

Combined Denoising and Suppression of Transient Artefacts in Arterial Spin Labelling MRI Using Deep Learning

Patrick W. Hales (PhD)¹, Josef Pfeuffer (PhD)², Chris Clark (PhD)¹

¹ Developmental Imaging & Biophysics Section, UCL Great Ormond Street Institute of Child Health, London WC1N 1EH

² Siemens Healthcare GmbH, MR Application Development, Erlangen, Germany

Corresponding author: Patrick W. Hales, Developmental Imaging & Biophysics Section, UCL Great Ormond Street Institute of Child Health, 30 Guilford Street, London, WC1N 1EH. Email: p.hales@ucl.ac.uk. Tel: 020 7905 2125.

Acknowledgements: This work was supported by the National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children NHS Foundation Trust and University College London.

Funding: This work was funded by Children with Cancer UK (grant number CwCUK-15-203).

Abstract

BACKGROUND

Arterial spin labelling (ASL) is a powerful tool for measuring cerebral blood flow (CBF). However, due to the low signal-to-noise ratio (SNR) of the technique, multiple repetitions are required, which results in prolonged scan times and increased susceptibility to artefacts.

PURPOSE

To develop a deep learning-based algorithm for simultaneous denoising and suppression of transient artefacts in ASL images.

STUDY TYPE

Retrospective.

SUBJECTS

129 paediatric neuro-oncology patients for model training. 11 healthy adult subjects for model evaluation.

FIELD STRENGTH/SEQUENCE: 3T/ASL

ASSESSMENT

Data were acquired at 3.0 T, using pseudo-continuous and pulsed ASL sequences. A denoising autoencoder (DAE) model was designed with stacked encoding / decoding convolutional layers. Gold standard images were generated by averaging 10 pairwise ASL subtraction images. The model was trained to produce perfusion images of a similar quality using a single subtraction image. Performance was compared against Gaussian and non-local means (NLM) filters. Evaluation metrics included SNR, peak SNR and structural similarity index (SSIM) of the CBF images, compared to the gold standard.

STATISTICAL TESTS

One-way ANOVA tests for group comparisons.

RESULTS

The DAE and NLM models both provided increases in SNR, by a factor of 3.2 ± 1.5 and 6.2 ± 0.7 respectively ($p < 0.01$). The DAE model was also effective at suppressing transient artefacts, and was the only model to show a significant improvement in accuracy in the generated CBF images, as assessed using PSNR. CBF images generated by the DAE model also had the highest SSIM values (0.85 ± 0.03) in comparison to the gold standard. In addition, the DAE images produced the lowest fitting error to the Buxton kinetic model, using data from multiple inflow time acquisitions.

DATA CONCLUSION

Deep learning-based algorithms provide superior accuracy when denoising ASL images, due to their ability to simultaneously increase SNR and suppress artefactual signals in raw ASL images.

Keywords

Deep learning, arterial spin labelling, ASL, denoising, autoencoder, CNN

Introduction

Arterial spin labelling (ASL) is a non-invasive imaging modality which provides a powerful means of measuring cerebral blood flow (CBF). One of the key advantages of ASL is that it utilises water in the blood as an endogenous tracer to measure perfusion, eliminating the need for an exogenous contrast agent. Instead, a series of label (L) and control images (C) are acquired, in which inflowing blood-water proximal to the imaging volume is 'tagged' using radio-frequency (RF) pulses during the label acquisition. After a delay, to allow labelled blood to flow into the tissue (the post-labelling delay, PLD), the perfusion signal (dM) is determined by the pair-wise subtraction of control and label images ($dM = C - L$). These are subsequently converted into CBF images, in physiological units of ml / 100g / min, using the method described in (1).

Due to the quantitative nature of the technique, the lack of exposure to ionising radiation, and the avoidance of a contrast agent injection, ASL has excellent clinical potential. However, an inherent limitation is the comparatively low signal-to-noise ratio (SNR) of the technique. This is due to a number of factors. Firstly, T_1 recovery of the tagged bolus during the PLD reduces the signal available from the tracer itself. In addition, in normal grey matter, perfusion replaces only ~1% of the brain water with in-flowing blood water every second (1). As such, inflowing blood can only perturb a very small fraction of the total magnetization in a typical voxel, and un-wanted signal fluctuations in the static tissue can easily outweigh the perfusion signal. To counteract this, background suppression pulses are often used to null the signal from the static tissue prior to image acquisition (2). Nonetheless, generally multiple repetitions of an ASL acquisition must be acquired in order to provide sufficient SNR, which leads to increased scan times.

In addition to the inherent limitations in SNR, ASL images can be corrupted by a number of artefacts (4, 5). Some of these are related to the acquisition protocol and / or the physiology of the subject, such as arterial transit time artefacts resulting from an insufficiently long PLD (5). Others are transient, and may occur sporadically during the series of repetitions. These can include artefacts related to subject motion (5), and CSF 'shine-through' in the ventricles

due to RF instabilities (4, 6). As these artefacts typically occur in only a small number of the total repetitions, their impact is less conspicuous after signal averaging (Figure 1).

