The intriguing effect of frequency disentangled learning on medical image segmentation

Guanghui Fu, Gabriel Jiménez, Sophie Loizillon, Lydia Chougar, Didier Dormont, Romain Valabrègue, Ninon Burgos, Stéphane Lehéricy, Daniel Racoceanu, Olivier Colliot

The Intriguing Effect of Frequency Disentangled Learning on Medical Image Segmentation

Guanghui Fu[a], Gabriel Jimenez[a], Sophie Loizillon[a], Lydia Chougar[a,b,c], Didier Dormont[a,c], Romain Valabregue[b,d], Ninon Burgos[a], Stéphane Lehéricy[b,c,d], Daniel Racoceanu[a], Olivier Colliot[a], and the ICEBERG Study Group[b]

[a]Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, Paris, France
[b]ICM, Centre de NeuroImagerie de Recherche-CENIR, Paris, France.
[c]AP-HP, Pitié Salpêtrière, DMU DIAMENT, Dep. of Neuroradiology, Paris, France
[d]Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France

ABSTRACT

Deep models have been shown to tend to fit the target function from low to high frequencies (a phenomenon called the frequency principle of deep learning). One may hypothesize that such property can be leveraged for better training of deep learning models, in particular for segmentation tasks where annotated datasets are often small. In this paper, we exploit this property to propose a new training method based on frequency-domain disentanglement. It consists of three main stages. First, it disentangles the image into high- and low-frequency components. Then, the segmentation network model learns them separately (the approach is general and can use any segmentation network as backbone). Finally, feature fusion is performed to complete the downstream task. The method was applied to the segmentation of the red and dentate nuclei in Quantitative Susceptibility Mapping (QSM) data and to three tasks of the Medical Segmentation Decathlon (MSD) challenge under different training sample sizes. For segmenting the red and dentate nuclei and the heart, the proposed approach resulted in considerable improvements over the baseline (respectively between 8 and 16 points of Dice and between 5 and 8 points). On the other hand, there was no improvement for the spleen and the hippocampus. We believe that these intriguing results, which echo theoretical work on the frequency principle of deep learning, are of interest for discussion at the conference. The source code is publicly available at: https://github.com/GuanghuiFU/frequency_disentangled_learning.

Keywords: Disentangle representation, Frequency domain, Segmentation, Deep Learning

1. INTRODUCTION

Deep learning is particularly powerful for medical image segmentation but its performance can be limited by the lack of training data.[1] Xu et al.[2] and Rahamman et al.[3] found that deep networks tend to fit from low to high frequency information during training. This phenomenon was referred to as the frequency principle (F-principle) of deep learning.[2] The unbalanced learning of high- and low-frequency information during training requires a large amount of data to produce reliable results. Tang et al.[4] analyzed from the frequency domain and proved that cascaded convolutional decoder networks are more likely to weaken high-frequency components. From the above references, we can conclude that different frequencies play different roles in the learning process of deep networks. Therefore, to effectively learn information, the learning process must balance between high- and low-frequency components. Furthermore, it has been shown that CNN decoders (which are a part of most deep learning segmentation methods) weaken the impact of high-frequency information during training.[4] One may thus hypothesize that disentangling image information into high- and low-frequency components may lead to improved segmentation results.

Further author information: (Send correspondence to Guanghui Fu, guanghui.fu@inria.fr)

Various recent works have proposed to exploit disentanglement in deep learning. Azad et al.[5] proposed a frequency re-calibration U-Net for medical image segmentation, by introducing the Laplacian pyramid in the U-shaped structure. It allowed to better generalize with few training data. Liu et al.[6] used frequency refinement to improve adversarial defense for several biomedical image segmentation tasks. McIntosh et al.[7] proposed a wavelet transform-based model and showed that extra high-frequency components can increase performance. Charstias et al.[8] proposed to decompose cardiac images into spatial anatomical factors and non-spatial modality factors using a variational autoencoder. Liu et al.[9] proposed a method for optical coherence tomography angiography (OCTA) segmentation based on disentangling images into the anatomy component and the local contrast component from paired OCTA scans. The disentangling module is implemented by a conditional variational autoencoder (CVAE). Furthermore, several works have used frequency decomposition for domain adaptation, generalization, or prior knowledge introduction.[10–12] However, the above approaches may be complex to train and implement. Moreover, they may be specific to a given architecture. Finally, most of them did not specifically assess the impact in the low-training size regime.

