2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 978-1-7281-6215-7/20/$31.00 ©2020 IEEE DOI: 10.1109/BIBM49941.2020.9313560

# Do Noises Bother Human and Neural Networks In the Same Way? A Medical Image Analysis Perspective

Shao-Cheng Wen[1], Yu-Jen Chen[1], Zihao Liu[2], Wujie Wen[3], Xiaowei Xu[4],
Yiyu Shi[5], Tsung-Yi Ho[1], Qianjun Jia[4], Meiping Huang[4], Jian Zhuang[4]
[1]*Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan*
[2]*Department of Electrical and Computer Engineering, Florida International University, FL, USA*
[3]*Department of Electrical and Computer Engineering, Lehigh University, PA, USA*
[4]*Guangdong Provincial People's Hospital, Guangdong Acadamic of Medical Science, Guangzhou, China*
[5]*Department of Computer Science and Engineering, University of Notre Dame, IN, USA*

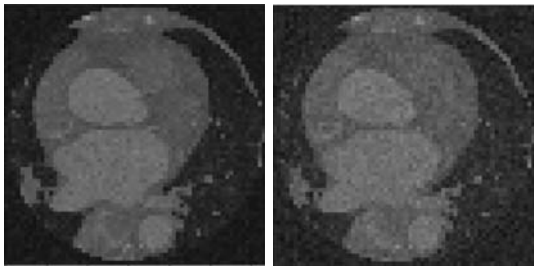
***Abstract*—Deep learning had already demonstrated its power in medical images, including denoising, classification, segmentation, etc. All these applications are proposed to automatically analyze medical images beforehand, which brings more information to radiologists during clinical assessment for accuracy improvement. Recently, many medical denoising methods had shown their significant artifact reduction result and noise removal both quantitatively and qualitatively. However, those existing methods are developed around human-vision, i.e., they are designed to minimize the noise effect that can be perceived by human eyes. In this paper, we introduce an application-guided denoising framework, which focuses on denoising for the following neural networks. In our experiments, we apply the proposed framework to different datasets, models, and use cases. Experimental results show that our proposed framework can achieve a better result than human-vision denoising network.**

***Index Terms*—Denoising, Deep Learning**

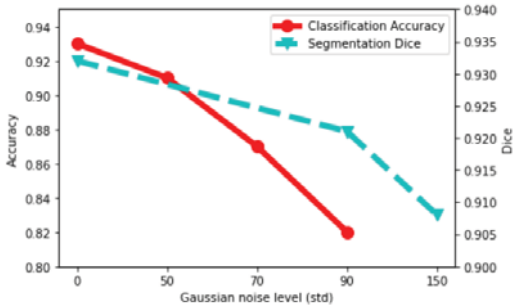
## I. Introduction

THE prevalence of deep learning in medical image computing and analysis has greatly reduced the human effort and enhanced the efficiency of diagnosis and treatment [1], [2], [3], [4]. To achieve superb performance in such tasks, high-quality medical images are often indispensable for training and testing state-of-the-art deep learning models [5], [6]. Unfortunately, these raw images inevitably suffer from high-intensity noises (see Fig. 1 (a)) due to complex clinical scenarios [7], [8], [9], which significantly jeopardizing the capability of machine learning models on image segmentation and classification. As the example in Fig. 1 (b) shows, both image segmentation performance (Dice) and classification accuracy drop dramatically with the increase of noise level. Even neural network models are trained for better generalization by using images containing the same level of noise as that of testing ones, the performance can be decreased. In contrast, testing images in the medical domain usually can be much noisier than the training dataset. This further aggravates the accuracy problem for deep learning assisted medical imaging.

This research was approved by the Research Ethics Committee of Guangdong General Hospital, Guangdong Academy of Medical Science with the protocol No. 20140316.



(a) Noise-affected examples




(b) Accuracy/Dice vs Noise level

Fig. 1. (a) Demonstrates the noise-affected test images with $\mu = 0$, $\sigma = 90$ (left) and $\mu = 0$, $\sigma = 150$ (right), respectively. (b) segmentation and classification Dice/Accuracy w.r.t. Gaussian noise level. Note that we use the dirty Multi-Modality Whole Heart Segmentation (MM-WHS) dataset to train the segmentation model No-New-Net [10] and Classification Convolutional Neural Network (CCNN) [11]. Detailed experimental settings can be found in Section III.

