Window-Level is a Strong Denoising Surrogate

Ayaan Haque[1,2], Adam Wang[2], and Abdullah-Al-Zubaer Imran[2]

[1] Saratoga High School, Saratoga, CA, USA
[2] Stanford University Department of Radiology, Stanford, CA, USA

Abstract. CT image quality is heavily reliant on radiation dose, which causes a trade-off between radiation dose and image quality that affects the subsequent image-based diagnostic performance. However, high radiation can be harmful to both patients and operators. Several (deep learning-based) approaches have been attempted to denoise low dose images. However, those approaches require access to large training sets, specifically the full dose CT images for reference, which can often be difficult to obtain. Self-supervised learning is an emerging alternative for lowering the reference data requirement facilitating unsupervised learning. Currently available self-supervised CT denoising works are either dependent on foreign domains or pretexts that are not very task-relevant. To tackle the aforementioned challenges, we propose a novel self-supervised learning approach, namely Self-Supervised Window-Leveling for Image DeNoising (SSWL-IDN), leveraging an innovative, task-relevant, simple, yet effective surrogate—prediction of the window-leveled equivalent. SSWL-IDN leverages residual learning and a hybrid loss combining perceptual loss and MSE, all incorporated in a VAE framework. Our extensive (in- and cross-domain) experimentation demonstrates the effectiveness of SSWL-IDN in aggressive denoising of CT (abdomen and chest) images acquired at 5% dose level only. [1]

Keywords: Computed Tomography · Image Denoising · Self-Supervised Learning · Window-Leveling · VAEs

1 Introduction

Computed Tomography (CT) imaging is one of the fundamental imaging modalities in medical practice. However, X-ray radiation is a clinical concern, as high radiation can be harmful to patients [3]. Low dose CT (LDCT) images could be acquired to reduce radiation dose as an alternative to full dose CT (FDCT). However, lowering radiation dose introduces higher noise and various imaging artifacts, resulting in degraded diagnostic and other image-based performance. To address this tradeoff, deep learning-based *denoising* methods have been investigated to improve and enhance CT imaging. Conventionally, denoising models map noisy LDCT (input) to cleaner FDCT (target). CT denoising is a popular field of research because of its clinical importance, as being able to denoise, and thus use, low dose CT provides improved patient safety and diagnostic performance. Approaches include new architectures [7, 5, 21, 1] and training procedures [6, 22].

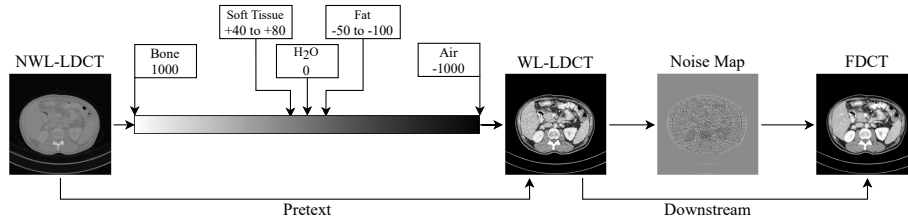[1] Code available at https://github.com/zubaerimran/SSWL-IDN

Fig. 1. Window-Leveling is the process of using CT numbers to adjust the contrast and brightness of the image. This image modification as a pretext learns important representations of the data, improving downstream denoising of predicting FDCT from LDCT by removing noise.

Acquiring reference images is challenging due to the harmful nature of radiation as well as the difficulty of performing two identical scans at different radiation doses. Thus, it is desirable to train denoising models with limited reference data. Self-Supervised Learning (SSL) has emerged as a promising alternative to fully-supervised learning in order to utilize large unlabeled training examples. In an SSL scheme, synthetic labels can be generated from the data itself, for both labeled and unlabeled data. Similar to transfer learning, SSL pre-trains a model on a surrogate task, but on the same dataset instead of one from a foreign domain, and then fine-tunes the pretrained model on a downstream, or main evaluation, task [17]. This SSL is not to be confused with other methods that are also called self-supervised which use no reference scans. Common surrogates available in literature include rotation prediction, colorization/restoration, and patch prediction. In general-purpose SSL denoising, popular works include [20, 14, 16, 19]. In SSL CT denoising, popular works include [13, 18, 24]. However, these specific methods do not use any reference scans or have a downstream task, meaning they are more unsupervised than self-supervised. Thus, as argued by these papers, a method using FDCT references, like ours, is not comparable.

