

Super-resolution and segmentation of 4D Flow MRI using Deep learning and Weighted Mean Frequencies

Simon Perrin¹[0009-0005-2008-2802], Sébastien Levilly¹[0000-0002-2704-0315],
Harold Mouchère¹[0000-0001-6220-7216], and Jean-Michel
Serfaty²[0000-0001-8801-1414]

¹ Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000
Nantes, France

`{simon.perrin,sebastien.levilly,harold.mouchere}@univ-nantes.fr`

² CHU Nantes, CNRS, INSERM, L'Institut du Thorax, Nantes Université, 44000
Nantes, France

`jeanmichel.serfaty@chu-nantes.fr`

Abstract. 4D Flow MRI is a promising imaging sequence that provides 3D anatomy and velocity along the cardiac cycle. However, hemodynamic biomarkers are susceptible to degradation due to the low resolution of the imaging modality, which can compromise vessel segmentation. In this study, we propose a novel deep-learning approach, named SURFR-Net, that combines both super-resolution and segmentation tasks, leading to a super-resolved segmentation. SURFR-Net is based on the RCAN super-resolution network, modified to handle a multi-task problem. A novel handcraft feature, named Weighted Mean Frequencies (WMF), has been introduced with the objective of assisting the network in differentiating between pulsatile and non-pulsatile fluid regions. Moreover, we demonstrate the use of WMF feature as input to enhance super-resolution and provide a more relevant segmentation on 4D Flow MRI images. The proposed solution has been shown to outperform the state-of-the-art solution, SRFlow, in terms of direction and quantification error on systolic and diastolic times with a maximum gain of 4.1% in relative error. Furthermore, this study demonstrates the benefit of combining the super-resolution with the segmentation in a multi-task framework on both outcomes. In conclusion, the proposed solution has the capacity to facilitate a super-resolved segmentation of the aorta, thereby potentially addressing the primary concern regarding 4D Flow MRI parietal biomarkers, such as wall shear stress.

Keywords: 4D Flow MRI · Super-resolution · Segmentation · Multi-Task · Deep Learning

1 Introduction

In the last decades, 4D Flow MRI emerged as a promising solution to observe both the anatomy and velocity field within a 3D volume and along the cardiac

cycle [8]. In practice, technological constraints lead to a compromise between acquisition time, resolution and signal-to-noise ratio. Due to constrained acquisition time in the clinical routine, the image resolution is generally the most degraded parameter, *eg.* for a thoracic aorta approximately 2.5mm isotropic (ISO) [8], along with the signal-to-noise ratio. Furthermore, some hemodynamic biomarkers, based on velocity derivative, suffer particularly from noise and the lack of resolution. In the case of the wall shear stress estimation, [6] demonstrated the negative impact of both noise and resolution from 4D Flow MRI images, and also the segmentation accuracy.

In the literature, the resolution and noise enhancement task has been tackled by different strategies: image processing by inverse problem [7, 3] or deep learning [5, 12]. These solutions have been designed to provide a super-resolved velocity field without any segmentation information. Actually, these studies have demonstrated quantitatively their performances particularly on the velocity field super-resolution.

Hemodynamic biomarkers being impacted by both the resolution and the segmentation accuracy [6], super-resolution approaches do not consider this part of the error sources. One can consider using recent deep learning solutions for segmentation that present promising performance [1, 2]. However, these solutions rely on the 4D Flow MRI images and then provide a segmentation at the same resolution.

As a contribution, we propose to combine both tasks of super-resolution and segmentation in a deep learning framework named SURFR-Net. Both tasks have to deal with the fluid and non-fluid areas, and we expect the features to better generalize these areas. Moreover, the resulting segmentation is obtained at the super-resolved resolution contrary to state-of-the-art solutions [1, 2]. Besides, we introduce the handcraft feature, called Weighted Mean Frequencies, as an input to provide temporal information in our time independent solution. The proposed model SURFR-Net is presented in Sec. 2 and results are discussed in Sec. 3.