In this paper, we propose to perform medical image segmentation using a simple disentanglement into high- and low-frequency parts. The method has two advantages: it is conceptually simple and it can be used with any type of segmentation network.

2. METHODS

The proposed method consists of two steps: i) frequency domain disentangling and feature learning; ii) frequency domain fusion. Two types of fusion are considered. In early fusion, the fusion is done before feeding the result to a segmentation network. In late fusion, only the high frequency information is fed to the segmentation network and the result is fused with low frequency features. The overall workflow of the approach can be seen in Figure 1.

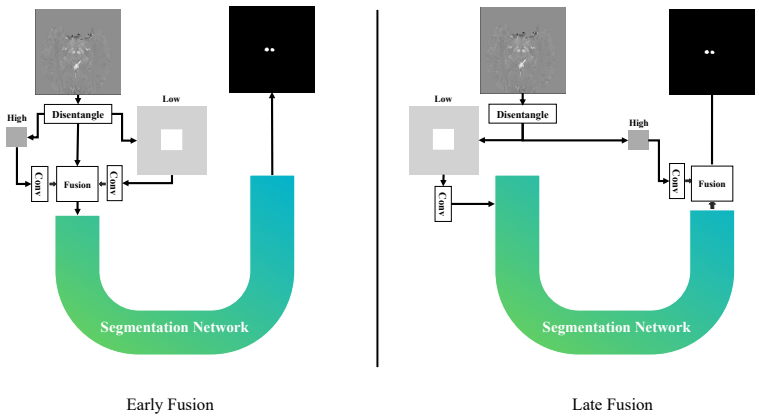

Figure 1. Processing flow of the proposed method. We introduce two ways of fusion: early fusion and late fusion.

2.1 Frequency domain disentangling and feature learning

Given samples $i \in I$, where $I \subset \mathbb{R}^{N_x \times N_y \times N_z}$ is a set of images, the disentangling operation is achieved by first transferring to Fourier space, and separating the high- and low-frequency components as follows:

$$\begin{aligned} \mathcal{H}^{\theta}(i) &= \mathcal{F}(i)\left[\frac{N_x \times (1-\theta)}{2} : \frac{N_x \times (1+\theta)}{2}, \frac{N_y \times (1-\theta)}{2} : \frac{N_y \times (1+\theta)}{2}, :\right] \\ \mathcal{L}^{\theta}(i) &= \mathcal{F}(i) - \mathcal{H}(i) \end{aligned} \tag{1}$$

where $\mathcal{F}(i)$ represents the Fourier transform of i, $\mathcal{L}^{\theta}(i)$ is the extraction of the low-frequency part of i, $\mathcal{H}^{\theta}(i)$ is the high-frequency part and $\theta \in (0, 1)$ is a parameter that controls the high/low frequency separation. We then

apply the inverse Fourier transform ($\mathcal{F}^{-1}$) to obtain high- and low-frequency parts in image space:

$$\begin{aligned} L^{\theta}(i) &= \mathcal{F}^{-1}(\mathcal{L}^{\theta}(i)) \\ H^{\theta}(i) &= \mathcal{F}^{-1}(\mathcal{H}^{\theta}(i)) \end{aligned} \tag{2}$$

$L^{\theta}(i)$ and $H^{\theta}(i)$ are then each fed to a convolutional layer and the outputs are respectively denoted as O_L^{θ} and O_H^{θ}. Figure 2 illustrates the transformation of the QSM image into frequency domain space, aimed at distinguishing between high and low frequencies. In this representation, the central region corresponds to low frequencies, while the borders indicate high frequencies. Notably, the low frequency segment is defined as encompassing a 10% width of the image.