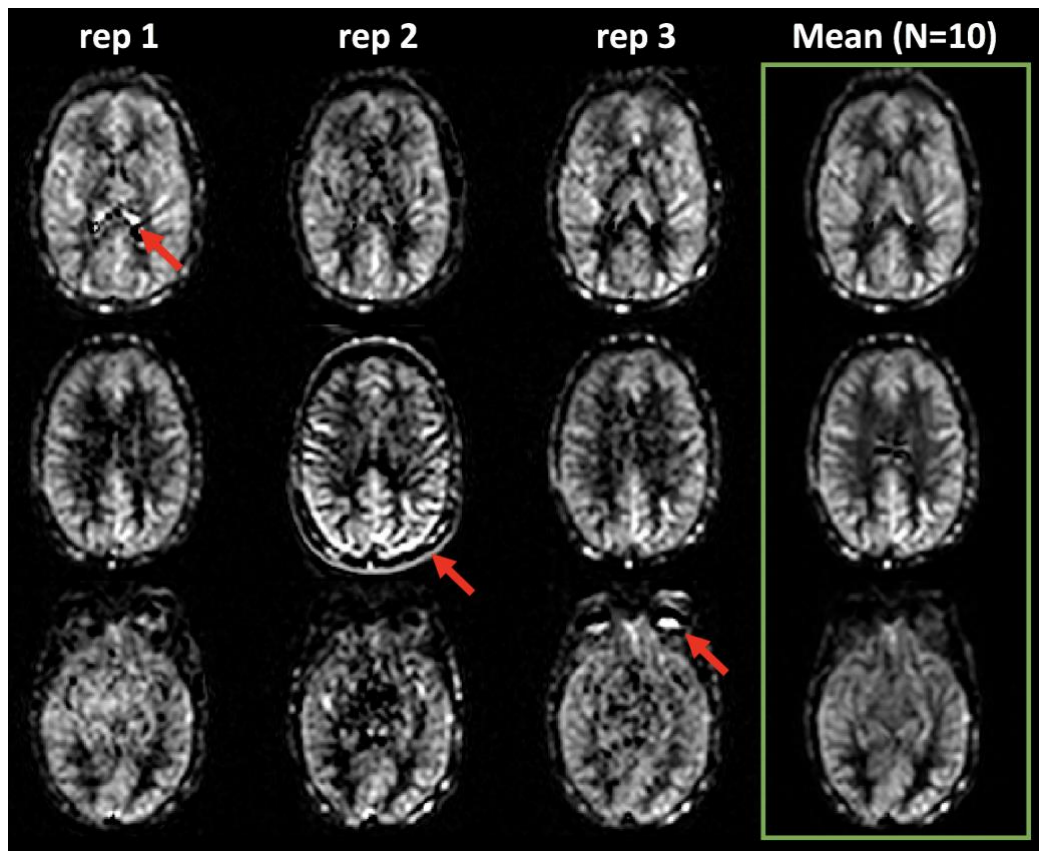


Figure 1 Illustration of transient artefacts affecting image quality in ASL datasets. Individual dM images for 3 repetitions are shown, along with the corresponding image after averaging over 10 repetitions (green box). Individual artefacts are illustrated with red arrows. Top row: CSF shine-through, demonstrating artefactual high signal in the lateral ventricles. Middle row: subject motion artefact, resulting in artefactual signal modulation within the brain, and a peripheral ring of high signal intensity. Bottom row: increased dM signal due to the subject's eye-motion. Given the transient nature of these artefacts, their impact is less pronounced after averaging over multiple repetitions (right column).

A number of post-processing techniques have been investigated to improve image quality in ASL data. Techniques to suppress transient artefacts include outlier rejection to remove hardware instabilities and motion-corrupted repetitions (6–10), physiological noise correction (11), and temporal filtering techniques (12–14). Techniques to increase SNR have focused on established image denoising techniques. These range from simple Gaussian smoothing (12, 15, 16), to more complex denoising strategies, such as wavelet-based

techniques (12, 17), Wiener filters (12), adaptive filters (12), non-local means filters (18, 19), and total generalized variation regularization (14).

In recent years, deep learning has emerged as a powerful tool for image processing and reconstruction. Within this field, convolutional neural networks (CNNs) have become a popular choice for processing imaging data, due to their ability to learn important features of images in a translationally invariant way. These techniques have been successfully applied to image denoising (20–22), and several studies have applied deep learning approaches to improving SNR in ASL images (23–28). Kim et al. developed a denoising CNN with two pathways, for extracting local low-level features and large-scale global features in parallel (23). This was shown to provide improvements in SNR and CBF accuracy in both single-PLD and Hadamard-encoded multiple-PLD data. Ulas et al. developed a CNN which was trained using a custom loss function, which enforced CBF estimates to be close to model-based reference values (24). Xie et al. recently developed a model combining dilated convolutions with wide activation residual blocks, which provided improved denoising compared to existing CNN architectures (26). Owen et al. introduced a joint filtering CNN model, in which maps of the mean and temporal variance of the ASL signal were used as dual inputs, in order to improve SNR and partially suppress transient artefacts (28). Finally, Gong et al. have demonstrated an unsupervised-deep-learning based framework that incorporates a subject's T1-weighted anatomical image as a structural prior.

The aforementioned studies have shown promising results for denoising low-SNR ASL images. However, previous studies have generally applied averaging over a subset of the total acquired repetitions, in order to generate low-SNR inputs for model training. Some previous studies have also applied motion correction (24, 26, 28) and Gaussian smoothing (24, 26) to the input data. In doing so, the presence of transient artefacts in the input data will be reduced, and the ability of these models to identify and suppress these artefacts may be compromised.

In this study, we explored the possibility of using all 'raw' repetitions (without averaging, motion correction or smoothing) from pseudo-continuous labelling ASL (pCASL) datasets as inputs to a deep learning-based denoising model. The model was based on an autoencoder

architecture, with stacked encoding / decoding convolutional layers, as this has previously shown promising results for image denoising (29–33). The denoising autoencoder (DAE) was trained to learn the transformation between the raw dM images from every repetition (dM_{raw}), and their corresponding ‘gold-standard’, which represent the same image after averaging over all acquired repetitions (dM_{mean}). Our hypothesis was that transient artefacts would occur in only in a subset of the dM_{raw} images, and as such the model would learn to identify and suppress these, while simultaneously increasing SNR.

The proposed model was trained using a large clinical database of pCASL images, acquired in paediatric brain tumour patients. In doing so, a further aim was to develop a model which could differentiate abnormal ASL signals associated with pathology from abnormal signals associated with transient artefacts. The signal associated with pathology would be expected to vary between subjects but remain consistent across repetitions; whereas the signal associated with transient artefacts should only occur in a small subset of repetitions.

Following training, we aimed to evaluate the model’s performance in healthy adult volunteers, to determine its performance under normal conditions (i.e. adult subjects with no pathology). The trained model was evaluated using both pCASL data and multiple inflow-time (multi-TI) pulsed-ASL (PASL) data, to evaluate the model’s performance when faced with data acquired with an alternative acquisition modality. In the latter, the SNR of individual images is often reduced in exchange for greater inflow-time coverage, making this an ideal application in which to test the proposed model. The performance of the model was also demonstrated in illustrative clinical cases.