2 Method

2.1 Data

The proposed solution relies on two different image types for both training and evaluation: synthetic images for the velocity super-resolution task and 4D Flow MRI images along with their associated masks for the segmentation task.

In fluid mechanics, Computational Fluid Dynamics (CFD) simulations are considered as reference to characterize small scale velocity pattern [10]. In this work, CFD simulations have been used to generate synthetic 4D Flow MRI images. As CFD data, we choose the public dataset and the data preprocess used in Ferdian *et al.* (2020) [4] (data are under licence CC-BY 4.0). Dataset is composed of 3 CFD simulations where each of them used a different geometry. First geometry, named *aorta01*, was acquired from a healthy volunteer meanwhile *aorta02* and *aorta03* present a coarctation of the aorta. For CFD data, we use the

same data preparation method as Ferdian *et al.* (2020) [4]. CFD are downsampled directly in the k-space (MRI domain acquisition) to produce low resolution and noisy CFD, *ie.* synthetic 4D Flow MRI data.

MRI data are constituted of 9 1.5T images from patients who presented a cardiovascular event. The research was carried out following the principles of the Declaration of Helsinki. These images are only used for the segmentation task. The segmentations were made for one timeframe of the images, and then recopied for each timeframe. Segmentations were done on clinical resolution MRI since high-resolution images were not available. The low resolution segmentations are upsampled by a factor of 2 using nearest neighbour interpolation. As these images are only used for the segmentation task, we don't need any phase high resolution image.

For the split between training and validation, *aorta01*, *aorta02* and seven MRI volumes are used as training data meanwhile *aorta03* and two MRI images for validation. Patches with a sized of $32 \times 32 \times 16$ are extracted from CFD and MRI data. CFD images are cut into 10 patches per timeframe leading to a total number of patch of 710 for one CFD simulation. By timeframe, 9 of the 10 patches have to be covered by at least 20% of the fluid domain. For validation on CFD, one patch is extracted per timeframe respecting the above fluid covering condition. For MRI data, five patches are taken per timeframe where each one has to be covered by 5% of fluid domain for training and validation.

Data augmentation is carried out on CFD patches. We apply 5 rotations on the patches to increase the data. The five rotations are those that keep the patch at a size of $32 \times 32 \times 16$.

The testing is made on the full image of *aorta03*. The entire image is cut into $32 \times 32 \times 16$ patches with a stride of $[28, 28, 12]$, respectively, for each axis, to avoid border effects. As mentioned before, this image is also used in the validation set. It creates an overlap between sets, but since only one patch per time frame is present in the validation set, the overlap is very small (about 1%).

2.2 Weighted Mean Frequencies

The proposed super-resolution approach deals only with 3D volume. In order to provide temporal and quantitative information for each voxel in the volume, we developed a new feature called Weighted Mean Frequencies (WMF). The advantage of that feature has been assessed in a former study [9]. WMF provides an image of the most contributing frequency at the voxel scale. Thus, WMF is particularly efficient to assist neural networks in identifying every voxel with is a pulsatile velocity.

More precisely, WMF is the mean of frequencies weighted by Fourier energies. Firstly, we compute the Fourier transform energies applied on u defined as

$$E(u) = |\text{FT}(u)|^2, \quad (1)$$

where u is a phase component and FT is the Fast Fourier Transform. Then, we calculate the mean of the frequencies weighted by the energies of the Fourier

transform as

$$\text{WMF}(u) = \frac{\sum_{i=1}^n E_i(u) f_i}{\sum_{i=1}^n E_i(u)}, \quad (2)$$

where $\text{WMF}(u)$ is the weighted mean result of the u component, f_i is the i -th frequency and n is the number of strictly positive frequencies. This image is computed for each velocity component. Lowest frequencies reveal the pulsatile flow areas. Thus, the three WMF are merged into one feature WMF_{\min} by taking the minimum frequency at the voxel scale such as:

$$\text{WMF}_{\min} = \min(\text{WMF}_u, \text{WMF}_v, \text{WMF}_w) \quad (3)$$

Consequently, WMF_{\min} provides a representation of the overall temporal dynamics of the image. In the presented framework, WMF_{\min} is computed for every CFD and 4D Flow MRI images.