Variational autoencoders (VAEs) [12] are an extension of autoencoders (AEs), which use encoders and decoders to deconstruct inputs to low-dimensional representations and then reconstruct it. VAEs are generative as they use the *reparameterization trick* to inject randomized noise into the latent code. For denoising, VAEs have not been extensively used [10, 25]. Additionally, residual learning [9] has gained interest in deep learning, so Residual (ResNet) VAEs have been proposed for image generation [11]. In medical image denoising, to our best knowledge, there is little literature using VAEs [2], and the use of RVAEs are even more scarce. Additionally, recent works support the use of Perceptual Loss because it optimizes high-level feature learning [1, 21] as opposed to Mean-Squared Error (MSE) which optimizes on a pixel-wise scale for precise noise removal. Only a few CT denoising methods have used hybrid losses [15, 8].

In this paper, we use SSL to improve performance of deep denoising models with limited reference FDCT. We propose a novel denoising surrogate to predict *window-leveled* CT images from non-window-leveled images. Window-leveling in CT is the process of modifying the grayscale of an image, using the CT
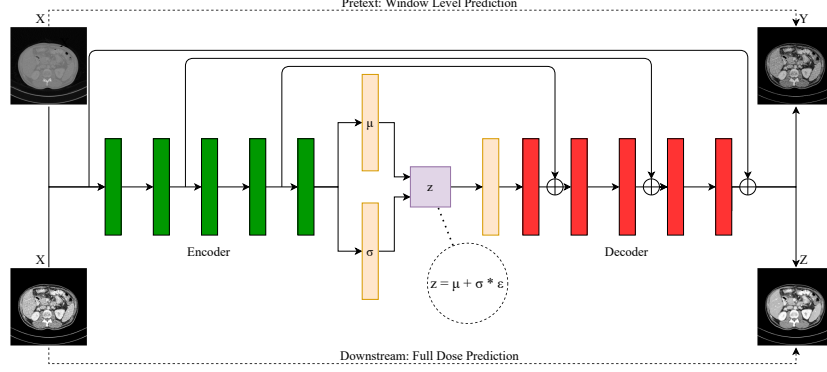
Fig. 2. Schematic of the proposed SSWL-IDN model. As a pre-text, the model predicts window-leveled images from non-window-leveled images. The model architecture uses residuals between the encoder and decoder with a VAE bottleneck.

numbers, to highlight, brighten, and contrast important structures. Unlike many other existing self-supervised learning methods, our proposed self-supervised window-leveling (SSWL) is a task-relevant surrogate, as it is directly related to the downstream task, prioritizing similar feature learning. Furthermore, we limit all our experiments to 5% dose level potentially towards an aggressive dose reduction mechanism to demonstrate effectiveness even at such low dose settings.

Our primary contributions can be summarized as follows:

– A novel and task-relevant self-supervised window-level prediction CT denoising surrogate which is related to the downstream task.
– An innovative residual-based VAE architecture coupled with a hybrid loss function to simultaneously penalize the model pixel-wise and perceptually.
– Extensive experimentation with varied quantities of labeled data on different proposed components on in- and cross-domain data demonstrating improved and effective denoising even from extremely low dose (5%) CT images.

2 Methods

2.1 Denoising

To formulate the problem, we assume unknown data distribution $p(X, Y)$ over LDCT X and FDCT Y. We also assume access to labeled training set $\mathcal{D}_l$ sampled i.i.d. from $p(X, Y)$ and unlabeled training set $\mathcal{D}_u$ sampled i.i.d. from $p(X)$ after marginalizing out Y. In CT denoising, the input images are LDCT and the reference images are FDCT. This relationship can be represented by the equation

$$X = Y + n \tag{1}$$

where n is the resultant noise due to lowering dose. The deep denoising model is trained to remove n by encoding the input LDCT and recovering the FDCT.