Materials and Methods

Arterial Spin Labelling Acquisition

All ASL datasets were acquired using a 3T Siemens MRI scanner (MAGNETOM Prisma, Siemens, Erlangen, Germany), equipped with a 20-channel head receive coil. PCASL data were acquired using a Siemens prototype sequence, with background suppression RF pulses and a 3D gradient-and-spin-echo readout. The labelling duration was 1800ms, with a 1500ms post-labelling delay, and 10 repetitions were acquired. Additional sequence parameters were:

TR=4620ms, TE=21.8ms, field of view=220mm, matrix size=64 x 62, in-plane resolution=1.7 x 1.7mm (after zero-filling), number of partitions=24, slice thickness=4.0 mm, turbo factor=12, EPI factor=31, segments=2 (with parallel imaging, GRAPPA=2). A proton-density weighted (M_0) image was also acquired (TR=4000ms), with identical readout to the ASL acquisition but with the labelling and background suppression RF pulses removed, for CBF quantification. Total acquisition time was 3 min 19 s.

Multiple inflow time (multi-TI) pulsed ASL (PASL) data were acquired using the same prototype sequence. Acquisitions were acquired at 10 TIs, ranging from 350 to 2600ms in 250 ms steps, with a single acquisition per TI. The TR was 3300ms; all other readout parameters were identical to the pCASL acquisition. Q2TIPS RF pulses were applied 700ms after the labelling pulse to define the temporal width of the bolus. The total acquisition time (for all TIs and the M_0 acquisition combined) was 2 min 25 s.

Training, Validation and Testing Data

Retrospective anonymised pCASL data were accrued from the clinical database at our institution, following approval from our Institutional Ethics Board. This was comprised of a cohort of 129 paediatric neuro-oncology patients (mean age = 7.1, range = 0.4 – 17.1 years), all of whom received the pCASL acquisition described above as part of their clinical imaging. Each pCASL dataset consisted of 24 axial slices, each with 10 repetitions, resulting in a set of 240 'raw' difference images (dM_{raw}) per subject. Each dM_{raw} image was matched to the corresponding mean image (dM_{mean}), averaged over 10 repetitions, which represented the gold-standard in this study. Over the entire cohort, this provided a set of 30,960 noisy dM_{raw} images (single repetition), each matched to their corresponding gold-standard images (mean of 10 repetitions). 80% of this dataset was used for model training, with 20% retained for model validation. The mean and standard deviation (SD) of the raw image set were used for Z-normalization of all images before they were entered into the model.

For model testing, new pCASL data were acquired in 11 healthy adult subjects (mean age 32 years, range 23-40 years). Additional multi-TI PASL data (using the protocol described in Section 2.1) were acquired in 7 healthy subjects (mean age 30 years, range 21-40 years).

Model Architecture

A schematic of the denoising autoencoder (DAE) model architecture is shown in Figure 2.

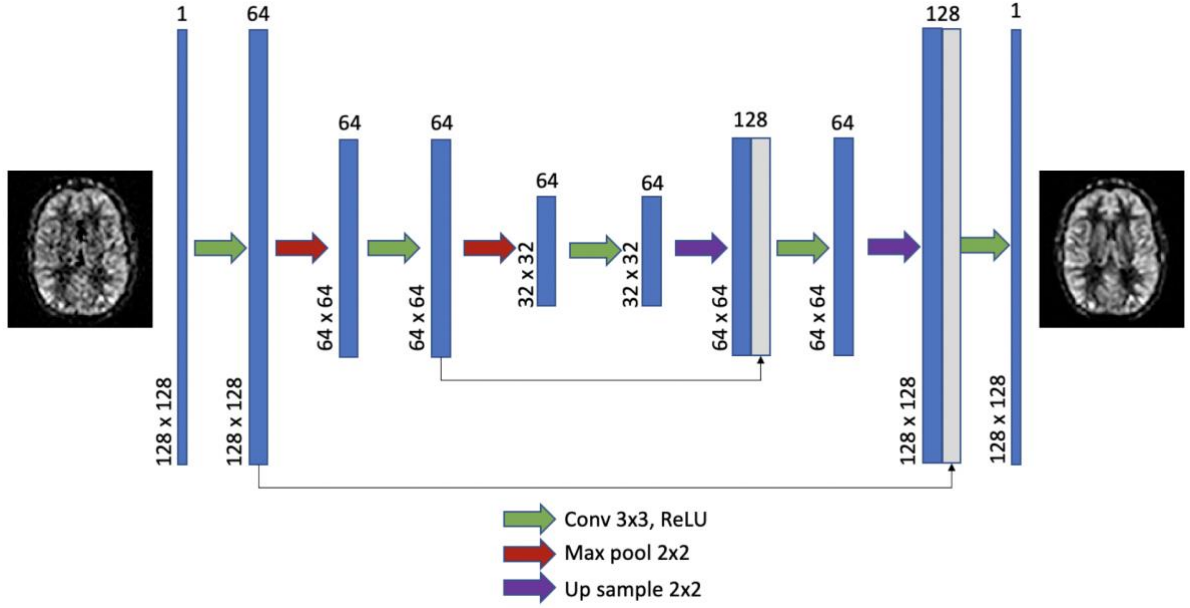


Figure 2. Architecture of the denoising auto-encoder model, with an example low-SNR, single repetition dM_{raw} image (left), and the corresponding high-SNR dM_{mean} image (right; same axial slice averaged over 10 repetitions). A total of 30,960 image pairs were used to train and validate the model. Image dimensions are shown for each step, along with the number of filter layers used. Skip connections are illustrated as horizontal lines, convolution operations (with subsequent ReLU activation) as green arrows, and max-pooling / up-sampling operations as red / purple arrows respectively.

The encoder component consisted of three convolution steps. Each convolution step employed 64 filter layers, each of which applied a 3x3 kernel, with padding used to maintain consistent image dimensions between the input and output. Following each convolution step, a rectified linear unit (ReLU) activation layer was added, followed by a 2x2 max pooling layer, in order to sub-sample the output by a factor of two. The decoder component mirrored the encoder architecture, with 2x2 up-sampling used between convolution operations, in order to reconstruct an output image with the same dimensions as the input. The last convolution step consisted of 1 filter layer only, with no ReLU activation, to produce the final image. Skip connections were added between the first two convolution steps on the encoding side and their counterparts on the decoding side. This allows image details captured in the feature maps from the encoding components to be concatenated with the feature maps produced during decoding, improving image restoration and the ability to train deeper networks (34). The use of batch normalisation was explored, but similar to (26), we found this did not

improve the performance of our model during training and validation, hence batch normalisation layers were not included in the final model.

Model Training

A batch size of 100 was used for model training. We employed the *RMSProp* optimizer with default settings in *Keras* (using the *TensorFlow* backend) to update the network's weights, and the mean squared error (MSE) was used for the loss function. Training was performed over 100 epochs, with an early-stopping criteria to interrupt the training when the loss in the validation data failed to improve over 5 consecutive epochs.

