Deep convolutional filtering for spatio-temporal denoising and artifact removal in arterial spin labelling MRI

David Owen[1], Andrew Melbourne[1], Zach Eaton-Rosen[1], David L Thomas[1,2], Neil Marlow[3], Jonathan Rohrer[2], and Sebastien Ourselin[4]

[1]Translational Imaging Group, University College London, UK
[2]Dementia Research Centre, Institute of Neurology, University College London, UK
[3]Institute for Women's Health, University College London, UK
[4]School of Biomedical Engineering and Imaging Sciences, King's College London

Abstract. Arterial spin labelling (ASL) is a noninvasive imaging modality, used in the clinic and in research, which can give quantitative measurements of perfusion in the brain and other organs. However, because the signal-to-noise ratio is inherently low and the ASL acquisition is particularly prone to corruption by artifact, image processing methods such as denoising and artifact filtering are vital for generating accurate measurements of perfusion. In this work, we present a new simultaneous approach to denoising and artifact removal, using a novel deep convolutional joint filter architecture to learn and exploit spatio-temporal properties of the ASL signal. We proceed to show, using data from 15 healthy subjects, that our approach achieves state of the art performance in both denoising and artifact removal, improving peak signal-to-noise ratio by up to 50%. By allowing more accurate estimation of perfusion, even in challenging datasets, this technique offers an exciting new approach for ASL pipelines, and might be used both for improving individual images and to increase the power of research studies using ASL.

1 Introduction

Arterial spin labelling (ASL) is an MR imaging technique that allows quantitative, noninvasive measurements of perfusion in the brain and other organs. ASL has demonstrated its utility in both research and clinical use, and has the potential to be used as a biomarker in several diseases [1]. However, ASL suffers from the twin problems of having low signal-to-noise ratio (SNR) and being prone to artifacts from patient motion, RF coil instability and several other sources.

Typically, to address these problems, denoising and artifact filtering are used. Denoising uses statistical properties of the ASL signal to improve the effective SNR, for example by modelling the signal using total variation priors, a wavelet basis, or anatomy-derived spatial correlation [2, 3]. Denoising methods tend to assume Gaussian noise, and are not usually robust to non-Gaussian artifacts, for example due to patient motion or hardware instability. Artifact filtering

methods, conversely, remove or down-weight parts of the ASL signal that have severe artifacts, allowing subsequent processing to assume Gaussian noise [4–7].

Denoising and artifact removal are usually considered in isolation from one another, but are overlapping problems: noise is often neither strictly Gaussian nor spatially homogeneous, and artifact filtering often results in the rejection of entire image volumes when only a fraction of the image is thoroughly corrupted. In this work, we develop a deep convolutional neural network (CNN) for simultaneous denoising and artifact filtering, making full use of the available data. Inspired by cutting-edge developments in computer vision, we create a novel deep learning architecture that can relate noisy, artifact-corrupted ASL images to the true underlying perfusion. This architecture uses joint convolutional filtering [8] in order to efficiently extract spatio-temporal information from the ASL signal, allowing our method to distinguish artifact from noise. We present results from our method in ASL data from 15 healthy volunteers, showing that our method improves on the state of the art for both artifact filtering and denoising, increasing the peak SNR by up to 50%. These promising initial results show that deep convolutional joint filtering holds great promise for ASL processing, and suggest our approach might be useful both for improving individual subjects' images and for increasing the statistical power of neuroimaging studies.

2 Methods

2.1 Arterial spin labelling

ASL images are acquired by tagging blood magnetically – applying inversion pulses at the neck before the blood flows to the brain. Images are acquired with and without this tagging, with the difference between these images being a function of the blood flow. Standard models exist in the literature to relate the measured signal to the underlying perfusion [9, 1].

In this work, we use an ASL dataset from 35 healthy 19-year-old volunteers (F/M=17/18). Images were acquired on a 3T Phillips Achieva with 2D EPI pseudo-continuous ASL using 30 control-label pairs, PLD=1800ms+41ms/slice, τ = 1650ms, $3 \times 3 \times 5$mm. We also acquired M_0 images and 3D T_1-weighted volumes at 1mm isotropic resolution. All ASL data were motion corrected via rigid registration before being used – note, however, that motion correction often does not fully compensate for subject motion, and this is one of the artifacts that should ideally be filtered when estimating the true perfusion.