Figure 1 shows the WMF images applied to 4D Flow MRI data. Darker areas represent pixels with the lower WMF value, *ie.* pixel where the most contributing frequency is the lower. In fact, this area is representative of the hull of all pulsatile velocity voxels. For the sake of clarity, we will refer to WMF_{\min} as WMF in the rest of the paper.

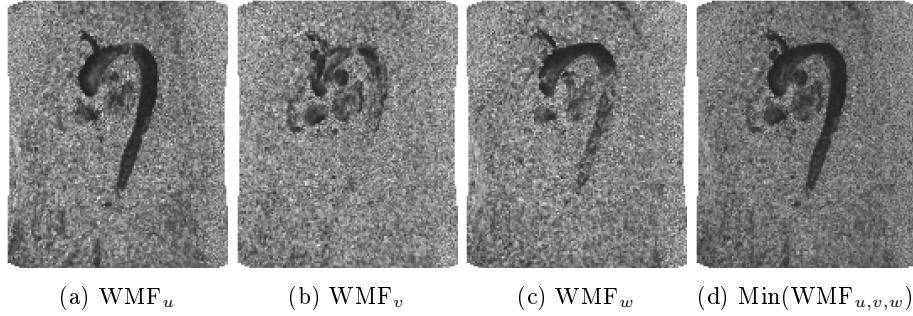


Fig. 1: Images (a), (b), (c) represent the result of WMF for each phase component and (d) is the image obtained using WMF_{\min} . Darker areas represent pixels with the lower WMF value.

2.3 Multi-task Network SURFR-Net

The SURFR-Net (*Segmentation, sUp-Resolution, Flow Reconstruction NETWORK*) is using as input the three velocity components (u, v, w) and the WMF image. Inputs are passed to two consecutive 3D convolutions with kernel $3 \times 3 \times 3$ to extract features (each convolution is followed by a ReLU layer). Features are then passed through a RCAN backbone [13] to extract deep and high-level features. RCAN uses a long skip connection to maintain a link between low- and

high-level features. The number of residual blocks and residual groups for RCAN backbone is set at 3 for both.

For upsampling module, Pixel Shuffle [11] is used as RCAN does in the original architecture. After upsampling module, a convolution $3 \times 3 \times 3$ and a ReLU activation function is used to reduce checkerboard artefacts. Four output branches are set, with the first three branches dedicated to the reconstruction of the three velocity components u , v and w . Each of them is composed of 2 convolutions $3 \times 3 \times 3$ with a ReLU activation layer between them. The last branch is dedicated to the segmentation task. It is composed of a convolution of kernel size $1 \times 1 \times 1$ that merges features to create the segmentation map, followed by a sigmoid layer to hold values between 0 and 1.

2.4 Loss function

In this study, a composed loss function was employed to train the segmentation and super-resolution tasks, given the objective of identifying two distinct tasks. The \mathcal{L}_1 loss was employed for the super-resolution task and the segmentation task was trained using a Binary Cross Entropy loss. The total loss is defined as

$$\text{Loss} = \mathbb{1}_{\text{CFD}} \mathcal{L}_1 + \alpha \mathcal{L}_{\text{seg}}, \quad (4)$$

where α weights the loss dedicated to the segmentation. Super-resolution loss has been exclusively used for CFD data (using the indicator function $\mathbb{1}_{\text{CFD}}$), and its implementation in the context of 4D has been omitted, primarily due to the absence of a ground truth.