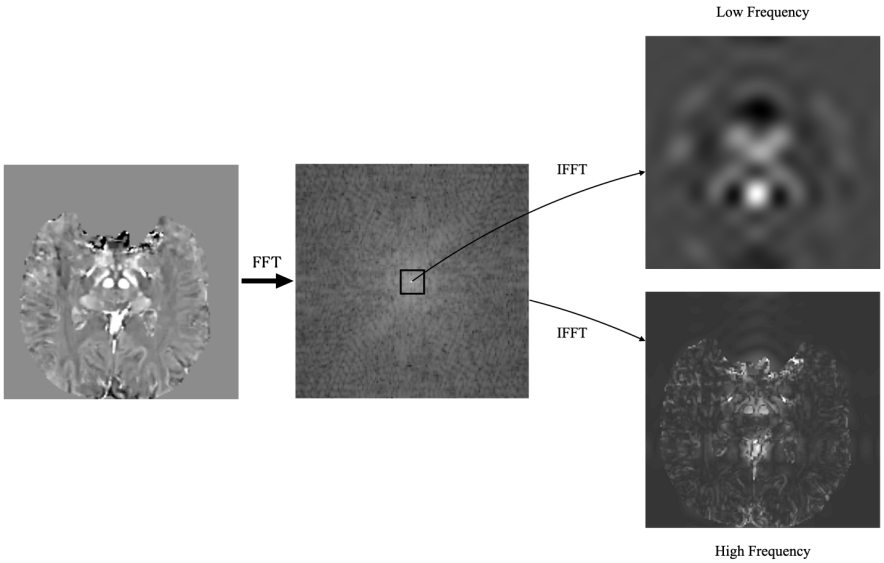

Figure 2. An example of QSM image frequency disentanglement. This figure depicts the process involving FFT (Fast Fourier Transform) and IFFT (Inverse Fast Fourier Transform). It highlights how high frequencies correlate with brain structures within the image, whereas low frequencies are associated with image contrast.

The F-principle indicates that models tend to first fit low-frequency information. Thus, in particular with low sample size, there is a risk that high-frequency information is not adequately learnt. High-frequency information represents structural details that are essential in medical tasks. Our approach addresses this issue by utilizing a simple disentanglement operation that forces the model to balance the learning process for both high and low-frequency information.

2.2 Feature fusion after disentangled learning

The fusion operation can be done at two different stages: early fusion and late fusion. In the case of early fusion, low- and high-frequency outputs O_L^{θ} and O_H^{θ} are fused before being fed to a segmentation network. In the case of late fusion, only the high frequency is fed to a segmentation network, resulting in a result denoted as S_H^{θ} which is fused with O_L^{θ}.

3. EXPERIMENTS AND RESULTS

3.1 Implementation details

Our code is developed based on the PyTorch framework.[13] We used the open-source Python library TorchIO[14] for reshaping images to the same size (for a given task) and for min-max normalization. We used Adam[15] as optimizer with a learning rate of 1e-3. We did not apply any hyperparameter selection techniques or data augmentation. In our experiments, we used a 3D-UNet[16] as segmentation network and Dice as loss function[17] but the approach is general and could be applied to other models.

3.2 Datasets and experiments

We conducted experiments on four segmentation tasks. The first task is the segmentation of the red nucleus and dentate nuclei from MRI quantitative susceptibility mapping (QSM) data. The three other tasks are the segmentation of the spleen, of the heart and of the hippocampus from the publicly available Medical Segmentation Decathlon (MSD).[18] For red nucleus segmentation, we studied a total of 80 participants including 18 healthy subjects, 46 patients with early Parkinson's disease (i.e. disease duration below 4 years), and 16 patients with prodromal parkinsonism (idiopathic rapid eye movement sleep behavior disorder-iRBD), recruited between May 2015 and January 2019 as part of the ICEBERG cohort. In some participants, the boundaries of the dentate nucleus were heavily affected by artifacts, making them impossible to distinguish. Such participants were excluded from the dentate nucleus segmentation task which included 67 participants including 17 healthy subjects, 39 patients with early Parkinson's disease, and 11 patients with prodromal parkinsonism (training, validation and test sets comprised 42, 11, and 14 participants, respectively). The QSM were generated from multi-echo 3D GRE (12 echo times ranging from 4 ms to 37 ms) with a full brain coverage at an isotropic voxel resolution of 1 mm^3.