2.2 Deep convolutional joint filtering

Convolutional neural networks (CNNs) are a well-established means of processing images for a variety of tasks. Here, we focus on *pixel-to-pixel* networks: in general terms, we wish to take input images and produce a higher quality output image. Although CNNs show state of the art performance in natural image processing [8], their application to ASL images has, to date, been limited to

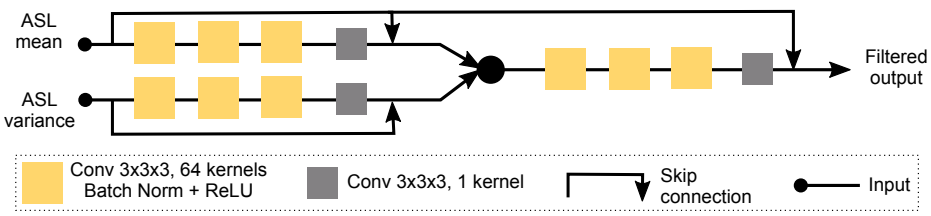

Fig. 1: Architecture diagram for our deep convolutional joint filter. For the mean-only filter, the ASL variance input is not used.

Hadamard-encoded ASL, and has had limited success [10]. Part of the reason may be the great benefit to ASL processing of *spatio-temporal* models: it is easier to distinguish artifact from perfusion when one is aware of whether a particular part of the signal was transient or permanent [2]. Processing a sequence of several 3D images rapidly becomes computationally expensive, however. Our solution is to explicitly feed in temporal variance information so the network has two inputs: the mean ASL signal, and the ASL signal variance over time.

The naive approach to using spatio-temporal information would be to feed the signal mean and variance maps into the same CNN, similarly to colour channels in a natural image. However, initial testing showed poor performance, as the network was unable to translate voxel-level features into meaningful cross-channel information. To achieve this end, we created a novel *joint convolutional filter* architecture, inspired by image processing approaches to integrating RGB cameras and depth sensors [8]. In our architecture, information is extracted and processed in parallel from the mean ASL image and the ASL temporal variance. These images are combined at a later stage in the network, with several more layers used to extract meaningful features from their combination. Skip connections in the parallel stages improve network convergence and robustness, as well as transferring global information in the learning process [8].

To train our models, we first identified artifact-corrupted volumes using the filtering method of Tan et al [4]. We generated gold standard high-quality perfusion maps by removing these outlying volumes and using all of the remaining data to fit perfusion according to literature recommendations [1]. These gold standard images were used as ground truth. The inputs to our network were derived by taking 10 random volumes from the ASL series, including artifact-corrupted volumes to train the network to correct for artifacts in addition to denoising. The loss function used was mean squared error within the brain mask.

We implemented our CNNs in Keras,using the Adam optimiser with learning rate 0.01 with 20 subjects for training and 15 subjects for validation. To avoid overfitting and improve generalisation, we augmented with random translations sampled uniformly up to 5mm in each dimension. We also augmented input images with Gaussian noise, magnitude approximately 1% of the ASL noise as estimated from gold standard data. We trained to convergence (approximately 1000 iterations), which took 12 hours using an NVIDIA K80 graphics card.

2.3 Comparison to pre-existing methods and validation

For both denoising and artifact filtering, we compare our method with a state of the art spatial regularisation technique using total generalised variation (TGV), which has been shown to produce reliable and accurate denoising with built-in artifact rejection via robust statistics [2]. For reference, we also compare against voxelwise fitting with no spatial regularisation – this remains a very common way to process ASL data, and acts as a representative baseline.

We evaluate our method, using the full spatio-temporal information as discussed in Section 2.2, and we also evaluate a simpler CNN architecture using only spatial information (see Figure 1), to show the benefit of the joint filter. We evaluate the performance by examining filtered images for residual artifacts, and by producing maps of absolute error relative to the gold standard perfusion map. These are shown in Figure 2. We show slices from subject 7, where there is extreme artifact; and subject 4, with less severe artifact. Subsequently we perform quantitative validation by calculating the PSNR for each denoising method, again calculated relative to the gold standard[1].

Often, outlier filtering is performed as a separate step prior to denoising; so we also compare against TGV and voxelwise fitting with explicit outlier rejection via z-score filtering [4]. Our validation dataset was chosen such that each subject contains one or more artifact volumes, as identified by z-score filtering on the full dataset. We remove artifact volumes for the reference methods, showing how they would perform in conjunction with z-score filtering. For our joint filter, however, we do not remove the volumes in this comparison, as the purpose of the joint filter was to use the non-artifact information within partially-corrupted images. As before, we evaluate example images from subjects 4 and 7, and then we present quantitative validation over all subjects using PSNR calculations.