2.5 Metrics

To evaluate super-resolution results, we choose four metrics from the state-of-the-art [4, 12]: the relative error \mathcal{E}_{rel} , the peak velocity-to-noise ratio (PVNR), the normalized root mean squared error of speed $\text{nRMS}_{\text{speed}}$ and the direction error \mathcal{E}_{dir} . Relative error \mathcal{E}_{rel} and PVNR are metrics that measure the reconstruction precision. The first one is particularly sensitive to errors on low velocity voxels. Besides, PVNR provides a global evaluation of the quantification error of all velocity vectors in magnitude and direction. The $\text{nRMS}_{\text{speed}}$ metric calculates the quantification error in terms of velocity norm only; in contrast, \mathcal{E}_{dir} quantifies the vector direction error. Segmentation is evaluated with the IoU metric.

3 Results and Discussion

3.1 Implementation

The network was implemented using Python (v3.12.6) and PyTorch (v2.4.1) as deep learning library. A learning rate of 0.0001, Adam optimizer and a batch size of 32 were set for training hyper-parameters. A Nvidia A100 80 GB and

CUDA (v12.1) were used for training. For CFD data downsampling, we used the Python script of Ferdian *et al.* (2020) [4].

As comparison, SRFlow [12] has been implemented and trained only with CFD data, a patch size of $16 \times 16 \times 16$ and a data augmentation using the nine possible rotations. SRFlow network is only dedicated to super-resolution task. For the sake of fidelity, the original losses were employed: \mathcal{L}_1 for velocity mismatch and $mp\mathcal{L}_1$ that penalize directional error. This approach was selected due to the presence of a close network architecture with a channel attention mechanism and long skip connection.

3.2 Quantitative performance

Table 1 shows the results obtained for different networks and loss configurations at the peak systolic frame (10th), the frame with the highest blood speed during the systolic phase, and at a low diastolic frame (32nd), a frame arbitrary chosen during the diastolic phase. Performance indicators are evaluated on the *aorta3* geometry which is used as a test set. In order to set the segmentation loss weight, a grid search has been conducted and led to an optimal α of 0.02.

Considering the loss function (eq. 4), the multi-task architecture, and the available ground-truth, the velocity super-resolution criterion relies only on the CFD simulation and the segmentation loss has no direct impact on velocity quantification. Consequently, the optimal performance with $\alpha = 0.02$ reveal the benefit of the segmentation task to the velocity super-resolution task. Thus, the velocity super-resolution is guided by the segmentation that explicit the fluid and non-fluid areas.

The proposed solution SURFR-Net outperforms the state-of-the-art deep learning approach SRFlow [4]. SURFR-Net is 0.6 and 4.13% better than SRFlow in relative error, respectively, in systolic and diastolic phases. It demonstrates the ability of SURFR-Net to properly quantify low velocity field. Moreover, this improvement is confirmed by the quantification metrics PVNR and $nRMS_{speed}$ with a gain of up to 2.57 dB and 1% in the diastolic phase. The directional error metric \mathcal{E}_{dir} shows that SURFR-Net is particularly more efficient to estimate low velocity with complex directional patterns.

Then, the super-resolution performance has been evaluated with regard to the handcraft feature WMF. That feature is essential to outperform the state-of-the-art solution SRFlow. Without WMF, the network relies only on velocity to identify fluid regions. Thus, the network can be misled by low and steady velocity regions which might be either a fluid recirculation or a non-fluid area with high anatomic signal (with a lower velocity standard deviation). This point is discussed particularly in the next section.

Based on the IoU performances in Table 1, one can observe the high segmentation performance with a minimum IoU of 0.959. However, this evaluation is focused on the synthetic 4D Flow MRI images created from the CFD simulation. SURFR-Net performs particularly well in that context because the velocity in the non-fluid areas is only characterize by noise. Consequently, one can assume

Table 1: Results of different approaches for systolic peak and a diastolic frame on CFD test set. **Best** and second best results are bold and underlined respectively. *SRFlow is trained without 4D Flow MRI images, no segmentation task and a patch size of $16 \times 16 \times 16$.