Each dataset was split into training, validation and test sets. The splits were done at the participant level to avoid any data leakage.[19] We studied the performance when varying the size of the training set, ranging from very small size (4 samples) to full training set (166 participants for the hippocampus task), while the validation and test sets were left unchanged. In order to avoid being biased by a lucky (or unlucky) subsampling, for a given training set size, we randomly drawn 10 training subsamples, trained separately on each of the 10 subsamples and averaged the results. The datasets and the splits are summarized in Table 1. The performance metrics were the Dice coefficient and the 95% Hausdorff distance. We assessed whether the average Dice was significantly higher with the frequency disentanglement than without using paired Student's t t-tests on the test set (with Bonferroni correction across the four independent tasks). Figure 3 shows the data and the region of interest for each segmentation task. Furthermore, Figure 4 displays the separated visualizations of the high and low frequency regions within our experimental data.

Table 1. Characteristics of the 3D medical imaging datasets.

Data type	Dataset	Region(s)	Train+val	Test	Image Size
3D MRI	[Local] ICEBERG	Red nucleus	51+13	16	160,160,128
3D MRI	[Local] ICEBERG	Dentate nucleus	42+11	14	160,160,128
3D CT	[Public] Spleen	Full organ	25+7	9	256,256,128
3D MRI	[Public] Heart	Full organ	12+4	4	320,320,128
3D MRI	[Public] Hippocampus	Anterior, Posterior	166+42	52	56, 56, 40

3.3 Results

The results are displayed in Table 2 and Table 3. Frequency disentanglement (either early or late stage) resulted in substantial and statistically significant increases in performance for the red nucleus and dentate nucleus (between 8 and 16 points of Dice) and the heart (between 5 and 8 points). However, no improvement was observed for the spleen and the hippocampus. As can be expected, across all tasks, performances increased with the training set size. For the red and dentate nuclei, the strongest improvements were observed when training with very few samples (e.g. 16 points when training with four samples). There was no major difference between early and late fusion.

4. DISCUSSION

Our method is grounded on the frequency principle of deep learning. Conventional training process can result in asynchronous learning in the frequency domain. Our solution is to disentangle the data without the need to modify the model architecture or employ other training techniques. This method has the benefit of being simple to implement and applicable with any segmentation network. We propose two fusion strategies: early and late. In our experiments, we did not observe any major or systematic difference in performance between the two. However, the early fusion strategy offers more flexibility and is not restricted to an encoder-decoder architecture.