A similar relationship can be found in the CT window-leveling task—non-window-leveled (NWL) scans as inputs and window-leveled (WL) scans for references. This relationship can be similarly represented as,

$$Z = aX + b, \tag{2}$$

where X is NWL, Z is WL, and a and b are window-leveling parameters (determined by the DICOM metadata). Fig. 1 illustrates the process of window-leveling and its relation to denoising. Inspired from the relatedness of the two tasks, we leverage the first task to help learn or improve the second. The window-leveling labels enable us to train a deep denoising model as if it is fully-supervised. Specifically, formulating it as a pretext to the downstream denoising task is more appropriate when obtaining full dose reference images is difficult. Window-leveling is similar to denoising as both tasks from a computational view are image modifications. Since the task is domain-specific, it allows for more relevant feature learning than foreign or arbitrary surrogates.

Therefore, our proposed self-supervised learning training comprises of two steps: fully-supervised pre-training on the window-leveling task followed by fine-tuning on the small labeled denoising task. For pre-training, we prepare both a NWL and WL version of each LDCT scan for both labeled and unlabeled data. Loss is optimized for predicting the WL LDCT from input NWL LDCT. Our surrogate is end-to-end as opposed to many other methods which do not use related tasks, as no architectural or loss changes are required between tasks. After pre-training, we fine-tune the pre-trained network on the standard denoising task only with the limited pairs of LDCT and FDCT without freezing parameters.

2.2   Model Architecture and Training

For the model architecture, illustrated in Figure 2 along with the SSL algorithm, we propose a Residual Variational Autoencoder (RVAE), which is a combination of [5] and [12]. While Residual-based VAEs have been proposed [11], they use residuals in the encoder and decoder separately (ResNet as encoder and Transposed ResNet as decoder) instead of using residual connections *between* the encoder and decoder like [5].

We use the base architecture of [5] and add a bottleneck component with Global Average Pooling and Linear Layers. This downsamples the input to a latent representation, where we use the reparameterization trick on the latent code, z. This improves FD predictions, as adding randomized noise ϵ, which is tunable through learnable parameters μ (mean) and σ (standard deviation), in the bottleneck can decrease overfitting, improve generalization, and act as a regularizer, which is experimentally verified through cross-domain evaluation. To reparameterize, we use the calculation $z = \mu + \sigma * \epsilon$. A generative model can allow for better FD predictions, as the noise in LDCT may hide important details and features which can be more easily recovered through generative models, as denoising tends to oversmooth and remove subtle information. As opposed to traditional VAEs, we use constant convolutional filters of 96 instead of downsampling convolutional layers. We use residuals from feature maps in the

encoder and add them after corresponding layers in the decoder phase. We use standard convolutional layers in the encoder and transpose convolutional layers in the decoder.

For our loss, we use a hybrid loss combining MSE and perceptual loss. Perceptual loss encourages high-level visual feature learning/matching, while MSE optimizes precise, pixel-wise noise removal. Thus, we combine both into one loss to obtain both benefits. We use [27]'s perceptual loss where the model prediction and target are passed to an Image-Net pre-trained VGG-19 and the extracted features from hidden convolutional layers are used to calculate perceptual distance between the features. The perceptual loss can be defined as

$$\mathcal{L}_{perceptual} = \frac{1}{M} \sum_{i=1}^{M} \left\| \mathcal{F}(\hat{y}_i) - \mathcal{F}(y_i) \right\|^2, \tag{3}$$

where M is the mini-batch size, $\hat{y}$ are the model predictions, y are the labels, and $\mathcal{F}$ is a feature extractor.