Comparison with Alternative Denoising Methods

Two alternative denoising techniques were used to compare the performance of the DAE against more established methods. The first was a Gaussian filter, in which the signal in a voxel is replaced by the weighted average of nearby voxels. This provides a simple method for increasing SNR in noisy data, albeit at the cost of a loss of sharpness in the resulting image. The second was a non-local means (NLM) filter (35). The principle of NLM is to average the value of a given voxel with values of other voxels in a limited neighbourhood, provided that the patches centred on the other voxels are similar enough to the patch centred on the voxel of interest. This provides effective image denoising, while potentially preserving fine structures and details in the image.

A subset of 500 training datasets were used to optimise the parameters for these alternative denoising methods, with the filter parameters which gave the minimum root-mean-square error (RMSE) between the denoised and the gold standard (dM_{mean}) images being optimal. For the Gaussian filter, the optimum window size was determined. For the NLM filter, the patch size, patch distance, and cut-off distance (h) were optimized. All filters were applied using Python 3.7: the *cv2* package was used for the Gaussian filter, and the *skimage* package was used for the NLM filter. Multi-parametric optimization of the NLM filter was performed using non-linear least-squares minimization, using the *lmfit* package.

Model Testing

Single PLD pCASL Data

Following model training, the new pCASL data acquired in 11 healthy subjects was used to test the efficacy of the DAE, Gaussian filter and NLM models on unseen data. The first repetition from each subject's raw dM dataset (using all axial slices) was used as the noisy input to the models (dM_{raw}). The denoised version of this was calculated for each model (dM_{Gauss} , dM_{NLM} , dM_{DAE}), for comparison with the gold standard dM_{mean} images. These dM images were then used to calculate CBF maps for each dataset (CBF_{raw} , CBF_{Gauss} , CBF_{NLM} , CBF_{DAE} , CBF_{mean}), using the standard method described in (1), with $\lambda=0.9$, $\alpha=0.85$, and $T_{1\text{bl}}=1.65\text{s}$. The CBF_{mean} map was used as the gold standard, against which alternative CBF maps were compared.

Multi-TI PASL Data

The performance of each model was also evaluated on unseen, multi-TI PASL data, acquired in 7 healthy subjects as described above. Here, the raw multi-TI difference images were denoised using each model. These datasets were then fit to the Buxton kinetic model (36), with CBF and bolus arrival time (BAT) as fitted parameters. The temporal width of the bolus was fixed at 700ms, due to the use of Q2TIPS saturation pulses during acquisition. Model fitting was performed using the *lmfit* Python package, and the goodness of fit in each voxel was calculated using χ^2 values (sum of squared residuals between the observed and fitted values over all TIs). Models were compared by calculating the mean χ^2 over all brain voxels within each subject.

Evaluation Metrics

In order to compare the denoising performance of each model, the SNR of each dM dataset was calculated. The noise in each image was determined by defining a noise mask, comprising a 15x15 voxel square patch at each corner of the axial images (containing non-brain, background signal only), and calculating the standard deviation of the signal in this area (σ). The *bet* algorithm in FSL (37) was used to define a brain mask for each subject, using the M_0 calibration image, and this was used to calculate the mean dM signal in the brain (S), over all axial slices. The SNR was defined as S/σ .

Similar to previous studies, the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) of the CBF images were used as additional evaluation metrics. PSNR was used to

define the accuracy of each CBF image in comparison to the gold standard, and was defined as $PSNR = 20 \cdot \log_{10} \cdot (CBF_{\max}/RMSE)$, where CBF_{\max} is the maximum value in the gold-standard CBF image, and $RMSE$ is the root mean square error between each CBF image and the gold-standard. Higher values of PSNR indicate that CBF images are more accurate, when compared to the gold standard. SSIM was used to quantify the visual quality of CBF maps, in comparison to the gold-standard (38). SSIM is thought to mimic the perceived quality of an image by a human observer, with values of 0 indicating no similarity, and 1.0 indicating perfect similarity. The *skimage* Python package was used to calculate SSIM values, using the default settings.

Statistics

The *SciPy* Python package was used for all statistical analysis. For comparison of evaluation metrics between models, the Levine test was used to test for equal variances (39). In cases of equal variance, one-way ANOVA tests were used for group comparisons, followed by a Tukey honestly significant difference post-hoc test. For un-equal variance, a Welch ANOVA test was used, followed by a Games-Howell post-hoc analysis. All p values are reported after correcting for multiple comparisons, with significance defined as $p < 0.05$.

Results

During model training, the early-stopping criterion was met after 18 epochs, resulting in a training time of 293 minutes, using a Mac with a 3.5 GHz Intel Core i5 CPU and 4 GB GPU. Following training, implementation of the DAE model on new pCASL datasets took approximately 340ms.

For the Gaussian filter, the optimum window size was 5 voxels. For the NLM filter, the optimum patch size was 6 voxels, the optimum patch distance was 13 voxels, and optimum cut-off distance was 6.0.

Exemplary results of the DAE are shown in Figure 3, demonstrating the model's ability to suppress the transient artefacts illustrated in Figure 2. The clinical example shown on the bottom row of Figure 3 illustrates a bright artefactual signal in the lateral ventricle, which

could be misinterpreted as a metastasis of the tumour in the temporal lobe. This artefactual signal is suppressed in both the gold standard and the denoised image.