To tackle this issue, image denoising is typically introduced as a pre-processing step before neural-network-based image classification or segmentation. Recently, most existing denoising methods [12], such as Residual Encoder-Decoder Convolution Neural Network (RED-CNN) [13] and Multi-Channel Denoising Convolution Neural Network (MCDnCNN) [14] utilized the power of deep neural network, which attempt to learn the distributions of the noise, so as to eliminate the noises in a more elaborate manner.

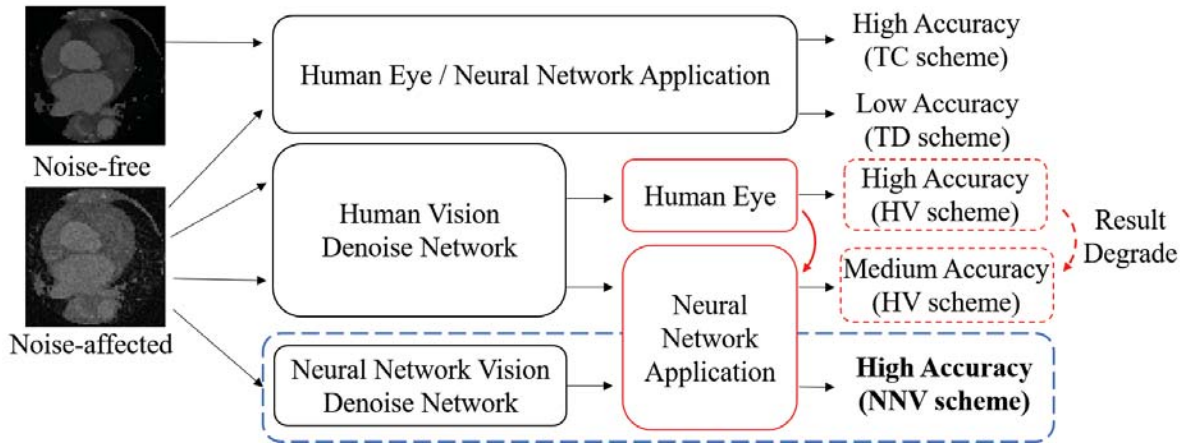
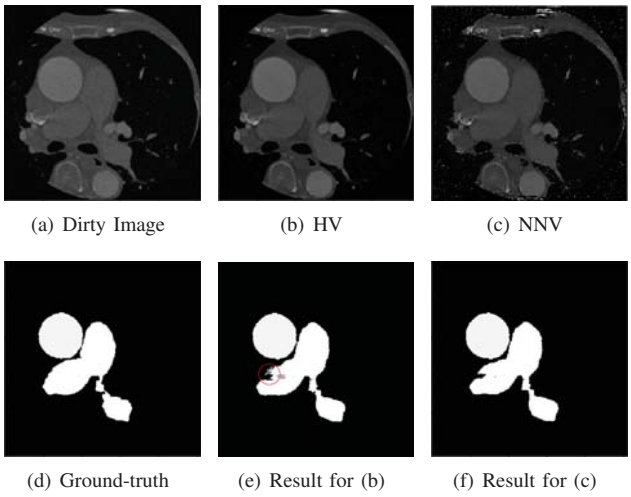



Fig. 2. Workflow comparison for input images with different noise levels and different predict sources (human and neural network application). The detail of all four schemes will be introduced in section III. We would like to point out that the result of the human-vision scheme may be degraded as the predict target changed (block marked with red). Note that the blue rectangle marked in the last row is the proposed workflow.



(a) Dirty Image (b) HV (c) NNV

(d) Ground-truth (e) Result for (b) (f) Result for (c)

Fig. 3. Segmentation result comparison for (a) dirty image and the denoised image through (b) Human-Vision (HV) and (c) Neural-Network-Vision (NNV). The corresponding dice score are 1.0 and 0.704, and 0.849. As shown in (d), the ground-truth segment the image into two classes, ascending aorta and the pulmonary artery. However, the region circled with red in (e) is the result of HV misclassified the pulmonary artery into the left atrium blood cavity, while (f) is correctly classified.