3 Results

3.1 Example images

Figure 2 shows example axial slices from subjects 4 and 7, as well as maps of absolute error. For subject 7, there is a strong hyperintense ring artifact near the front of the brain. Similarly, for subject 4, several artifacts present as extreme intensity changes, mostly near the edges of the brain. Voxelwise fitting shows these most plainly in both subjects, as the fitting has no implicit artifact removal. TGV results in heavily smoothed images, removing some of the artifact seen in the voxelwise images, but also losing detail in the image. CNN mean-only smooths away even more spatial detail than TGV, and shows a similar pattern of artifact to voxelwise fitting. However, the joint filter produces a significantly less artifact-prone image, as well as improved denoising.

Figure 3 shows example axial slices for subjects 4 and 7 again, this time preprocessed with artifact removal as a separate step. This is a more realistic

[1] $PSNR = 20 \log_{10}(S_{max}/RMSE)$, where S_{max} is the maximum ASL signal over all voxels and $RMSE$ is the root mean square error

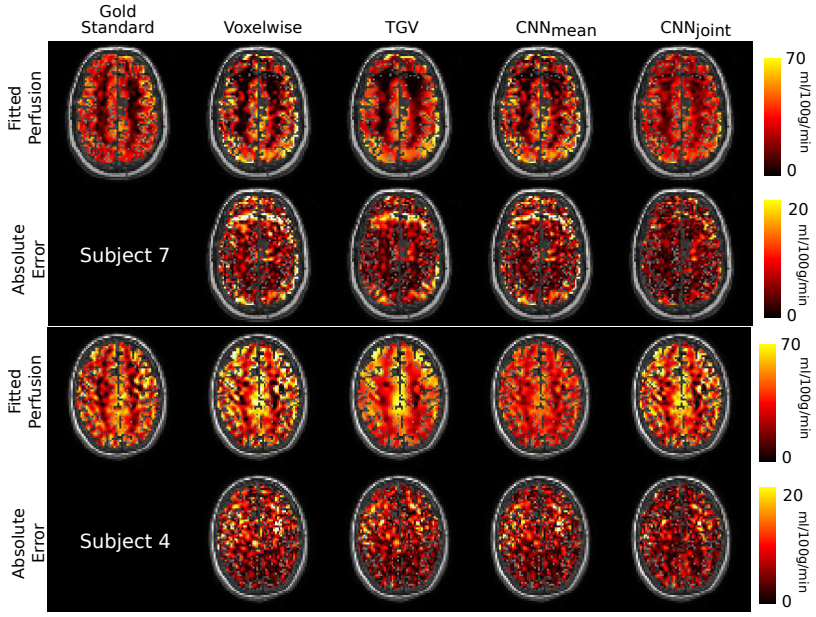

Fig. 2: Example fits and errors, no separate artifact filtering step.

comparison – certainly for voxelwise smoothing, which has no built-in artifact rejection. Here, the mean-only CNN performs closer to the joint CNN, although the joint CNN continues to produce visibly better denoising. Moreover, the remaining artifacts have been better removed by the joint CNN, despite the joint CNN being the only method to have no explicit artifact rejection before fitting.

3.2 Quantitative evaluation via PSNR

Figure 4 shows the PSNR for each subject and method, when there is no explicit artifact filtering. Because there are relatively few ASL images, and there is large inter-subject variability in the artifacts and global perfusion, PSNR varies greatly across subjects. To assist comparison between methods, Figure 4 shows change in PSNR relative to voxelwise fitting for each subject. The joint CNN produces the best result in all subjects except subject 15, where TGV performs marginally better. The average per-subject improvement of the joint CNN over TGV is 1.25dB ($p < 0.01$), although for some subjects with extreme artifact the improvement can be as high as 6dB. Crucially, while the mean-only CNN is often worse than TGV, the joint filter outperforms it significantly ($p < 0.05$ for each) in 11/15 subjects, marginally in 3/15 subjects, and is marginally worse (0.72dB, $p = 0.07$) only in subject 15. Although joint filtering does not always produce significantly better results than the mean-only CNN, it is significantly better than the mean-only CNN ($p < 0.05$) in nine subjects.