Network	Peak Systolic				
	$\mathcal{E}_{\text{rel}} \downarrow$	PVNR \uparrow	nRMS _{speed} \downarrow	$\mathcal{E}_{\text{dir}}(10^{-3}) \downarrow$	IoU \uparrow
Our ($\alpha = 0.05$)	8.82 ± 16.48	27.39	0.0360	5.5 ± 58.0	0.936
Our ($\alpha = 0.03$)	8.13 ± 15.51	27.61	0.0350	5.2 ± 58.0	0.944
Our ($\alpha = 0.02$)	<u>6.71 ± 12.73</u>	<u>29.10</u>	<u>0.0281</u>	<u>4.4 ± 51.1</u>	<u>0.959</u>
Our ($\alpha = 0.01$)	8.70 ± 16.51	27.42	0.0359	5.1 ± 57.0	0.930
Our ($\alpha = 0.005$)	8.60 ± 15.78	<u>27.73</u>	<u>0.0340</u>	5.0 ± 54.3	0.942
Our (<i>w/o</i> WMF, $\alpha = 0.05$)	8.63 ± 16.20	26.86	0.0390	5.5 ± 57.6	0.935
Our (<i>w/o</i> WMF, $\alpha = 0.02$)	9.72 ± 17.79	25.47	0.0472	6.3 ± 63.4	0.917
Our (<i>w/o</i> WMF, $\alpha = 0.005$)	9.23 ± 16.95	25.70	0.0464	5.9 ± 61.6	0.931
SRFlow*	<u>7.31 ± 14.25</u>	27.34	0.0367	5.2 ± 58.1	/

Network	Low Diastolic				
	$\mathcal{E}_{\text{rel}} \downarrow$	PVNR \uparrow	nRMS _{speed} \downarrow	$\mathcal{E}_{\text{dir}}(10^{-3}) \downarrow$	IoU \uparrow
Our ($\alpha = 0.05$)	18.60 ± 19.61	<u>26.54</u>	<u>0.0388</u>	23.5 ± 121.4	<u>0.947</u>
Our ($\alpha = 0.03$)	18.63 ± 19.91	25.98	0.0429	<u>22.2 ± 115.9</u>	0.944
Our ($\alpha = 0.02$)	<u>15.46 ± 16.85</u>	<u>28.46</u>	<u>0.0314</u>	<u>16.8 ± 100.9</u>	<u>0.965</u>
Our ($\alpha = 0.01$)	18.50 ± 19.99	26.28	0.0396	24.9 ± 128.4	0.941
Our ($\alpha = 0.005$)	<u>18.33 ± 19.84</u>	26.32	<u>0.0388</u>	24.8 ± 128.5	0.946
Our (<i>w/o</i> WMF, $\alpha = 0.05$)	20.18 ± 20.99	25.63	0.0435	27.8 ± 132.7	0.921
Our (<i>w/o</i> WMF, $\alpha = 0.02$)	21.47 ± 22.48	23.99	0.0530	33.0 ± 153.0	0.932
Our (<i>w/o</i> WMF, $\alpha = 0.005$)	20.13 ± 21.13	25.22	0.0463	28.0 ± 137.9	0.945
SRFlow*	19.59 ± 20.31	25.89	0.0418	25.5 ± 123.5	/

lower performance on 4D Flow MRI images. Interpolated mask from low resolution 4D Flow MRI images, were used in the training, instead of ground truth super-resolved mask, to fuel the network with more complex velocity field. Thus, it has been considered as too imprecise to present IoU performance over 4D Flow MRI images.

3.3 Qualitative evaluation of SR segmentation

Figure 2 shows segmentations obtained by using our approach on a real 4D Flow MRI. Due to the unavailability of high-resolution MRI and the necessary metrics for evaluating the results, a qualitative evaluation will be conducted. Segmentations are obtained by training the SURFR-Net with the hyperparameter α set to 0.02 and, the WMF is used as input in Fig. 2b or not as in Fig. 2c.