Our final loss function can be represented by

$$L(y, \hat{y}, \mu, \sigma) = L_{MSE}(\hat{y}, y) + \beta \mathcal{L}_{perceptual}(\mathcal{F}(\hat{y}), \mathcal{F}(y)) + \alpha \mathcal{L}_{KL}(\mu, \sigma), \tag{4}$$

where L_{MSE} is standard MSE loss, the $\mathcal{L}_{perceptual}$ is perceptual loss, and β is $\mathcal{L}_{perceptual}$ weight. For the VAE, $\mathcal{L}_{KL}$ represents the KL divergence loss, and α is the $\mathcal{L}_{KL}$ weight. μ is the mean term, and σ is the standard deviation term, both from the latent space. The KL divergence attempts to reduce divergence of μ and σ in the training distribution from those of the target distribution. Both the surrogate and downstream task are trained with the same loss and architecture.

3   Empirical Evaluation

3.1   Data

We primarily collect abdomen scans from the publicly available Mayo CT data [4, 23]. The dataset includes CT scans originally acquired at routine dose level (full dose), so simulated quarter dose images are reconstructed through inserting Poisson noise into each projection dataset. For thorough denoising evaluation, we generate the CT scans at 5% dose level using the full dose and quarter dose data (scaling the zero-mean independent noise from 25% to 5% dose level). While from a clinical perspective it is more ideal to have a well-denoised quarter dose as opposed to a lower quality denoised 5% dose, from a computational perspective, showing the ability to remove high volumes of noise can more appropriately evaluate the model's full potential to accurately remove noise. We use 15 full dose abdomen CT and the corresponding quarter (25%) dose CT scans: 10 scans (1,533 slices, 15% for validation) for training and 5 (633 slices) for testing. 5 chest scans (1,061 slices) are selected from the same library for cross-domain evaluation.[3]

---

[3] For chest scans, the 5% dose level is simulated from routine and 10% dose level scans available in the Mayo data library

3.2   Implementation Details

Baselines: For architectural baselines, we used a VAE [12], DnCNN [26], and RED-CNN [5]. For SSL surrogate task baselines, we use simple reconstruction (Rec), a recent general method Noisy-as-Clean (NAC) [20] as it uses a surrogate to downstream training like us, and Noise2Void [13], even though it is a no reference method, meaning a comparison is not fitting. **Training:** Models were trained at varying levels of supervision in terms of the number of labeled data (250, 500, 1000, full). Each experiment was repeated 5 times and mean scores were reported. All inputs were normalized and resized to $256 \times 256 \times 1$. **Hyperparameters:** The code was written in Python and PyTorch (newest versions) and trained on an NVIDIA K80 GPU with 12GB RAM. We used the Adam optimizer with learning rates of 1e-5, momentum of 0.1 per 8 epochs, and a minibatch size 10. For loss, $\beta = 0.6$ and $\alpha = 1.0$ (as per VAEs). Based on tuning experiments, choice of hyperparameters do not noticeably affect performance. All parameters are identical for pre-text and downstream training. **Evaluation:** For evaluation, we used Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Mean-Squared Error (MSE), and Normalized RMSE (NRMSE). To evaluate statistical significance, we performed two-sample t-tests.

3.3   Results & Discussion

Table 1. Comparison of our RVAE architecture against other conventional and SoTA architectures on in- and cross-domain evaluation. The best fully-supervised scores are bolded, and best semi-supervised scores are underlined.