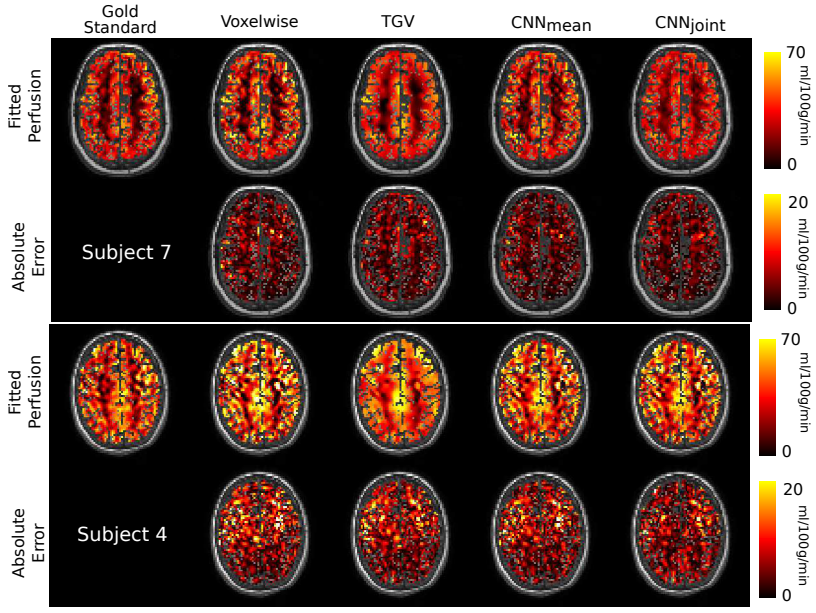

Fig. 3: Example fits and errors, separate artifact filtering performed before fitting.

Figure 5 shows the PSNR for each subject and each method, when there is an explicit artifact filtering step as described in Section 2.3. Joint filtering again performs the best, always better than or comparable to the runner-up. Joint filtering is significantly ($p < 0.05$) better than TGV in 13/15 subjects, and marginally superior in subjects 9 and 13. The average improvement over TGV per subject is 1.64dB ($p < 0.001$). Moreover, the second-best method is typically the mean-only CNN – with this explicit filtering step, even the mean-only CNN consistently outperforms TGV filtering ($p < 0.05$ for 12/15 subjects). This is reasonable: when there is less artifact influence, temporal information is less important and the problem becomes one of spatial regularisation, where CNNs excel. Additionally, TGV often performs worse than voxelwise fitting – over-regularising the fits based on the scarce data remaining after filtering.

4 Discussion and conclusions

As demonstrated by the visible improvements in image quality (Figure 2) and the significant increase in PSNR (Figure 4), our joint filtering approach performs better than state of the art for denoising in the presence of artifact. Of particular note is the filter's strong performance in artifact removal – in subject 7, for example, a prominent edge artifact is removed completely from the output image without any appreciable drop in denoising. The superior performance of the joint filter, compared with a mean-only CNN, shows the value in providing temporal variance information when processing artifact-prone data.

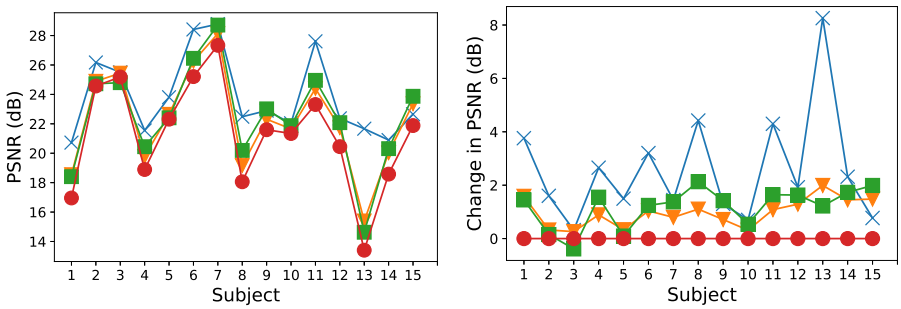

Fig. 4: PSNR for each subject and method, no separate artifact filtering step. Key: × joint CNN, ▼ mean-only CNN, ■ TGV, ● voxelwise.

Compared with pipelines involving separate filtering and denoising, our method again outperforms state of the art (Figures 3 and 5). By retaining parts of a corrupted volume, more information can be used in denoising, meaning the joint filter performs better than mean-only CNN filtering in most subjects. This is evidence the joint filter is able to perform better, on average, than combining a simpler CNN approach with explicit artifact filtering. Moreover, even the mean-only CNN is itself an advance on state of the art: this approach outperforms TGV in 12/15 subjects when filtering is applied separately.

