On the performance of learned and fixed-framelet shrinkage networks for low-dose CT denoising

Luis A. Zavala-Mondragon[1] L.A.ZAVALA.MONDRAGON@TUE.NL
Peter H. N. de With[1] P.H.N.DE.WITH@TUE.NL
Fons van der Sommen[1] FVDSOMMEN@TUE.NL
[1] *Video coding architectures group, Eindhoven University of Technology*

Editors: Under Review for MIDL 2022

Abstract

The recently introduced wavelet shrinkage networks (WSNs) are models with a performance close to state-of-the-art CT denoising CNNs, but they are faster and have less parameters. Here, we compare elements of two WSNs. The DHSN2 where the encoding-decoding (ED) path is composed by fixed convolution filters/framelets and the noise reduction is achieved through a CNN in the skip connection. Alternatively, the LWFSN where the ED path is learned and denoising is achieved by an ensemble of semi-hard thresholds. Although both models have been used for CT denoising, heterogeneities in data partitioning, training strategies and overall design, do not allow for direct evaluation of the benefits of having a trainable ED path and using a more elaborated design of a shrinkage CNN. This paper compares these issues by evaluating WSNs under common conditions. Our results show that the configuration with the best trade-off between performance and total trainable parameters is the combination of a learned framelet in the ED path with a simple thresholding layer in the skip connection. In addition, we observe that the CNN with fixed ED improves the most from using a CNN in the skip connection, but a careful design is required of the intermediate CNN to avoid extreme increases in trainable parameters.

Keywords: CNNs, shrinkage, denoising, CT.

1. Introduction

Noise reduction to improve image quality of low-dose CT has received significant attention from the deep-learning community, which has resulted in the development of diverse CNNs, among which the wavelet shrinkage networks (WSNs) (Zavala-Mondragon et al., 2021, 2022) that were recently proposed. The WSNs have linear encoding-decoding (ED) paths which are wavelet frames/framelets and a nonlinear shrinkage section positioned in the middle of the ED path. This simple structure allows for interpretable and fast operation, less trainable parameters and performance similar to state-of-the art models.

The WSNs proposed so far are the learned wavelet frame shrinkage network (LWFSN) and the dual-Haar shrinkage network (DHSN2). The LWFSN and the DHSN2 differ in two key aspects. First, the DHSN2 uses a fixed ED path, while this is trainable in the LWFSN. Second, the shrinkage CNNs of the DHSN2 are designed loosely based on NeighShrink (Chen et al., 2005). In contrast, the LWFSN uses an ensemble of semi-hard thresholds. Figure 1 visualizes the differences between the LWFSN and DHSN2.

Despite the fact that the learned/fixed framelets in WSNs are similar and have been applied to noise reduction in CT, the different training strategies, data partitioning and complexities of the shrinkage CNNs in (Zavala-Mondragon et al., 2021) and (Zavala-Mondragon

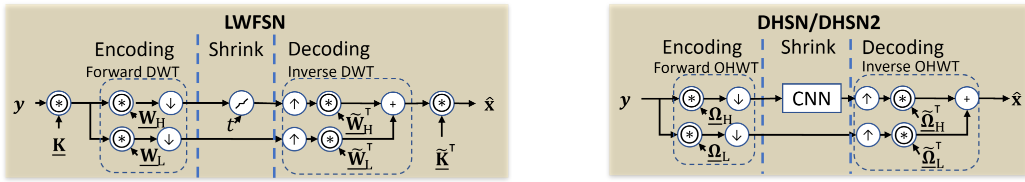

Figure 1: Simplified LWFSN and DHSN2. Variables $\mathbf{y}$ and $\mathbf{x}$ are noisy/estimated noiseless images, tensors $\underline{\mathbf{\Omega}}_{\mathrm{L}}$, $\underline{\mathbf{\Omega}}_{\mathrm{H}}$ and $\underline{\mathbf{W}}_{\mathrm{H}}$, $\underline{\mathbf{W}}_{\mathrm{H}}$ are low- and high-frequency filters of a redundant framelet and the discrete wavelet transform (DWT), while $\underline{\mathbf{K}}$ is a convolution filter and $\underline{t}$ is a learned threshold.

et al., 2022) do not allow for direct comparison of the (dis)advantages of using a fixed/learned framelet. This paper addresses this limitation by comparing the noise reduction performance of WSNs using the same shrinkage CNNs in fixed- and learned-ED, while training the CNNs with a common dataset, the same data partitioning, training loop and image quality metrics as in (Zavala-Mondragon et al., 2022).

2. Methods

To assess the advantages/disadvantages of using a learned/fixed frame as ED path, we propose four different architectures, which are listed as follows. (1) Fixed framelet with soft thresholds (FF-ST), (2) fixed framelet with CNN threshold (FF-CN), (3) learned framelet with soft thresholds (LF-ST) and finally, (4) learned framelet with CNN threshold (LF-CN). The ED path of the CNNs with fixed frame is adopted from (Zavala-Mondragon et al., 2021), while for the learned framelet, the ED path is adopted from (Zavala-Mondragon et al., 2022). For reference, we show the FF-ST, FF-CN, LF-ST, LF-CN with one decomposition level, and adopt the thresholding CNN in Figure 2.