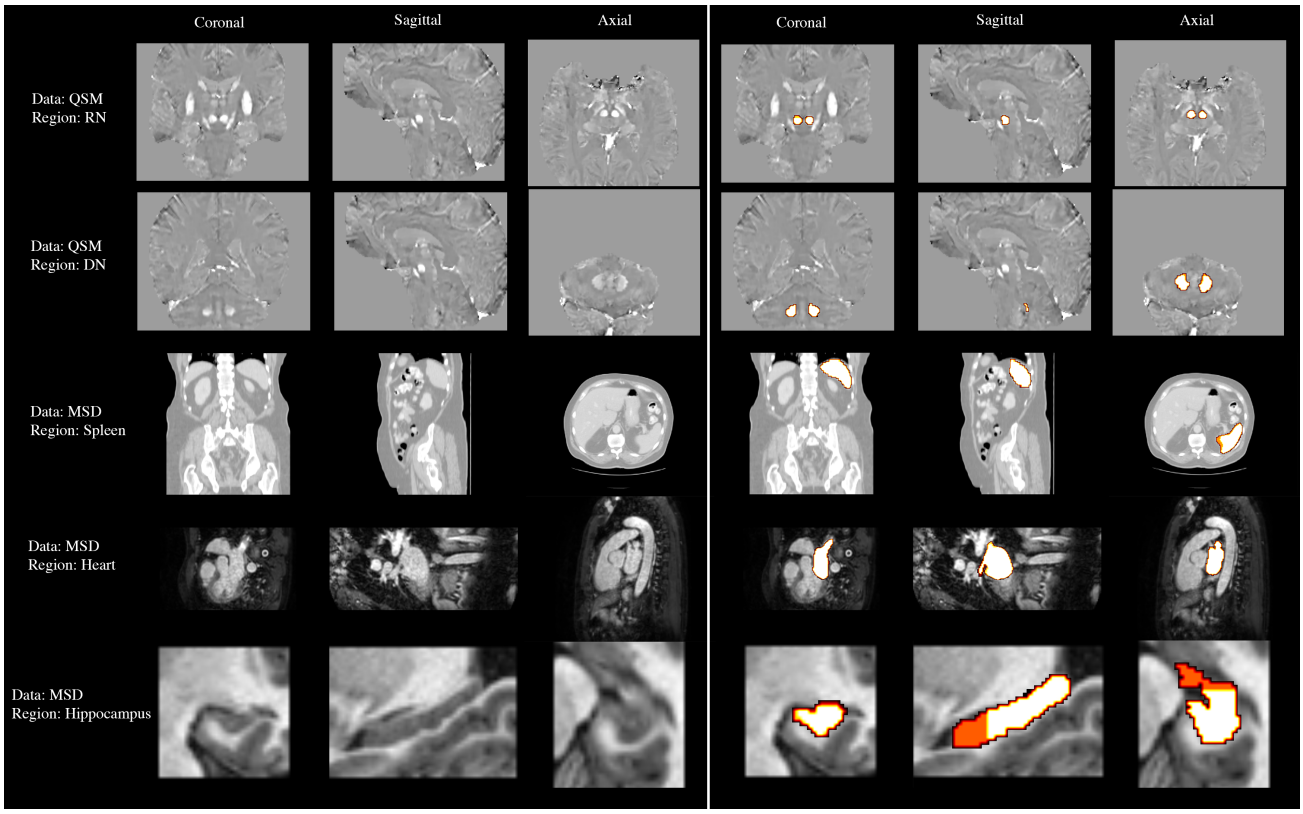

Figure 3. These are examples of all the datasets we used for model evaluation. We overlap the region of interest (right part) and visualize them in three planes.

We performed a rigorous evaluation using an independent test set separated from the very beginning, reporting unbiased SEMs and statistical tests computed on the test set and using multiple resamplings of the training set to obtain robust estimates. The use of frequency disentanglement (FD) led to considerable improvement in performances for the red nucleus and the heart segmentation. We believe that it is quite remarkable that such a simple strategy results in such gains of performance. However, this was not the case for the spleen and hippocampus tasks. It remains unclear why FD is beneficial in some cases but not in others. One can only speculate that this is due to some specific characteristics in the shape or appearance of the target objects. Further experiments will be needed to explain this phenomenon.

In this preliminary work, we only tested our approach with a simple backbone model, the U-Net. Future work should assess whether FD is also beneficial to more advanced segmentation architectures. Another limitation of our approach is that one needs to choose the parameter θ which controls the separation between high and low frequencies. We did not experiment with varying values of θ. It is possible that other values would have been more adapted for some tasks. Future work could aim to integrate the parameter into the loss function as in reference.[20]

In summary, we presented a novel, yet simple, segmentation approach based on disentangling of frequency components. When applied to the red and dentate nuclei and the heart, it provided considerable improvements in performance. We believe that these intriguing results are of interest for the community.