The segmentation of the vessels (fluid domain) obtained using WMF as input is visually close to the vessels, particularly the aorta, as seen in the anatomical image. In Fig. 2b, the network appears to encounter greater difficulty in isolating

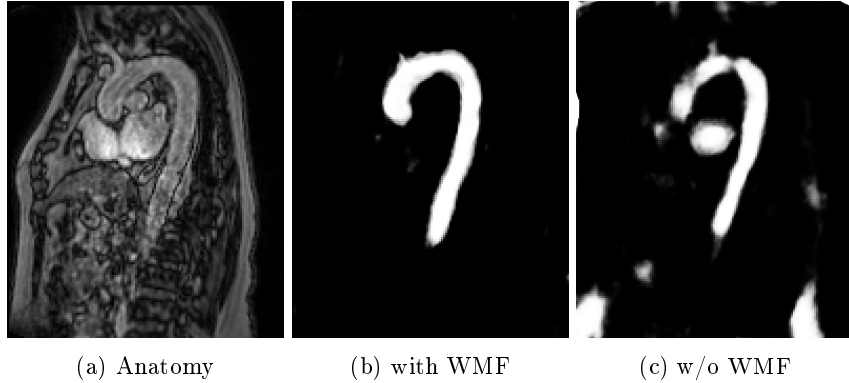


Fig. 2: Anatomical image and segmentations produced by the network with $\alpha = 0.02$ with and without WMF as input. Segmentations are made for phase images at the beginning of the diastolic phase.

the brachiocephalic trunk on the upper left side of the aortic arch, the pulmonary trunk and the heart ventricles. This phenomenon may be attributed to the utilisation of phase images captured during the diastolic phase, which inherently exhibit lower velocities. Moreover, in this multi-task framework, SURFR-Net relies primarily on the velocity to determine the voxel containing fluid. Concurrently, WMF input assists in the exclusion of voxels where flow pulsatility is negligible (see the bottom of Fig. 2c). As observed in Fig. 2b, the absence of the heart in the segmentation using WMF can be attributed to the absence of ventricle segmentation in the training. As illustrated in Fig. 2c, in the absence of WMF as an input, the network experiences significant challenges in delineating fluid and non-fluid zones. The network successfully identifies the region of the heart ventricles. However, it incorrectly identifies other areas where blood flow is absent.

As evidenced by this observation, the WMF contributes significantly to the segmentation of fluid zones in 4D Flow MRI. It is important to provide information on global fluid dynamics in MRI scans to ensure that the network does not mistake noise for highly turbulent or low-speed fluid zones.

4 Conclusion

In the present study, a novel approach for 4D Flow MRI is proposed, which has the capacity to facilitate super-resolution and high-resolution segmentation. The capacity of the network is utilised to denoise the non-fluid domain, thereby facilitating the extraction of features. These features are then employed to guide a segmentation process, which is further constrained by a super-resolution and a segmentation loss. The Weighted Mean Frequencies (WMF) technique is a manually crafted feature that delineates the pulsatility of the fluid domain in

the volume. This facilitates network segmentation of the 4D Flow MRI data set exclusively through phase images.

Moreover, the proposed approach is based on a relatively early attention architecture. A more recent backbone could enhance the latent space representation and the resulting performance. In that study, the dataset size is limited and a larger one could facilitate more complex contrast contexts. Further investigation should be conducted to evaluate the impact on biomarker quantification.

