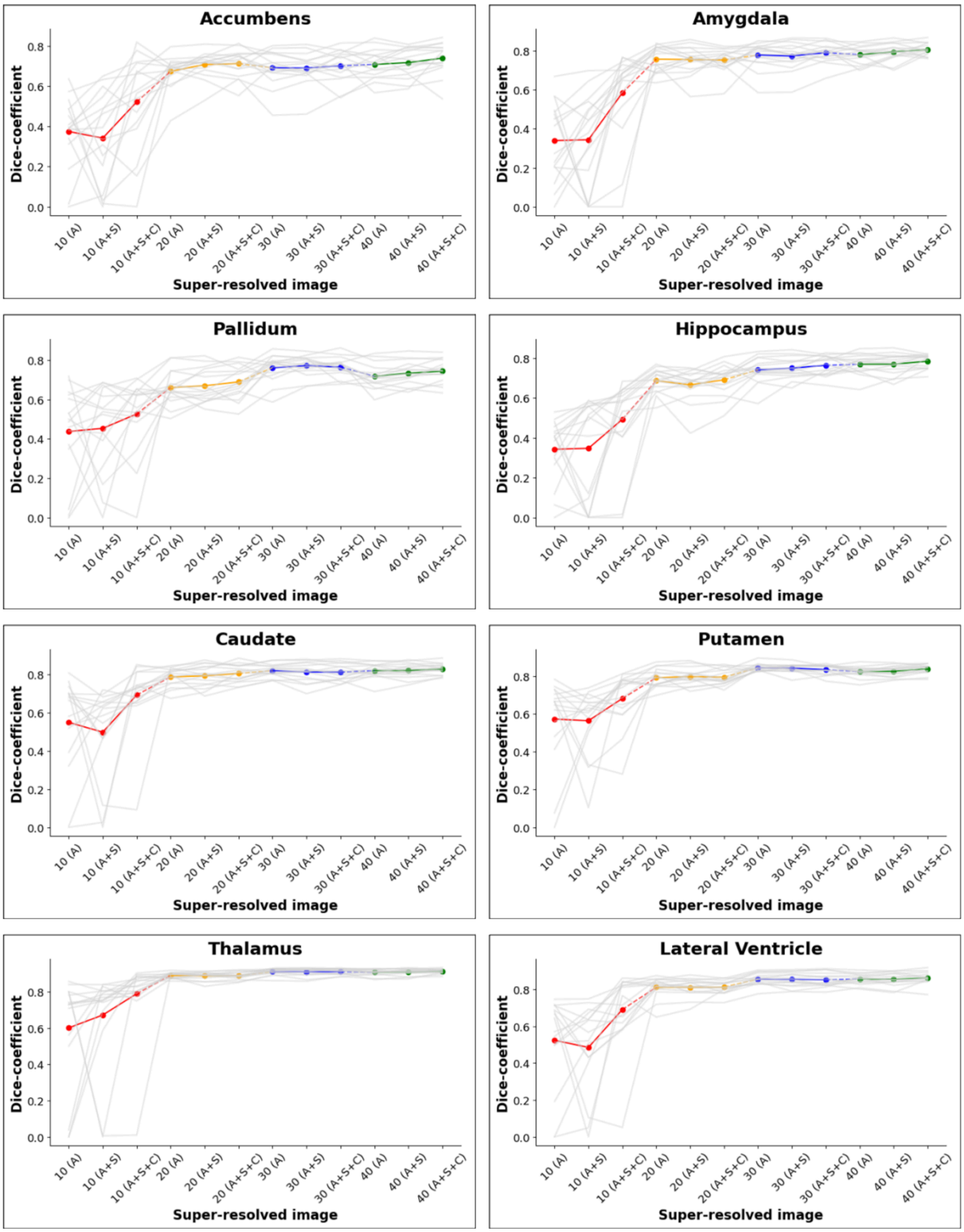

Figure S3) Region-specific Dice coefficients between HF scan and outputs of U-Nets as a function of increasing training set and increasing number of inputs given to the models. A = axial, A+S = axial and sagittal, A+S+C = axial, sagittal and coronal.

S4 – Interquartile range for Spearman's rank correlation

The IQR for each Spearman's rank correlation value was calculated via the following equation:

$$CI = \tanh\left(\text{atanh} * r \pm \frac{0.6745}{\sqrt{n-3}}\right) \quad (3)$$

where r is the estimate of correlation and n is the sample size. The Fisher transformation, obtained via ***atanh***, ensures that the sampling distribution of the estimate is approximately normal. As such, a 75% confidence interval is generated by taking this transformed estimate and adding/subtracting 0.6745 times the standard error

Table S4) Spearman's rank correlation and 95% confidence interval across regions

Region	**MRR predictions**			**UNet predictions**		
Subcortical	r_s	*Lower*	*Upper*	r_s	*Lower*	*Upper*
Accumbens	0.21	0.026	0.38	0.44	0.28	0.58
Amygdala	0.74	0.64	0.81	0.62	0.49	0.72
Pallidum	0.63	0.50	0.73	0.69	0.58	0.78
Hippocampus	0.74	0.64	0.81	0.80	0.72	0.85
Caudate	0.74	0.64	0.81	0.90	0.86	0.93
Putamen	0.70	0.59	0.78	0.75	0.66	0.82
Thalamus	0.96	0.94	0.97	0.96	0.94	0.97
Lateral ventricles	0.93	0.90	0.95	0.96	0.94	0.97
Cortical	r_s	*Lower*	*Upper*	r_s	*Lower*	*Upper*
Grey Matter	0.88	0.83	0.92	0.83	0.76	0.88
White Matter	0.92	0.89	0.94	0.94	0.91	0.96
CSF	0.86	0.80	0.90	0.88	0.83	0.92

Khula SA Study Team

Michal R. Zieff
Donna Herr
Chloë A. Jacobs
Sadeeka Williams
Zamazimba Madi
Nwabisa Mlandu
Tembeka Mhlakwaphalwa
Lauren Davel
Reese Samuels
Zayaan Goolam
Thandeka Mazubane
Bokang Methola
Khanyisa Nkubungu
Candice Knipe