These denoising techniques could largely remove the noises based on the image visual quality measurements defined by human eyes, e.g., peak signal-to-noise ratio (PSNR) calculated by pixel-by-pixel difference between clean and its dirty version, and thereby enhance neural-network-based medical image segmentation or classification performance. We would like to argue that their advocated high denoising efficiency (dedicated to "human-vision") may not be necessarily translated into impressive accuracy improvement for neural networks (or "neural-network-vision"). A clear workflows comparison for input images with different noise levels and predict environment (human eye or neural network application) could be found in Fig. 2. From this figure, we would like to point out that the result of the human-vision scheme may be degraded as the predict environment changed.

In addition, we observed that simply removing all noises perceivable by human eyes to make images completely noise-free, sometimes may result in inferior neural network decision making. Some noises could reinforce the information deemed to be important by neural networks for better image segmentation and recognition. This can be clearly observed from Fig. 3, where Fig. 3 (b) is obtained by denoising dirty image (a) with RED-CNN guided by an Human-Vision (HV) rule. Visually, Fig. 3 (b) has a much lower level of noise compared with Fig. 3 (c), which is denoised by an Neural-Network-Vision (NNV) manner tailored for deep learning by deliberately keeping some noises. Yet surprisingly, processing both denoised images with the same segmentation network No-New-Net, suggests that the noisier one, i.e., Fig. 3 (c) denoised by NNV, could achieve a much higher Dice score than the clean version, i.e., Fig. 3 (b) denoised by HV on image segmentation. The detailed segmentation result comparison is illustrated in Fig. 3 (e) and (f).

In this paper, we propose to redefine the framework of medical image denoising by integrating the concept of "neural-network-vision". Different from the human perceived visual distortion adopted by existing denoising solutions, the proposed framework evaluates the denoising efficiency directly through the perspective of neural network computation. As a result, such denoising can deal with the noises in a way that neural network favors [15], [16], so as to significantly boost the accuracy. We validate and compare our design with state-of-the-art denoising solutions, through comprehensive experiments on both image segmentation and classification tasks.

The main contributions of our work are as follows:

- We proposed the first neural-network-guided denoising framework for image denoising, which makes the denoising network can denoise images in a way desirable by any application network.
- Experimental results show that the proposed Neural-Network-Vision-based (NNV) image-denoising method

outperforms any existing Human-Vision-based (HV) image-denoising methods in both segmentation and classification tasks.

## II. The Difference between Human-Vision and Neural-Network-Vision

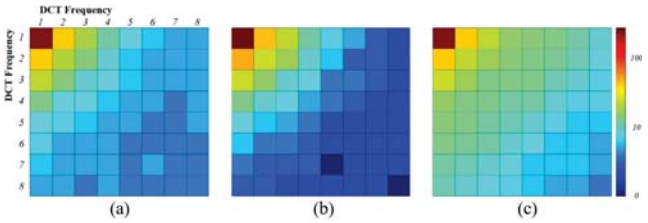



Fig. 4. DCT frequency distributions at different examples: (a) Test image; (b) HV-based denoising; (c) NNV-based denoising

**Frequency results** In order to know the difference between HV and NNV, we first transfer an image into the frequency domain by $8\times8$ 2D Discrete Cosine Transform (DCT). In this way, the image is split into multiple small $8 \times 8$ frequency coefficients blocks. We then put the frequency coefficients belong to the same frequency components together to measure its distribution of this image (totally 64 frequency distributions). Since all the distributions obey normal distribution (i.e., mean is 0), thus the standard deviation (SD) indicates the energy in each frequency component of this image (i.e., large SD means more energy in this frequency component). Fig. 4 shows the heat map of SD at each frequency component, where (a) is clean image, (b) indicates HV-based denoised image, and (c) represents the proposed NNV-based denoised image.

Obviously, the NNV-based denoised image (c) has more comprehensive information in high frequency domain compared with clean image (a).

