Exploring connections of spectral analysis and transfer learning in medical imaging – Supplemental material

1 Implementation Details

Pre-training The pre-training was carried out on ImageNet and RadImageNet separately. All the pre-training used SGD optimizer, with the initial and final learning rate being 10^{-1} and 10^{-3}, and the learning rate decay of 0.1 for every 30 and 15 epochs. The pre-training took approximately 90 epochs on ImageNet and 45 epochs on RadImageNet to reach full convergence. Random cropping, resizing, and flipping were applied to augment training samples. The top-5 validation accuracy is 92.7% on ImageNet and 94.5% on RadImageNet.

Fine-tuning The fine-tuning used Adam optimizer, with the initial and the minimum learning rate being 10^{-4} and 10^{-5}, and the learning rate decay of 0.1 if no further improvement for 50 epochs. Random cropping, flipping, affine transformation (degrees = 30, shearing = 0.2, scaling = (0.8, 1.2), and color jittering (brightness = 0.5, contrast = 0.5) were applied to augment training samples. An experiment on the target dataset with random weight initialization was also carried out for comparison, where the learning rate was 10× higher than that of fine-tuning.

Artifact synthesis For CT, the raw data collected by a scanner is presented as projections, in which the dominant noise is the element-wise independent Photon noise that follows the Poisson distribution. However, the CT image reconstruction pipeline breaks this independence and introduces structure-related patterns called streak artifacts. To simulate these operations, a virtual scanner with a parallel ray source and 1D flat detector array was created to convert the CT slices to projections. Specifically, the method from [1] was adopted, with the hyper-parameters listed in Table. 1.

Table 1. Hyper-parameters of CT Noise Synthesis

Parameter	Value	Description
I_0	0.5×10^5	Photon statistics of X-ray
μ_w	20	Attenuation ratio of water
μ_a	0.02	Attenuation ratio of air
N	1000	Number of projection angles
WL	-200	Display window level
WW	2000	Display window width

For MRI, the non-local means is a popular choice to suppress Rician noise [2], which was implemented in the Python library scikit-image [3] with the hyper-parameters listed in Table. 2.

Table 2. Hyper-parameters of non-local means denoising for MRI

Parameter	Value	Description
patch size	7	Size of extracted patches
patch distance	11	Maximal distance to search patches
h	0.2	Cut-off distance
σ	Estimated from input	Standard deviation of noise

References

1. Leuschner, J., Schmidt, M., Baguer, D.O., Maass, P.: Lodopab-ct, a benchmark dataset for low-dose computed tomography reconstruction. Scientific Data **8**(1), 109 (2021)
2. Manjón, J.V., Carbonell-Caballero, J., Lull, J.J., García-Martí, G., Martí-Bonmatí, L., Robles, M.: Mri denoising using non-local means. Medical image analysis **12**(4), 514–523 (2008)
3. Van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., Yu, T.: scikit-image: image processing in python. PeerJ **2**, e453 (2014)