Future work will involve validation across different ASL acquisitions and subject populations, leading the way for use in neuroimaging studies and the clinic. Future work might also explore alternative ways to exploit temporal information, for example through a recurrent-convolutional architecture. More importantly, any method should handle variations in ASL data such as readout and label type. Currently this requires retraining on each new dataset, but it may be possible for a single network to handle these different cases. Finally, a limitation of this work is the necessity for higher-quality data (e.g. more ASL volumes) in a subset of subjects for training; so we wish to explore how cross-validation derived loss

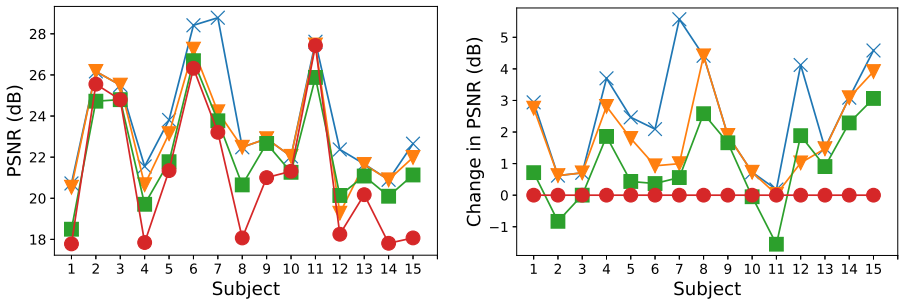

Fig. 5: PSNRs with artifact filtering before denoising as described in Section 2.3.

functions might ameliorate this. To this end, transfer learning may be helpful to reduce the computational cost of retraining in several cross-validation folds.

The innovative joint approach to denoising and artifact filtering presented here has the potential to substantially increase the quality of ASL images, even salvaging datasets that were previously considered unusable. By fusing temporal variance information with spatial information in a novel network architecture, our deep convolutional joint filter method outperforms state of the art in both denoising and filtering. Our method is applicable to any ASL data, subject to training requirements, and could even be used in other imaging modalities. Consequently, deep convolutional joint filtering presents an exciting future direction for medical image processing in noisy and artifact-prone modalities, and may eventually be used to improve the statistical power of neuroimaging studies.

Acknowledgements We acknowledge the MRC (MR/J01107X/1), the National Institute for Health Research (NIHR), the EPSRC (EP/H046410/1) and the NIHR University College London Hospitals Biomedical Research Centre (NIHR BRC UCLH/UCL High Impact Initiative BW.mn.BRC10269). This work is supported by the EPSRC-funded UCL Centre for Doctoral Training in Medical Imaging (EP/L016478/1) and the Wolfson Foundation.

References

1. Alsop, D., et al.: Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications. MRM **73**(1) (2015) 102–116
2. Spann, S., Kazimierski, K., Aigner, C., et al.: Spatio-temporal TGV denoising for ASL perfusion imaging. Neuroimage (2017)
3. Owen, D., Melbourne, A., Eaton-Rosen, Z., et al.: Anatomy-driven modelling of spatial correlation for regularisation of arterial spin labelling images. In: MICCAI, Springer (2017) 190–197
4. Tan, H., Maldjian, J.A., Pollock, J.M., et al.: A fast, effective filtering method for improving clinical pulsed arterial spin labeling MRI. JMRI **29**(5) (2009) 1134–1139
5. Wang, Z.: Improving cerebral blood flow quantification for arterial spin labeled perfusion MRI by removing residual motion artifacts and global signal fluctuations. MRM **30**(10) (2012) 1409–1415
6. Shirzadi, Z., Crane, D.E., Robertson, A.D., et al.: Automated removal of spurious intermediate cerebral blood flow volumes improves image quality among older patients: a clinical arterial spin labeling investigation. MRM **42**(5) (2015) 1377–1385
7. Tanenbaum, A.B., Snyder, A.Z., Brier, M.R., et al.: A method for reducing the effects of motion contamination in arterial spin labeling MRI. Journal of Cerebral Blood Flow & Metabolism **35**(10) (2015) 1697–1702
8. Li, Y., Huang, J.B., Narendra, A., Yang, M.H.: Deep joint image filtering. In: European Conference on Computer Vision. (2016)
9. Buxton, R., et al.: A general kinetic model for quantitative perfusion imaging with arterial spin labeling. MRM **40**(3) (1998) 383–396
10. Kim, K.H., Choi, S.H., Park, S.H.: Improving arterial spin labeling by using deep learning. Radiology (2017) 171154