This indicates how the segmentation network wants to change the denoised image, i.e., the additional information added in NNV-based denoised image is guided by segmentation network.

**Frequency analysis** Assume $x_k$ is a single pixel of a raw image $\mathbf{X}$, and $x_k$ can be represented by $8 \times 8$ DCT:

$$x_k = \sum_{i=0}^{i=7} \sum_{j=0}^{n=7} c_{(k,i,j)} \cdot b_{(i,j)} \tag{1}$$

where $c_{(k,i,j)}$ and $b_{(i,j)}$ are the DCT coefficient and corresponding basis function at 64 different frequencies, respectively.

Since the human visual system is less sensitive to high-frequency components, HV-based denoising is achieved by intentionally discarding the high-frequency parts $c_{(k,i,j)}$. On the contrary, Deep-Neural-Networks (DNN) examine the importance of the frequency information in a quite different way. The gradient of the DNN function $F$ with respect to a basis function $b_{(i,j)}$ can be calculated as:

$$\frac{\partial F}{\partial b_{(i,j)}} = \frac{\partial F}{\partial x_k} \times \frac{\partial x_k}{\partial b_{i,j}} = \frac{\partial F}{\partial x_k} \times c_{(k,i,j)} \tag{2}$$

Eq. 2 implies that the contribution of a frequency component ($b_{i,j}$) of a single pixel $x_k$ to the DNN learning will be mainly determined by its associated DCT coefficient ($c_{(k,i,j)}$) and the importance of the pixel ($\frac{\partial F}{\partial x_k}$). Here $\frac{\partial F}{\partial x_k}$ is obtained after the DNN training, while $c_{(k,i,j)}$ will be distorted by filtering before training. If $c_{(k,i,j)} = 0$, the frequency feature ($b_{i,j}$), which may carry important details for DNN feature map extraction, cannot be learned by DNN for weights updating, causing a lower accuracy.

As shown in Fig 4 (a), clean image has comprehensive information in all frequency domains, however (b) HV-based method discord the high frequency information which will make DNN hard to learn these features. The NNV-based method can add more features in all frequency components to make the DNN easier to learning or training.

## III. Experiments

### A. Datasets

We use two segmentation datasets and one classification dataset to evaluate the proposed framework. Multi-Modality Whole Heart Segmentation (MM-WHS) [17] dataset contains 2,557 images as the training set, and 363 images as the test set. This dataset aims to accurately segment all the substructures of the whole heart into seven categories and background, as eight classes.

Second, we examined the experiments with Brain Tumor segmentation challenge (BraTS) [18] segmentation dataset. This dataset includes 210 High Grade Glioma (HGG) cases, which consist of different MRI modalities for each patient. In the experiment, we chose post-contrast T1-weighted images as our input. Each tumor is segmented into four classes and background, as five categories. We split the dataset into the training set and test set with 1,125 images and 129 images, respectively.

For classification dataset, we utilized a brain tumor public dataset [19]. The objective of this dataset is to correctly classify the input into one of the three grades. This dataset contains 233 patients with a total of 3,064 brain T1-weighted contrast-enhanced MRI images. We split them into training set with 2,500 images and test set with 300 images.

### B. Experiment Schemes

In this paper, we compared our Neural-Network-Vision (NNV) based denoising scheme with three other schemes, segmentation or classification network Trained with Clean images (TC) and Trained with Dirty images (TD) which has the same noise level as test images, and the Human-Vision (HV) based denoising, respectively. To train all the four schemes, noises were added to datasets with different noise levels. For TC and TD schemes, they were trained using clean images and dirty images, respectively. For HV scheme, the denoising network was independently trained using paired clean images and dirty images and optimized using pixel-wise loss function, such as Mean Squared Error (MSE). Finally, the proposed framework, NNV scheme was trained with dirty images and optimized by the loss function which considered the difference between the output of the following neural networks and its ground-truth, such as cross-entropy loss.