5. ACKNOWLEDGMENTS

The research leading to these results has received funding from the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute) and reference ANR-10-IAIHU-06 (Agence Nationale de la Recherche-10-IA Institut Hospitalo-Universitaire-6). The ICEBERG study is supported by the European Research Council

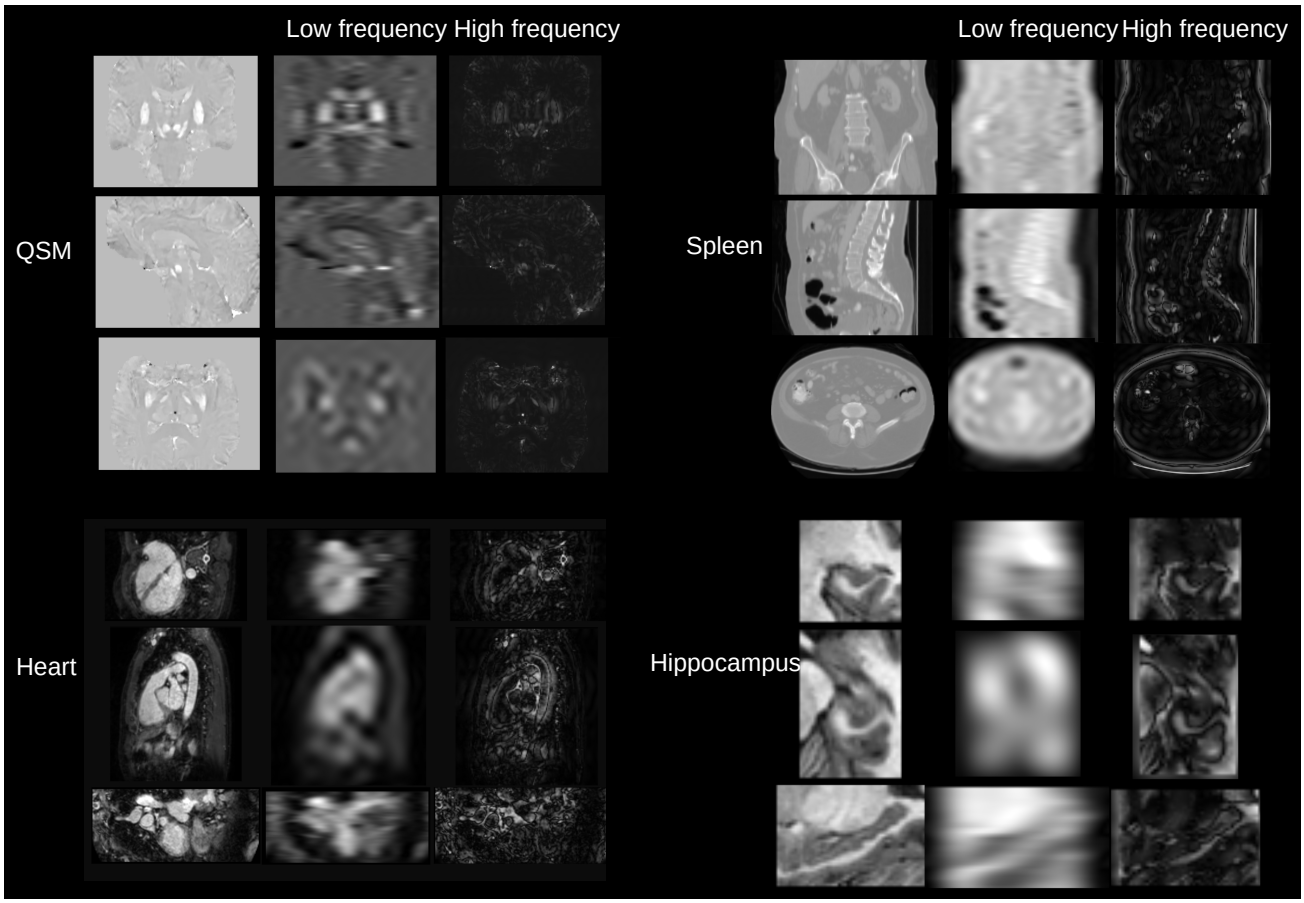

Figure 4. Displayed here are the outcomes following the disentangling of high and low frequencies for all datasets employed in our model evaluation.