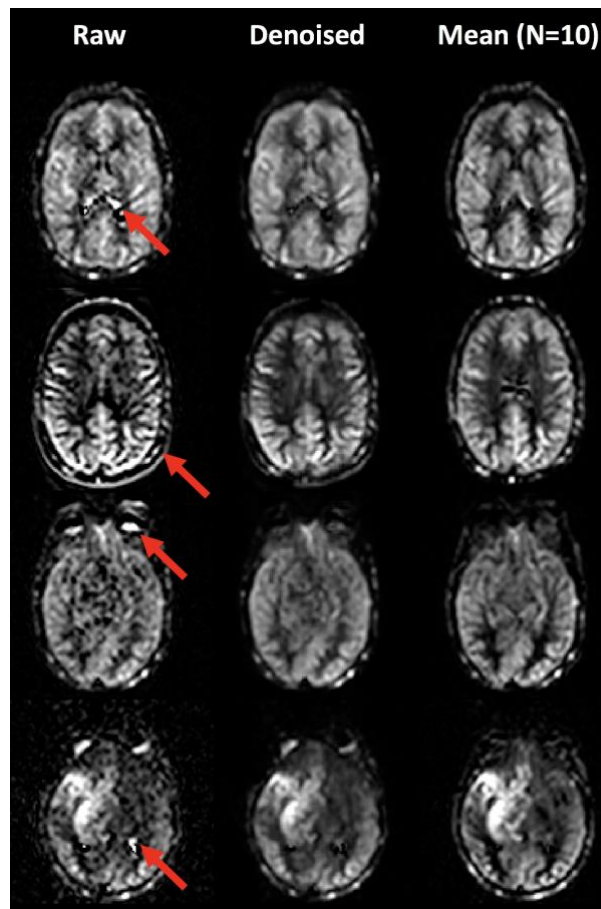


Figure 3. Example of the application of the DAE on un-seen data. The illustrative examples shown in Fig 1 are shown here before and after denoising with the DAE, along with a further clinical example (bottom row; paediatric brain tumour patient, showing diffuse hyper-perfusion in the right temporal lobe). dM images from a single repetition are shown in both their raw form (left column) and after denoising with the DAE (middle column). The equivalent image after averaging over 10 repetitions is shown in the right column. Transient artefacts are indicated with arrows.

Model Testing: pCASL Data in Healthy Subjects

The mean SNR of the dM_{raw} , dM_{Gauss} , dM_{NLM} , dM_{DAE} , dM_{mean} images acquired in 11 healthy subjects is shown in Figure 4A. Mean SNR was 28 ± 13 (\pm SD) in the raw images (range 10 - 49). The NLM and DAE denoising models produced dM images with significantly higher SNR than the raw images ($p < 0.01$), with the NLM filter producing the highest SNR, albeit with the greatest between-subject variance. Both the NLM (SNR = 198 ± 86) and DAE models (SNR = 101 ± 22) produced dM images with higher SNR than the gold standard (SNR = 75 ± 37). On

average, SNR in the raw dM images increased by a factor of 3.2 ± 1.5 after denoising with the DAE model, and by a factor of 6.2 ± 0.7 using the NLM model.

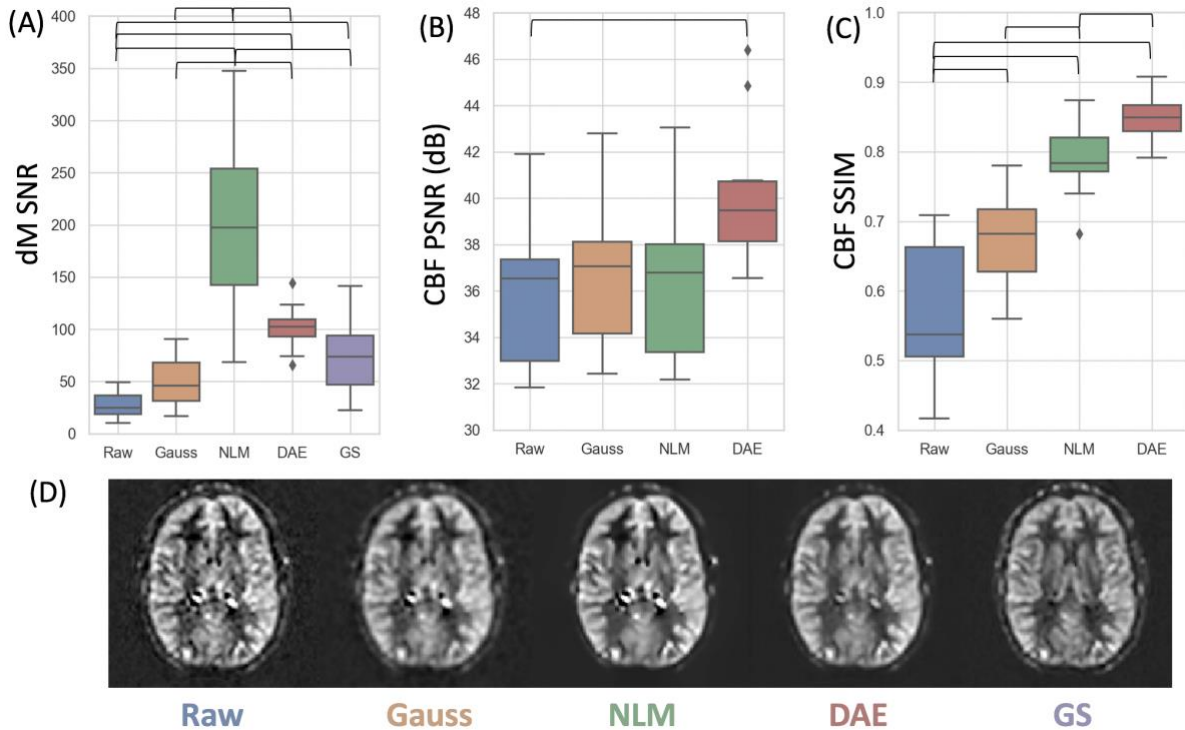


Figure 4 (A) Box and whisker plot of SNR in the dM images acquired in 11 healthy subjects, using the raw, Gaussian, non-local means (NLM), denoising autoencoder (DAE) and gold-standard (GS) datasets. (B) Peak SNR values, obtained using CBF images calculated from dM_{raw} , dM_{Gauss} , dM_{NLM} , dM_{DAE} datasets, compared to the gold-standard CBF images. (C) Structural similarity index (SSIM) values, demonstrating the visual similarity assessment between CBF maps generated by the different models, in comparison to the gold standard. Significant differences between groups ($p < 0.05$) are illustrated in each plot. (D) Example dM images from each model, in an axial slice from one representative subject. The artefactual CSF hyper-intensity seen in the raw dM image remains prominent in the dM_{Gauss} and dM_{NLM} images, but is attenuated in the dM_{DAE} image, which more closely resembles the gold standard.

The accuracy of the CBF images produced by each model, in comparison to the gold standard, is shown by the CBF PSNR values in Figure 4B. PSNR was highest in the CBF images produced using the DAE model (mean PSNR = 40.2 ± 3.0 dB), and this was the only model to produce a significant increase in PSNR compared to the raw CBF images (mean PSNR = 36.0 ± 3.4 dB, $p = 0.02$).