Model	$|\mathcal{D}_l|$	In-Domain						Cross-Domain					
		LDCT	U-Net	VAE	DnCNN	RED-CNN	RVAE	LDCT	U-Net	VAE	DnCNN	RED-CNN	RVAE
PSNR	250	20.179	17.325	17.278	19.814	23.414	24.829	16.876	16.434	16.318	16.861	18.625	19.033
	500	—	18.445	19.241	20.430	23.612	26.139	—	17.385	17.329	17.192	18.919	19.193
	1000	—	20.930	20.771	21.693	24.097	<u>26.286</u>	—	18.458	18.094	18.702	19.005	<u>19.259</u>
	Full	—	21.900	21.547	21.186	24.115	**26.574**	—	18.843	18.364	18.925	19.005	**19.288**
SSIM	250	0.7554	0.5685	0.4656	0.7031	0.8433	0.8483	0.6725	0.4781	0.3741	0.6132	0.6971	0.7438
	500	—	0.6606	0.5849	0.7372	0.8522	0.8538	—	0.6094	0.4859	0.6554	0.7264	0.7467
	1000	—	0.7697	0.6873	0.7460	0.8538	<u>0.8626</u>	—	0.6883	0.5634	0.6343	0.7327	<u>0.7497</u>
	Full	—	0.7918	0.7510	0.7907	0.8616	**0.8646**	—	0.7095	0.5971	0.6584	0.7467	**0.7490**
MSE	250	0.0107	0.0201	0.0192	0.0112	0.0054	0.0053	0.0207	0.0234	0.0238	0.0207	0.0139	0.0131
	500	—	0.0144	0.0121	0.0101	0.0051	0.0050	—	0.0187	0.0188	0.0202	0.0132	0.0127
	1000	—	0.0083	0.0086	0.0071	0.0052	<u>0.0046</u>	—	0.0147	0.0148	0.0157	0.0129	<u>0.0125</u>
	Full	—	0.0067	0.0073	0.0088	0.0047	**0.0045**	—	0.0135	0.0145	0.0136	0.0124	**0.0122**
NRMSE	250	0.3082	0.4204	0.4173	0.3171	0.2155	0.1889	0.4832	0.5186	0.5245	0.4777	0.4319	0.3861
	500	—	0.3629	0.3324	0.2981	0.2078	0.1889	—	0.4647	0.4671	0.4740	0.4019	0.3789
	1000	—	0.2746	0.2802	0.2533	0.1990	<u>0.1878</u>	—	0.4110	0.4266	0.4217	0.3819	<u>0.3763</u>
	Full	—	0.2470	0.2573	0.2749	0.1979	**0.1817**	—	0.3933	0.4091	0.3662	0.3745	**0.3690**

Table 1 shows our proposed RVAE alone is able to outperform all baselines, including two state-of-the-art denoising architectures in DnCNN [26] and RED-CNN [5], and even with minimum data, we are able to match the fully-supervised metrics of those two architectures. While DnCNN and RED-CNN are older methods, as architectures alone, they are still SoTA. Our RVAE outperforms RED-CNN with statistical significance (p-value of 0.032 for PSNR), and also significantly outperforms the other models ($p < 0.05$). For the cross-domain chest

dataset evaluation, similar improvements are shown from RVAE compared to the baselines and the state-of-the-art, proving the generalizatibility of the RVAE.

Table 2 displays the performance of RED-CNN and our RVAE with various SSL tasks (trained on MSE). SSL Reconstruction simply uses the LDCT as the input and reference, allowing the model to become familiar with the data. For the Noisy-As-Clean surrogate task [20], we follow their implementation where their surrogate task uses LDCT with additional injected Gaussian noise as the input and the plain LDCT as the reference, and subsequently perform denoising. As shown in the tables, SSWL has significantly improved performance over no pre-training and basic reconstruction (p-value of 0.021 and 0.039 respectively). Compared to NAC, the SoTA, we still have the best performance, confirming the importance and ability of our proposed SSL task. Against N2V, a SoTA no-reference method, we greatly outperform them (p-value < 0.5), proving the effectiveness of our new method. Similar to in-domain evaluations, the cross-domain metrics further confirm the superior performance and generalization of SSWL, showing the importance of a task-relevant surrogate.
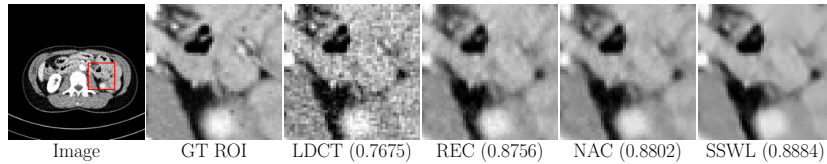

Image   GT ROI   LDCT (0.7675)   REC (0.8756)   NAC (0.8802)   SSWL (0.8884)

Fig. 3. ROI predictions with RVAE ($|\mathcal{D}_l|$=500) and the 3 SSL tasks show higher visual noise removal and SSIM from SSWL.