(ERC) under grant agreement No. 678304, the European Union's Horizon 2020 research and innovation program under grant agreement No. 826421 (TVB-Cloud), Agence Nationale de la Recherche (ANR) under grant agreements ANR-10-IAIHU-06 (IHU ICM), ANR-11-INBS-0006, and ANR-19-JPW2-000 (JPND E-DADS), association France Parkinson (PRECISE-PD project), the Fondation d'Entreprise EDF, Biogen Inc., Fondation Thérèse and René Planiol, Fondation Saint Michel. It received unrestricted support for Research on Parkinson's disease from Energipole (M. Mallart), M. Villain and the Société Française de Médecine Esthétique (M. Legrand). Guanghui Fu is supported by the Chinese Government Scholarship provided by China Scholarship Council (CSC). Lydia Chougar is supported by a Poste d'accueil Inria/AP-HP.

REFERENCES

[1] Tajbakhsh, N., Jeyaseelan, L., and et al., "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation," *Medical Image Analysis* **63**, 101693 (2020).

[2] Xu, Z.-Q. J., Zhang, Y., and et al., "Frequency principle: Fourier analysis sheds light on deep neural networks," *Communications in Computational Physics* **28**(5), 1746–1767 (2020).

[3] Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., Bengio, Y., and Courville, A., "On the spectral bias of neural networks," in [*Proc. ICML 2019*], 5301–5310, PMLR (2019).

[4] Tang, L., Shen, W., Zhou, Z., Chen, Y., and Zhang, Q., "Defects of convolutional decoder networks in frequency representation," *arXiv preprint arXiv:2210.09020* (2022).

[5] Azad, R., Bozorgpour, A., Asadi-Aghbolaghi, M., Merhof, D., and Escalera, S., "Deep frequency recalibration U-Net for medical image segmentation," in [*Proc. ICCV 2021*], 3274–3283 (2021).

Table 2. The performance of red nucleuse and dentate nucleus segmentation task on private dataset (ICEBERG data). Results are shown as mean±SEM (standard error of the mean). FD indicates whether frequency disentangled learning was performed. Results are marked in bold when the difference in Dice between the best FD method and the baseline (None) was statistically significant. n represents the size of the subsample used for training.

Task	n (percentage)	FD	Dice	95 Hausdorff
Red Nucleus	4 (7.5%)	None	60.17±2.34	44.84±5.62
		Early	76.65±2.59	23.05±6.3
		Late	**76.75±2.42**	27.53±7.0
	8 (15%)	None	74.73±1.95	23.99±5.01
		Early	**84.39±1.58**	10.58±4.88
		Late	83.84±1.59	12.22±4.94
	16 (30%)	None	81.47±1.35	10.61±5.04
		Early	86.63±1.25	7.37±4.23
		Late	**87.1±1.17**	6.6±3.45
	51 (100%)	None	87.96±0.63	4.91±3.79
		Early	95.67±1.09	0.38±0.12
		Late	**95.83±0.88**	0.38±0.12
Dentate Nucleus	4 (7.5%)	None	51.09±3.17	59.4±6.31
		Early	**79.59±1.61**	**2.16±0.4**
		Late	78.82± 1.53	8.3±2.37
	7 (15%)	None	69.23±2.97	37.65±8.67
		Early	81.44±1.46	7.77±3.1
		Late	**81.89±1.48**	**2.93±1.06**
	13 (30%)	None	75.63±2.34	11.23±5.93
		Early	82.65±1.7	5.22±2.9
		Late	**84.00±1.33**	**1.44±0.24**
	42 (100%)	None	64.56±2.83	49.7±8.6
		Early	83.61±2.01	3.41±1.69
		Late	**85.73±1.87**	**2.98±1.46**

[6] Liu, Q., Jiang, H., and et al., "Defending deep learning-based biomedical image segmentation from adversarial attacks: a low-cost frequency refinement approach," in [*Proc. MICCAI 2020*], 342–351, Springer (2020).

[7] McIntosh, D., Marques, T. P., and Albu, A. B., "Preservation of high frequency content for deep learning-based medical image classification," in [*Proc. CRV 2021*], 41–48, IEEE (2021).