The CBF_{raw} , $\text{CBF}_{\text{Gauss}}$, CBF_{NLM} , CBF_{DAE} images all had significantly different SSIM values (Figure 4C), and structural similarity was highest in CBF images produced by the DAE model (mean $\text{SSIM} = 0.85 \pm 0.03$). As such, the DAE model produced CBF images with the greatest perceived visual quality, in comparison to the gold standard.

All models added some degree of blurring to the input image, and as such all denoised images were not as sharp as the gold standard. As expected, the Gaussian filter resulted in the greatest degree of blurring, and, given the superior performance in terms PSNR and SSIM values, overall the DAE model produced the images which was most closely matched to the gold standard.

Model Testing: Multi-TI PASL Data in Healthy Subjects

The mean voxel-wise χ^2 values, after fitting the Buxton kinetic model to multi-TI PASL data in 7 healthy subjects, are shown in Figure 5A. Model fitting using the dM_{Gauss} , dM_{NLM} , dM_{DAE} datasets resulted in significantly lower voxel-wise χ^2 values compared to model fitting based with the dM_{raw} datasets ($p < 0.05$, all comparisons). The Buxton fit to the dM_{DAE} images produced significantly lower χ^2 values than all other dM images (mean $\chi^2 (\text{DAE}) = 1.0 \pm 0.3$, $p < 0.05$, all comparisons). Example CBF, BAT, and χ^2 maps in an axial slice from a representative subject are shown in Figure 5B.

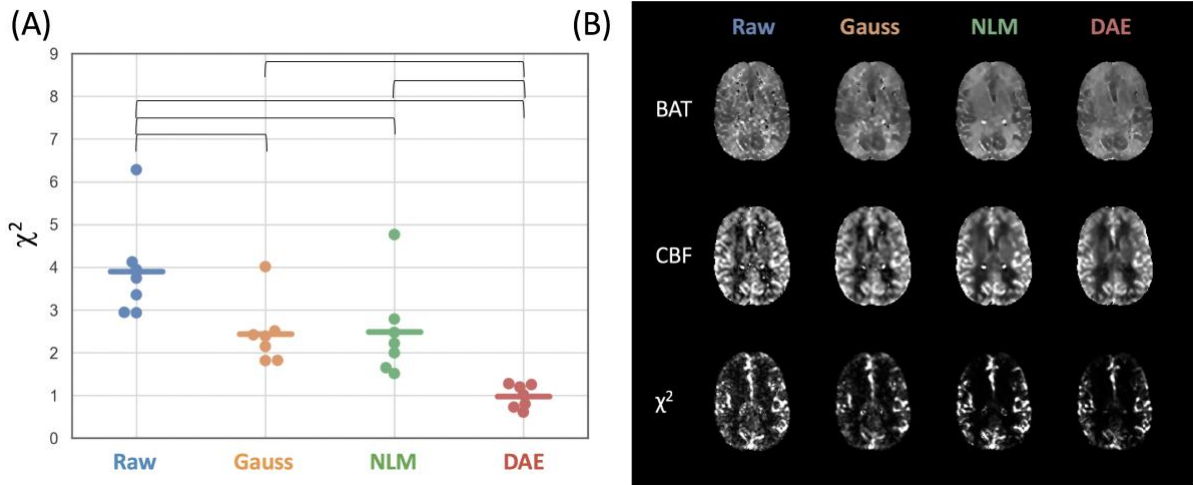


Figure 5 (A) Mean, voxel-wise χ^2 values (sum of squared residuals), after fitting the Buxton kinetic ASL model to dM images from the raw, Gaussian, non-local means (NLM), and denoising autoencoder (DAE) datasets. Data points represent the mean voxel-wise χ^2 values throughout the brain in individual subjects. Significant differences between groups are

indicated in the plot. (B) Example fitted maps of bolus arrival time (BAT), cerebral blood flow (CBF), and χ^2 values, in a representative subject.

Clinical Examples

Some further examples of the DAE model applied to illustrative clinical ASL images, none of which were used during model training, are shown in Figure 6. The brain tumour data shown in 6A and 6B were acquired at 1.5 T, using a similar pCASL protocol to that described in Section 2.1, but with thicker slices (5.0 mm). The data from the Sturge Weber patient shown in 6C were acquired at 3.0 T, again using a similar pCASL protocol to that described in Section 2.1, but with a PLD of 2000 ms.