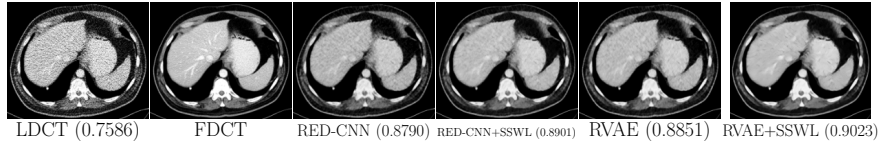LDCT (0.7586)   FDCT   RED-CNN (0.8790)   RED-CNN+SSWL (0.8901)   RVAE (0.8851)   RVAE+SSWL (0.9023)

Fig. 4. FDCT predictions from RED-CNN and RVAE ($|\mathcal{D}_l|$=500) with and without SSWL show strong visual improvements and improved SSIM from our proposed models.

The improvements from our hybrid loss are shown in our final model, SSWL-IDN, in Table 2. When compared to RVAE + SSWL, which is trained on MSE, we see improved performance of up to 1 to 2 scores higher for both PSNR and SSIM. A full ablation for our hybrid loss is available in the supplemental. Figs. 3 and 4 demonstrate precise removal of noise from whole scans as well as specific regions of interest (ROIs), proving the effectiveness of our model over baseline architectures and other self-supervised tasks. While certain structural details are lost, this is due to the ultra low dose, and our method recovers details best.

Table 2. SSWL is compared to Rec, NAC, and N2V trained on RED-CNN and RVAE architectures. SSWL is shown to be the most effective. Follows same key as Table 1.

Model	$|\mathcal{D}_l|$	In-Domain				Cross-Domain			
		PSNR	SSIM	MSE	NRMSE	PSNR	SSIM	MSE	NRMSE
RED-CNN + REC	250	23.414	0.8433	0.0054	0.2155	19.351	0.7551	0.0122	0.3726
	500	23.612	0.8522	0.0051	0.2078	19.430	0.7572	0.0120	0.3691
	1000	24.097	0.8538	0.0052	0.1990	19.499	0.7523	0.0117	0.3644
	Full	24.115	0.8616	0.0047	0.1979	19.546	0.7555	0.0120	0.3678
RVAE + REC	250	23.795	0.8522	0.0051	0.2120	19.417	0.7583	0.0120	0.3694
	500	24.088	0.8598	0.0050	0.2067	19.544	0.7602	0.0117	0.3640
	1000	24.071	0.8650	0.0048	0.2007	19.570	0.7598	0.0116	0.3632
	Full	24.301	0.8669	0.0047	0.1958	19.609	0.7613	0.0115	0.3614
RED-CNN + NAC	250	23.989	0.8375	0.0050	0.2050	19.122	0.7527	0.0124	0.3750
	500	24.033	0.8541	0.0048	0.2015	19.401	0.7565	0.0127	0.3703
	1000	24.219	0.8549	0.0045	0.1961	19.494	0.7577	0.0118	0.3660
	Full	24.422	0.8611	0.0045	0.1958	19.579	0.7588	0.0116	0.3625
RVAE + NAC	250	23.890	0.8612	0.0052	0.2029	19.313	0.7545	0.0123	0.3739
	500	24.168	0.8579	0.0047	0.1996	19.338	0.7532	0.0123	0.3726
	1000	24.019	0.8647	0.0047	0.1980	19.439	0.7580	0.0122	0.3686
	Full	24.186	0.8649	0.0046	0.1956	19.520	0.7596	0.0118	0.3650
RED-CNN + N2V	250	23.145	0.7899	0.008	0.2505	18.310	0.7170	0.0148	0.4098
	500	23.743	0.8000	0.0075	0.2390	18.075	0.7212	0.0157	0.4206
	1000	24.020	0.8032	0.0073	0.2342	18.169	0.7101	0.0153	0.4162
	Full	24.116	0.8083	0.0072	0.2318	18.417	0.7215	0.0145	0.4047
RVAE + N2V	250	23.928	0.8057	0.0073	0.2352	18.241	0.7177	0.0151	0.4129
	500	24.139	0.8107	0.0070	0.2297	18.307	0.7193	0.0149	0.4010
	1000	24.072	0.8111	0.0070	0.2301	18.313	0.7186	0.0148	0.4099
	Full	24.321	0.8113	0.0069	0.2269	18.174	0.7155	0.0153	0.4162
RED-CNN + SSWL	250	26.119	0.8509	0.0050	0.1890	19.550	0.7611	0.0117	0.3635
	500	26.300	0.8542	0.0050	0.1865	19.460	0.7614	0.0120	0.3679
	1000	26.710	0.8627	0.0046	0.1786	19.520	0.7617	0.0120	0.3652
	Full	26.747	0.8626	0.0045	0.1764	19.547	0.7619	0.0117	0.3642
RVAE + SSWL	250	26.150	0.8612	0.0051	0.1900	19.566	0.7632	0.0116	0.3630
	500	26.464	0.8659	0.0048	0.1820	19.619	0.7634	0.0115	0.3607
	1000	26.799	0.8669	<u>0.0043</u>	0.1781	19.549	0.7619	0.0117	<u>0.3505</u>
	Full	26.844	0.8701	0.0044	0.1774	19.617	0.7624	0.0115	0.3692
SSWL-IDN	250	26.581	0.8649	0.0046	0.1793	19.660	0.7645	0.0109	0.3556
	500	26.778	0.8723	0.0045	0.1783	19.854	0.7680	0.0107	0.3530
	1000	<u>27.018</u>	<u>0.8744</u>	<u>0.0043</u>	<u>0.1732</u>	<u>20.016</u>	<u>0.7706</u>	<u>0.0105</u>	<u>0.3505</u>
	Full	**27.800**	**0.8815**	**0.0042**	**0.1701**	**20.178**	**0.7739**	**0.0104**	**0.3458**