### C. Experiment Setup

In our paper, for 3D scenario, every experiment scheme involved required input volumes which were scaled into 64×64×64. As for the experiments based on the 2D model, the input images were scaled to the size of 256×256. While training, we followed the default settings mentioned in the referred paper. For HV and MV schemes, the number of epoch was set to 300. Xavier uniform initializer was used for all kernel in every convolution and deconvolution layer. The batch size was set to 1. Adam optimization was used with learning rate at 0.00001. Moreover, all experiments done in our paper were implemented in Python3 with TensorFlow 1.14 over NVIDIA GeForce RTX 2080 Ti GPU.

For segmentation results, we follow existing works [10], [20], [21] applying Dice score and Hausdorff Distance $d_H$ for evaluation. For classification, top-1 accuracy was applied and reported in Section IV.

## IV. Results

In this section, we will discuss the experimental results which were completed using two datasets, two noise types, and two denoising networks on segmentation and classification for the four experiment schemes, as TC, TD, HV, and NNV schemes mentioned in Section III.

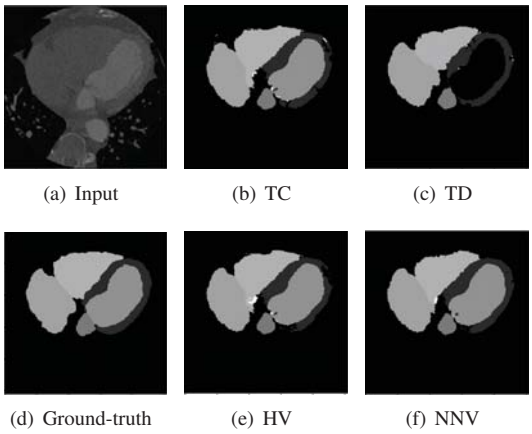
### A. Results Analysis for Segmentation

We start our discussion on No-New-Net [10] 2D segmentation. Table I reports the mean and the standard deviation (SD) of the test result using MM-WHS and BrasTS datasets, with two different noise added, Gaussian noise and Poisson noise for four experiment schemes. To show the flexibility of the proposed framework, two denoise methods, RED-CNN [13] and MCDnCNN [14] were implemented to both HV and NNV schemes.

For MM-WHS dataset, first of all, TD scheme achieved 0.134 better Dice than TC scheme on average. This is as expected since the network could learn the feature extracted from noisier images. Secondary, since denoising method was included, we believe that HV denoise network is still somehow effective. Thus, for both denoising methods, RED-CNN and MCDnCNN, HV scheme improved the Dice and the Hausdorff over both schemes without denoising network by 0.103 and 0.107, 0.77 and 0.85, respectively. Finally, NNV scheme achieved another improvement over HV scheme by 0.019 higher Dice and 0.054 lower Hausdorff distance for RED-CNN and 0.018 and 0.069 for MCDnCNN. Fig. 5 shows the input, ground-truth, and the segmentation results of four schemes.

As for the comparison with Poisson noise added in MM-WHS dataset, we can notice that TD scheme achieves higher Dice than TC scheme with 0.062 improvement. Compared with the TD scheme, the proposed NNV scheme achieves the optimal performance in both metrics which yields an improvement of 0.126 higher Dice and 0.895 lower Hausdorff distance for RED-CNN and 0.117 and 0.835 for MCDnCNN.

We also applied the same experiment to BraTS dataset. We can notice that the improvement trend of all the schemes on MM-WHS dataset is the same as that on BraTS dataset. For both experiments on Gaussian noise and Poisson noise, TD scheme outperforms TC scheme on Dice score and Hausdorff distance. With the denoising network included, HV scheme beats TD scheme with slight improvement. And finally, the proposed NNV scheme lead the optimal score in both Dice score and Hausdorff distance for RED-CNN and MCDnCNN.



(a) Input (b) TC (c) TD

(d) Ground-truth (e) HV (f) NNV

Fig. 5. Segmentation comparison for (a) input image from MM-WHS dataset. (b-c) are the segmentation results of TC and TD schemes, which denoising network is not included. (d) is the ground-truth and (e-f) are the segmentation results of HV and NNV schemes using RED-CNN denoising network. Note that Gaussian white noise $\sigma = 70$ is added to simulate dirty images in this case.