Acknowledgments. This research used resources of the GLiCID Computing Facility (Ligerien Group for Intensive Distributed Computing, <https://doi.org/10.60487/glicid>, Pays de la Loire, France).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

- Berhane, H., Scott, M., Elbaz, M., Jarvis, K., McCarthy, P., Carr, J., Malaisrie, C., Avery, R., Barker, A.J., Robinson, J.D., Rigsby, C.K., Markl, M.: Fully automated 3D aortic segmentation of 4D flow MRI for hemodynamic analysis using deep learning. *Magn. Reson. Medicine* **84**(4), 2204–2218 (2020). <https://doi.org/https://doi.org/10.1002/mrm.28257>
- Bustamante, M., Viola, F., Engvall, J., Carlhäll, C.J., Ebberts, T.: Automatic time-resolved cardiovascular segmentation of 4D flow MRI using deep learning. *J. of Magn. Reson. Imaging* **57**(1), 191–203 (2023). <https://doi.org/10.1002/jmri.28221>
- Fathi, M.F., Perez-Raya, I., Baghaie, A., Berg, P., Janiga, G., Arzani, A., D'Souza, R.: Super-resolution and denoising of 4D-Flow MRI using physics-Informed deep neural nets. *Computer Methods and Programs in Biomedicine* **197**, 105729. <https://doi.org/10.1016/j.cmpb.2020.105729>
- Ferdian, E., Suinesiaputra, A., Dubowitz, D.J., Zhao, D., Wang, A., Cowan, B., Young, A.A.: 4DFlowNet: Super-Resolution 4D Flow MRI Using Deep Learning and Computational Fluid Dynamics. *Frontiers in Physics* **8**, 138. <https://doi.org/10.3389/fphy.2020.00138>
- Ferdian, E., Suinesiaputra, A., Dubowitz, D.J., Zhao, D., Wang, A., Cowan, B., Young, A.A.: 4DFlowNet: Super-resolution 4D flow MRI using deep learning and computational fluid dynamics. *AIP Conf. Proc.* **8**, 138 (2020). <https://doi.org/10.3389/fphy.2020.00138>
- Levilly, S., Castagna, M., Idier, J., Bonnefoy, F., Le Touzé, D., Moussaoui, S., Paul-Gilloteaux, P., Serfaty, J.M.: Towards quantitative evaluation of wall shear stress from 4D flow imaging. *Magnetic Resonance Imaging* **74**, 232–243. <https://doi.org/10.1016/j.mri.2020.08.017>
- Levilly, S., Moussaoui, S., Serfaty, J.M.: Segmentation-free velocity field super-resolution on 4d flow mri. *IEEE Transactions on Image Processing* **33**, 5637–5649 (2024). <https://doi.org/10.1109/TIP.2024.3470553>
- Markl, M., Frydrychowicz, A., Kozerke, S., Hope, M., Wieben, O.: 4D flow MRI. *Journal of Magnetic Resonance Imaging* **36**(5), 1015–1036. <https://doi.org/10.1002/jmri.23632>

9. Perrin, S., Levilly, S., Sun, H., Mouchère, H., Serfaty, J.M.: Weighted Mean Frequencies: a handcraft Fourier feature for 4D Flow MRI segmentation (2025), <https://arxiv.org/abs/2506.20614>
10. Puiseux, T., Sewonu, A., Meyrignac, O., Rousseau, H., Nicoud, F., Mendez, S., Moreno, R.: Reconciling PC-MRI and CFD: An in-vitro study. *NMR Biomed* **32**(5), e4063 (2019). <https://doi.org/10.1002/nbm.4063>
11. Shi, W., Caballero, J., Huszar, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: 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). pp. 1874–1883. IEEE. <https://doi.org/10.1109/CVPR.2016.207>
12. Shit, S., Zimmermann, J., Ezhov, I., Paetzold, J.C., Sanches, A.F., Pirkel, C., Menze, B.H.: SRflow: Deep learning based super-resolution of 4D-flow MRI data. *Frontiers in Artificial Intelligence* **5**, 928181. <https://doi.org/10.3389/frai.2022.928181>
13. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) *Computer Vision – ECCV 2018*, vol. 11211, pp. 294–310. Springer International Publishing. https://doi.org/10.1007/978-3-030-01234-2_18