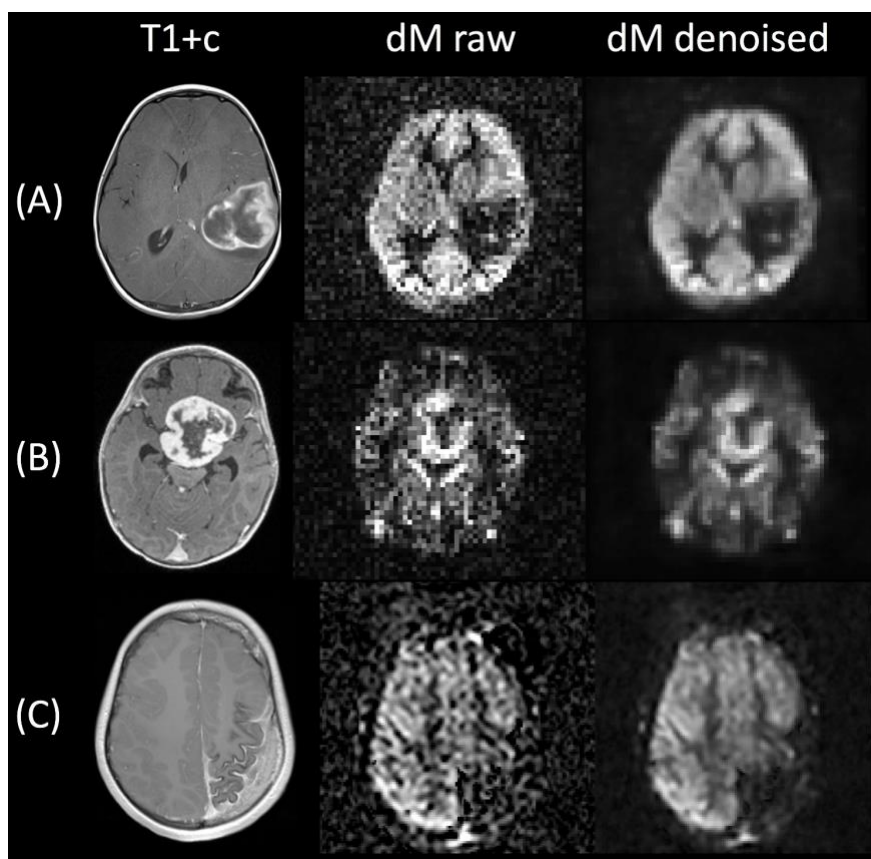


Figure 6 Clinical examples of implementation of the denoising autoencoder (DAE). T1-weighted images after injection of gadolinium contrast agent (T1+c) are shown in the left column. Raw pCASL dM images are shown in the centre column, with the equivalent denoised dM image shown on the right. Data shown in (A) and (B) were acquired at 1.5 T, data shown in (C) were acquired at 3.0 T. (A) 4 year old patient with a glioblastoma multiforme; (B) 0.9 year old patient with a pilocytic astrocytoma; (C) 11 year old patient with Sturge-Weber syndrome, demonstrating a cortical angioma and marked atrophy.

Discussion

In this work, we have developed a deep-learning model for denoising ASL images, based on an autoencoder architecture. Our model was effective at both increasing SNR and suppressing transient artefacts in low-SNR ASL images, producing CBF images with the greatest accuracy in comparison to the gold standard. Although the NLM model gave the greatest increase in SNR in this study, the CBF images generated by our proposed model had higher accuracy and greater structural similarity when compared to the gold standard images. This is due to the ability of our model to not only learn how to denoise images, but to identify artefactual signals in a single image. In comparison, traditional denoising approaches such as Gaussian and NLM filters are effective at improving SNR, but cannot learn to separate a prominent artefactual signal from a ‘true’ signal. As such, transient artefacts remain in the denoised CBF images, which results in reduced accuracy and poorer structural similarity.

ASL denoising models are particularly well suited to multi-TI ASL data, as typically fewer signal averages are acquired per TI in these acquisitions, in exchange for a wider coverage of inflow times. Although our DAE model was trained using single-PLD pCASL data, it also performed well on multi-TI PASL data, providing dM images which had the best fit to the widely used Buxton kinetic model. This represents a promising future application for our proposed model.

By training on large database of clinical pCASL scans, a further aim of this study was to produce a model which could differentiate between abnormal signals associated with pathology, and fluctuating abnormal signals associated with transient artefacts. As illustrated in Figures 3 and 6, our model performed well in this regard, producing denoised images in which artefactual signals were suppressed whilst pathological signals remained prominent. However, despite being trained on clinical paediatric datasets, our results also suggest the model performs well in healthy adult data, indicating our model also performs well under both pathological and non-pathological conditions.

In comparison to previous work, our study is unique in that no averaging, motion correction or smoothing was applied to the noisy images used as inputs during training. This was done to maximise the conspicuity of transient artefacts in the noisy images, so that the model could

effectively learn to suppress these, in conjunction with increasing SNR. One previous study also focussed on a deep-learning approach for joint denoising and suppression of transient artefacts (28), however, this required joint inputs relating to the mean and standard deviation of the ASL signal over multiple repetitions. In comparison, our proposed model can suppress transient artefacts in single subtraction images alone. Also, in contrast to some previous studies, our model does not rely on additional anatomical T1-weighted images (25) or a CBF signal model prior (24), which should improve the generalisability of our model.

Overall, the PSNR and SSIM values obtained in this study compare favourably to previous work (25, 26). This suggests that the DAE model presented here provides good denoising performance in comparison to previously published deep-learning-based denoising models, but could potentially eliminate the need to acquire multiple repetitions of the raw ASL data, which would have a significant impact on scan times.

Limitations

A potential limitation of our study was that we employed a relatively simple model architecture compared to some recent studies in this area (26). However, as mentioned above, the performance of our model compared favourably to previous work (25, 26). In addition, we did not perform model training and validation across noisy data generated using different levels of signal averaging in this work, as this would have suppressed the influence of transient artefacts in the input data. However, there was considerable variability in the noisy data used to test this model, with the SNR of the raw data in the healthy subjects ranging from 10 – 49. As such, our results indicate that our model performs well across a range of SNR levels.

Conclusion

In this work, we have proposed a deep-learning-based framework for simultaneous denoising and suppression of transient artefacts in ASL images. The model is effective at differentiating abnormal signals associated with biological variability from abnormal signals associated with transient artefacts. The model works effectively on low-SNR ASL data acquired without signal averaging, and produces CBF maps which show good agreement with those acquired with 10 signal averages. As such our model could provide a significant saving in the scan time required

to acquire ASL data. Our results also suggest the model would be well suited for multi-TI acquisitions, in which signal averaging is often sacrificed in exchange for greater inflow-time coverage.

References

1. Alsop DC, Detre JA, Golay X, et al.: Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: A consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. *Magn Reson Med* 2014; 73:102–116.
2. Ye FQ, Frank JA, Weinberger DR, McLaughlin AC: Noise reduction in 3D perfusion imaging by attenuating the static signal in arterial spin tagging (ASSIST). *Magnetic Resonance in Medicine* 2000; 44:92–100.
3. Günther M, Oshio K, Feinberg DA: Single-shot 3D imaging techniques improve arterial spin labeling perfusion measurements. *Magnetic Resonance in Medicine* 2005; 54:491–498.
4. Deibler AR, Pollock JM, Kraft RA, Tan H, Burdette JH, Maldjian JA: Arterial Spin-Labeling in Routine Clinical Practice, Part 1: Technique and Artifacts. *AJNR Am J Neuroradiol* 2008; 29:1228–1234.
5. Amukotuwa SA, Yu C, Zaharchuk G: 3D Pseudocontinuous arterial spin labeling in routine clinical practice: A review of clinically significant artifacts. *J Magn Reson Imaging* 2016; 43:11–27.
6. Tan H, Maldjian JA, Pollock JM, et al.: A fast, effective filtering method for improving clinical pulsed arterial spin labeling MRI. *Journal of Magnetic Resonance Imaging* 2009; 29:1134–1139.
7. Maumet C, Maurel P, Ferré J-C, Barillot C: Robust estimation of the cerebral blood flow in arterial spin labelling. *Magnetic Resonance Imaging* 2014; 32:497–504.
8. Dolui S, Wang Z, Shinohara RT, Wolk DA, Detre JA, Alzheimer's Disease Neuroimaging Initiative: Structural Correlation-based Outlier Rejection (SCORE) algorithm for arterial spin labeling time series. *J Magn Reson Imaging* 2017; 45:1786–1797.
9. Li Y, Dolui S, Xie D-F, Wang Z, Alzheimer's Disease Neuroimaging Initiative: Priors-guided slice-wise adaptive outlier cleaning for arterial spin labeling perfusion MRI. *J Neurosci Methods* 2018; 307:248–253.
10. Shirzadi Z, Crane DE, Robertson AD, et al.: Automated removal of spurious intermediate cerebral blood flow volumes improves image quality among older patients: A clinical arterial spin labeling investigation. *J Magn Reson Imaging* 2015; 42:1377–1385.
11. Behzadi Y, Restom K, Liao J, Liu TT: A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* 2007; 37:90–101.