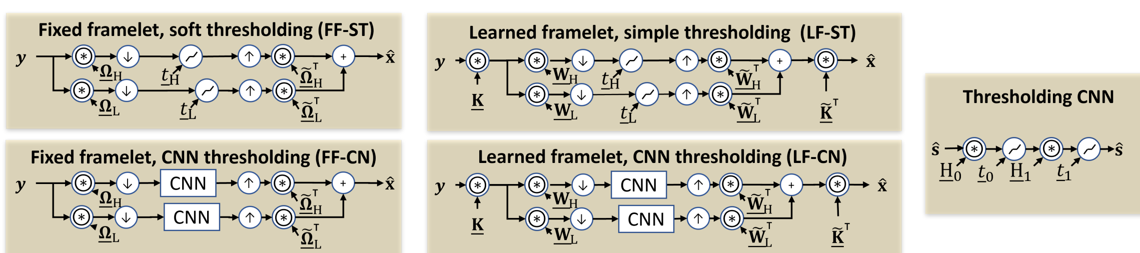

Figure 2: CNNs evaluated in this paper. Convolution kernels $\underline{\mathbf{H}}_0$ and $\underline{\mathbf{H}}_1$ have filters of 3×3. Kernel $\underline{\mathbf{H}}_0$ duplicates the feature maps and $\underline{\mathbf{H}}_1$ reduces the number of feature maps to the same amount as used in the input of the thresholding CNN. All the activation layers are soft thresholdings.

3. Experiments and discussion

For the experiments presented in this section, we use the full- and simulated, reduced-dose brain CT scans (FDCT and RDCT, respectively) (Moen et al., 2020; Clark et al., 2013) where the same partitioning of training and test scans is used as in (Zavala-Mondragon et al., 2022). All the CNNs use 4 decomposition levels and are trained for 500 epochs with the Adam optimizer with a linearly decaying learning rate and an initial value of 5×10^{-4}.

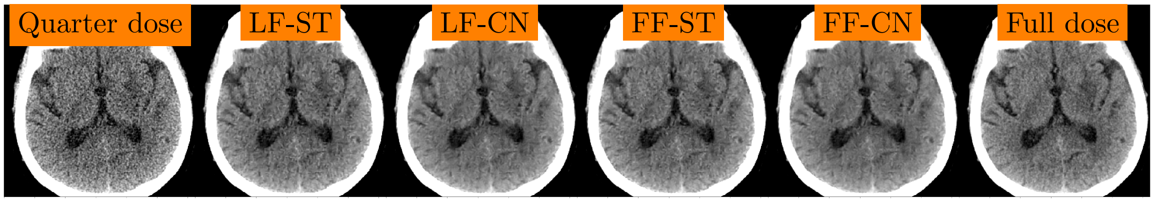

Figure 3: CN slice processed with the models presented in this paper.

Table 1 presents the resulting comparison between architectures, where NTrP stands for *number of trainable parameters* and AU for *arbitrary units*. Note that FF-CN performs best, but also has by far the largest amount of parameters. From the table, LF-ST has a similar performance, but with a considerably smaller number of parameters. This architecture is also closest to the original LWFSN. Note that the FF-CN and LF-CN have the largest amount of parameters, since the CNNs in the skip connection have many input/output channels, which requires convolution tensors with more elements. Interestingly, the more complicated shrinkage step in the skip connection improves significantly the performance of FF-CN, but only marginally for LF-CN.

Table 1: Average metrics for the models under comparison.

Model	HaarPSI [AU]	MSSIM [AU]	PSNR [dB]	NTrP
Input	0.796	0.654	25.65	
FF-ST	0.828	0.663	26.76	**1,276**
FF-CN	**0.837**	**0.664**	**27.73**	22,651,956
LF-ST	0.833	0.660	27.04	**32,904**
LF-CN	**0.838**	**0.670**	**27.510**	1,796,184

4. Conclusions

Our results show that the configuration with the best trade-off between performance and total trainable parameters is to use a learned framelet in the ED path, combined with a simple thresholding layer in the skip connection. In addition, we observe that the network with fixed ED improves the most when using a CNN in the skip connection, but a careful design of the intermediate CNN is required to decrease the trainable parameters (e.g. (Zavala-Mondragon et al., 2021)).

References

G. Y. Chen, T. D. Bui, and A. Krzyżak. Image denoising with neighbour dependency and customized wavelet and threshold. *Pattern recognition*, 38(1):115–124, 2005.

K. Clark et al. The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of digital imaging*, 26(6):1045–1057, 2013.

T. R. Moen et al. Low dose ct image and projection dataset. *Medical Physics*, 2020.

L. A. Zavala-Mondragon et al. Image noise reduction based on a fixed wavelet frame and cnns applied to ct. 30:9386–9401, 2021.

L. A. Zavala-Mondragon et al. Noise reduction in ct using learned wavelet-frame shrinkage networks. *IEEE Transactions on Medical Imaging*, 2022.