[8] Chartsias, A., Joyce, T., and et al., "Disentangled representation learning in cardiac image analysis," *Medical image analysis* **58**, 101535 (2019).

[9] Liu, Y., Carass, A., and et al., "Disentangled representation learning for OCTA vessel segmentation with limited training data," *IEEE transactions on medical imaging* **41**(12), 3686–3698 (2022).

[10] Yang, Y. and Soatto, S., "FDA: Fourier domain adaptation for semantic segmentation," in [*Proc. CVPR 2020*], 4085–4095 (2020).

[11] Huang, J., Guan, D., and et al., "FSDR: Frequency space domain randomization for domain generalization," in [*Proc. CVPR 2021*], 6891–6902 (2021).

[12] Fu, G., Jimenez, G., and et al., "Fourier disentangled multimodal prior knowledge fusion for red nucleus segmentation in brain MRI," *arXiv preprint arXiv:2211.01353* (2022).

[13] Paszke, A., Gross, S., and et al., "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems* **32** (2019).

[14] Pérez-García, F. and et al., "TorchIO: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning," **208**, 106236 (2021).

[15] Kingma, D. P. and Ba, J., "Adam: A method for stochastic optimization," *arXiv preprint* **1412.6980** (2014).

[16] Çiçek, Ö. and et al., "3D U-Net: learning dense volumetric segmentation from sparse annotation," in [*Proc. MICCAI 2016*], 424–432, Springer (2016).

[17] Milletari, F., Navab, N., and Ahmadi, S.-A., "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in [*Proc. 3DV 2016*], 565–571, IEEE (2016).

Table 3. Segmentation performance for the three public tasks from medical segmentation decalthon (spleen, heart, hippocampus). Results are shown as mean±SEM (standard error of the mean). FD indicates whether frequency disentangled learning was performed. Results are marked in bold when the difference in Dice between the best FD method and the baseline (None) was statistically significant. n represents the size of the subsample used for training.

Task	n (percentage)	FD	Dice	95 Hausdorff
Spleen	4 (15%)	None	57.39±6.27	205.48±32.65
		Early	54.82±6.63	143.26±26.95
		Late	55.75±6.3	154.96±25.39
	8 (30%)	None	69.13±6.25	165.88±35.85
		Early	70.46±5.23	88.5±27.07
		Late	66.91±5.89	97.29±27.33
	25 (100%)	None	86.01±2.74	50.39±24.93
		Early	90.8±1.25	7.7±3.05
		Late	88.25±2.34	30.76±16.98
Heart	4 (30%)	None	78.89±2.74	24.95±6.09
		Early	**84.01±1.83**	13.08±2.91
		Late	83.94±1.8	13.51±3.39
	12 (100%)	None	82.74±1.19	14.13±2.52
		Early	**90.55±0.74**	3.76±0.41
		Late	90.26±0.89	4.78±1.34
Hippocampus	4 (2.5%)	None	76.21±0.75	4.33±0.46
		Early	75.46±0.74	3.24±0.33
		Late	75.35±0.71	3.55±0.32
	17 (10%)	None	81.02±0.6	2.38±0.26
		Early	81.13±0.56	2.11±0.2
		Late	81.37±0.56	2.02±0.19
	83 (50%)	None	84.32±0.47	1.49±0.1
		Early	84.17±0.47	1.5±0.1
		Late	84.29±0.48	1.42±0.07
	166 (100%)	None	85.45±0.4	1.32±0.05
		Early	84.91±0.52	1.32±0.05
		Late	85.11±0.46	1.31±0.04

[18] Antonelli, M., Reinke, A., and et al., "The medical segmentation decathlon," *Nature communications* **13**(1), 4128 (2022).

[19] Thibeau-Sutre, E., Diaz, M., and et al., "ClinicaDL: an open-source deep learning software for reproducible neuroimaging processing," *Computer Methods and Programs in Biomedicine* **220**, 106818 (2022).

[20] Cai, M., Zhang, H., and et al., "Frequency domain image translation: More photo-realistic, better identity-preserving," in [*Proc. ICCV 2021*], 13930–13940 (2021).