12. Wells JA, Thomas DL, King MD, Connelly A, Lythgoe MF, Calamante F: Reduction of errors in ASL cerebral perfusion and arterial transit time maps using image de-noising. *Magn Reson Med* 2010; 64:715–724.
13. Zhu H, Zhang J, Wang Z: Arterial spin labeling perfusion MRI signal denoising using robust principal component analysis. *J Neurosci Methods* 2018; 295:10–19.
14. Spann SM, Kazimierski KS, Aigner CS, Kraiger M, Bredies K, Stollberger R: Spatio-temporal TGV denoising for ASL perfusion imaging. *Neuroimage* 2017; 157:81–96.
15. Wang J, Aguirre GK, Kimberg DY, Detre JA: Empirical analyses of null-hypothesis perfusion fMRI data at 1.5 and 4 T. *Neuroimage* 2003; 19:1449–1462.
16. Fazlollahi A, Bourgeat P, Liang X, et al.: Reproducibility of multiphase pseudo-continuous arterial spin labeling and the effect of post-processing analysis methods. *Neuroimage* 2015; 117:191–201.
17. Bibic A, Knutsson L, Ståhlberg F, Wirestam R: Denoising of arterial spin labeling data: wavelet-domain filtering compared with Gaussian smoothing. *MAGMA* 2010; 23:125–137.
18. Petr J, Ferre J-C, Gauvrit J-Y, Barillot C: Improving arterial spin labeling data by temporal filtering. In *Medical Imaging 2010: Image Processing. Volume 7623*. International Society for Optics and Photonics; 2010:76233B.
19. Liang X, Connelly A, Calamante F: Voxel-Wise Functional Connectomics Using Arterial Spin Labeling Functional Magnetic Resonance Imaging: The Role of Denoising. *Brain Connect* 2015; 5:543–553.
20. Jiang D, Dou W, Vosters L, Xu X, Sun Y, Tan T: Denoising of 3D magnetic resonance images with multi-channel residual learning of convolutional neural network. *Jpn J Radiol* 2018; 36:566–574.
21. Manjón JV, Coupe P: MRI Denoising Using Deep Learning. In *Patch-Based Techniques in Medical Imaging*. Edited by Bai W, Sanroma G, Wu G, Munsell BC, Zhan Y, Coupé P. Cham: Springer International Publishing; 2018:12–19. [*Lecture Notes in Computer Science*]
22. Benou A, Veksler R, Friedman A, Riklin Raviv T: Ensemble of expert deep neural networks for spatio-temporal denoising of contrast-enhanced MRI sequences. *Medical Image Analysis* 2017; 42:145–159.
23. Kim KH, Choi SH, Park S-H: Improving Arterial Spin Labeling by Using Deep Learning. *Radiology* 2018; 287:658–666.
24. Ulas C, Tetteh G, Kaczmarz S, Preibisch C, Menze BH: DeepASL: Kinetic Model Incorporated Loss for Denoising Arterial Spin Labeled MRI via Deep Residual Learning. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Edited by Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G. Cham: Springer International Publishing; 2018:30–38. [*Lecture Notes in Computer Science*]

25. Gong K, Han P, El Fakhri G, Ma C, Li Q: Arterial spin labeling MR image denoising and reconstruction using unsupervised deep learning. *NMR Biomed* 2019:e4224.
26. Xie D, Li Y, Yang H, et al.: Denoising arterial spin labeling perfusion MRI with deep machine learning. *Magnetic Resonance Imaging* 2020.
27. Xie D, Bai L, Wang Z: Denoising Arterial Spin Labeling Cerebral Blood Flow Images Using Deep Learning. *arXiv:180109672 [cs]* 2018.
28. Owen D, Melbourne A, Eaton-Rosen Z, et al.: Deep Convolutional Filtering for Spatio-Temporal Denoising and Artifact Removal in Arterial Spin Labelling MRI. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Edited by Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G. Cham: Springer International Publishing; 2018:21–29. [Lecture Notes in Computer Science]
29. Gondara L: Medical image denoising using convolutional denoising autoencoders. *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)* 2016:241–246.
30. Bermudez C, Plassard AJ, Davis LT, Newton AT, Resnick SM, Landman BA: Learning Implicit Brain MRI Manifolds with Deep Learning. *Medical Imaging 2018: Image Processing* 2018:56.
31. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol P-A: Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *J Mach Learn Res* 2010; 11:3371–3408.
32. Xie J, Xu L, Chen E: Image denoising and inpainting with deep neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. Lake Tahoe, Nevada: Curran Associates Inc.; 2012:341–349. [NIPS'12]
33. Cho K: Boltzmann Machines and Denoising Autoencoders for Image Denoising. *arXiv:13013468 [cs, stat]* 2013.
34. Mao X-J, Shen C, Yang Y-B: Image Restoration Using Convolutional Auto-encoders with Symmetric Skip Connections. *arXiv* 2016; 1606.08921 [cs.CV].
35. Buades A, Coll B, Morel J-M: A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Volume 2; 2005:60–65 vol. 2.
36. Buxton RB, Frank LR, Wong EC, Siewert B, Warach S, Edelman RR: A general kinetic model for quantitative perfusion imaging with arterial spin labeling. *Magnetic Resonance in Medicine* 1998; 40:383–396.
37. Smith SM: Fast robust automated brain extraction. *Hum Brain Mapp* 2002; 17:143–155.
38. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP: Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 2004; 13:600–612.

39. Olkin I: *Contributions to Probability and Statistics; Essays in Honor of Harold Hotelling*. Stanford, Calif.: Stanford University Press; 1960.