4   Conclusions

We present SSWL-IDN, a self-supervised denoising model with a novel, task-relevant, and efficient surrogate task of window-level prediction. We also propose a Residual-VAE specialized for denoising, as well as a hybrid loss leveraging benefits of both perceptual and pixel-wise optimization. We confirm each component of our method outperforms baselines on difficult 5% dose denoising for both in- and cross-domain evaluations, and when combined, the model significantly outperforms state-of-the-art methods. Improved denoising with limited reference data is of clinical significance to reduce harms to patients. Our future work will focus on developing cascaded and joint surrogate and downstream learning as well as 3D architectures to utilize information in the z-dimension.

References

[1] Ataei, S., Alirezaie, J., Babyn, P.: Cascaded convolutional neural networks with perceptual loss for low dose CT denoising. In: 2020 International Joint Conference on Neural Networks (IJCNN). pp. 1–5. IEEE (2020) 1, 2

[2] Biswas, B., Ghosh, S.K., Ghosh, A.: Dvae: deep variational auto-encoders for denoising retinal fundus image. In: Hybrid Machine Intelligence for Medical Image Analysis, pp. 257–273. Springer (2020) 2

[3] Brenner, D.J., Hall, E.J.: Computed tomography — an increasing source of radiation exposure. New England Journal of Medicine 357(22), 2277–2284 (2007), https://doi.org/10.1056/NEJMra072149, pMID: 18046031 1

[4] Chen, B., Duan, X., Yu, Z., Leng, S., Yu, L., McCollough, C.: Development and validation of an open data format for CT projection data. Medical physics 42(12), 6964–6972 (2015) 5

[5] Chen, H., Zhang, Y., Kalra, M.K., Lin, F., Chen, Y., Liao, P., Zhou, J., Wang, G.: Low-dose CT with a residual encoder-decoder convolutional neural network. IEEE Transactions on Medical Imaging 36(12), 2524–2535 (Dec 2017), http://dx.doi.org/10.1109/TMI.2017.2715284 1, 4, 6

[6] Chen, H., Zhang, Y., Zhang, W., Liao, P., Li, K., Zhou, J., Wang, G.: Low-dose ct denoising with convolutional neural network. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). pp. 143–146. IEEE (2017) 1

[7] Diwakar, M., Kumar, M.: A review on ct image noise and its denoising. Biomedical Signal Processing and Control 42, 73–88 (2018) 1

[8] Gholizadeh-Ansari, M., Alirezaie, J., Babyn, P.: Deep learning for low-dose ct denoising using perceptual loss and edge detection layer. Journal of digital imaging pp. 1–12 (2019) 2

[9] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016) 2

[10] Im Im, D., Ahn, S., Memisevic, R., Bengio, Y.: Denoising criterion for variational auto-encoding framework. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 31 (2017) 2

[11] Kingma, D.P., Salimans, T., Jozefowicz, R., Chen, X., Sutskever, I., Welling, M.: Improving variational inference with inverse autoregressive flow. arXiv preprint arXiv:1606.04934 (2016) 2, 4

[12] Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013) 2, 4, 6

[13] Krull, A., Buchholz, T.O., Jug, F.: Noise2void-learning denoising from single noisy images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2129–2137 (2019) 2, 6

[14] Laine, S., Karras, T., Lehtinen, J., Aila, T.: High-quality self-supervised deep image denoising. Advances in Neural Information Processing Systems 32, 6970–6980 (2019) 2

[15] Ma, Y., Wei, B., Feng, P., He, P., Guo, X., Wang, G.: Low-dose ct image denoising using a generative adversarial network with a hybrid loss function for noise learning. IEEE Access 8, 67519–67529 (2020) 2

[16] Quan, Y., Chen, M., Pang, T., Ji, H.: Self2self with dropout: Learning self-supervised denoising from single image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020) 2

[17] de Sa, V.R.: Learning classification with unlabeled data. In: Advances in neural information processing systems. pp. 112–119. Citeseer (1994) 2

[18] Wu, D., Ren, H., Li, Q.: Self-supervised dynamic CT perfusion image denoising with deep neural networks. IEEE Transactions on Radiation and Plasma Medical Sciences (2020) 2

[19] Xie, Y., Wang, Z., Ji, S.: Noise2same: Optimizing a self-supervised bound for image denoising. Advances in Neural Information Processing Systems 33 (2020) 2

[20] Xu, J., Huang, Y., Cheng, M.M., Liu, L., Zhu, F., Xu, Z., Shao, L.: Noisy-As-Clean: Learning self-supervised denoising from corrupted image. IEEE Transactions on Image Processing 29, 9316–9329 (2020) 2, 6, 7

[21] Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., Kalra, M.K., Zhang, Y., Sun, L., Wang, G.: Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss. IEEE Transactions on Medical Imaging 37(6), 1348–1357 (Jun 2018), http://dx.doi.org/10.1109/TMI.2018.2827462 1, 2

[22] Yi, X., Babyn, P.: Sharpness-aware low-dose ct denoising using conditional generative adversarial network. Journal of digital imaging 31(5), 655–669 (2018) 1

[23] Yu, L., Shiung, M., Jondal, D., McCollough, C.H.: Development and validation of a practical lower-dose-simulation tool for optimizing computed tomography scan protocols. Journal of computer assisted tomography 36(4), 477–487 (2012) 5

[24] Yuan, N., Zhou, J., Qi, J.: Half2half: deep neural network based ct image denoising without independent reference data. Physics in Medicine & Biology 65(21), 215020 (2020) 2

[25] Yue, Z., Yong, H., Zhao, Q., Zhang, L., Meng, D.: Variational denoising network: Toward blind noise modeling and removal. In: The Thirty-third Annual Conference on Neural Information Processing Systems (2019) 2

[26] Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE transactions on image processing 26(7), 3142–3155 (2017) 6

[27] Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR (2018) 5