We further applied the experiment to No-New-Net 3D version, the comparison are reported in Table II. In this experiment, Gaussian white noise with $\sigma = 90$ was superposed to the MM-WHS dataset. From the table, similar to 2D MM-WHS segmentation experiment, TD scheme again achieves 0.01 higher Dice over the TC scheme. With denoising network included, HV and MV schemes outperform TD scheme up to 0.014 Dice and 0.019 sensitivity on average. Furthermore, compared with HV scheme, MV scheme achieves slightly higher Dice and sensitivity up to 0.01 and 0.005, respectively. However, since the size of input volume is only 64×64×64, all the four schemes had originally achieved over 0.8 Dice score. Thus, the improvement, though smaller than that in the No-New-Net 2D segmentation presented in the paper, is still significant. Moreover, it can be clearly seen that all the experiments result in high specificity, which means that most true negative cases can be correctly segmented.

### B. Results Analysis for Classification

In this section, we also extend the evaluation to classification using CCNN. Similar to previous section, we explored how the NNV-based denoising framework works among different noise levels. Thus, three datasets were synthesized with $\sigma = 50, 70, 90$ Gaussian white noise superposed to Brain Tumor dataset. Note that both training set and test set contain the same noise level.

From Table III, we can again observe that in both TC and TD schemes, the higher level the noise is, the lower accuracy does the classification network results in. Obviously, TD scheme outperformed TC scheme in the experiment, which

TABLE I
STATISTIC RESULT COMPARISON (MEAN±SD) FOR NO-NEW-NET 2D (256×256) MODEL USING MM-WHS AND BRATS DATASETS. RED-CNN AND MCDNCNN DENOISING NETWORKS WERE USED. GAUSSIAN WHITE NOISE $\sigma = 70$ WAS ADDED TO SIMULATE DIRTY IMAGES.

| | | Gaussian Noise | | | | Poisson Noise | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | MM-WHS dataset | | BraTS dataset | | MM-WHS dataset | | BraTS dataset | |
| Schemes | | Dice | Hausdorff | Dice | Hausdorff | Dice | Hausdorff | Dice | Hausdorff |
| w/o Denoise | TC | 0.542±0.251 | 2.723±0.976 | 0.481±0.184 | 3.047±0.837 | 0.641±0.217 | 2.655±1.725 | 0.525±0.186 | 2.502±1.055 |
| | TD | 0.676±0.257 | 2.815±2.116 | 0.497±0.203 | 2.570±0.760 | 0.703±0.160 | 2.685±1.538 | 0.588±0.168 | 2.334±0.732 |
| RED-CNN | HV | 0.779±0.177 | 1.953±1.373 | 0.577±0.169 | 2.366±0.727 | 0.779±0.182 | 1.994±1.398 | 0.590±0.169 | 2.320±0.721 |
| | NNV | **0.798±0.167** | **1.899±1.333** | **0.585±0.164** | **2.340±0.781** | **0.829±0.150** | **1.790±1.227** | **0.602±0.163** | **2.253±0.740** |
| MCDnCNN | HV | 0.783±0.176 | 1.965±1.392 | 0.575±0.172 | 2.416±0.745 | 0.757±0.190 | 2.067±1.419 | 0.596±0.161 | 2.326±0.699 |
| | NNV | **0.802±0.171** | **1.896±1.340** | **0.585±0.163** | **2.341±0.774** | **0.820±0.155** | **1.850±1.293** | **0.600±0.169** | **2.281±0.770** |

TABLE II
STATISTIC RESULT COMPARISON (MEAN±SD) FOR NO-NEW-NET 3D (64×64×64) MODEL USING MM-WHS DATASET. BOTH TRAINING AND TEST SET CONTAINS GAUSSIAN WHITE NOISE WITH $\sigma = 90$.

| Schemes | | Dice | Sensitivity | Specificity |
|---|---|---|---|---|
| w/o Denoise | TC | 0.817±0.089 | 0.817±0.093 | 0.997±0.001 |
| | TD | 0.827±0.065 | 0.807±0.083 | 0.998±0.000 |
| RED-CNN | HV | 0.830±0.060 | 0.820±0.065 | 0.998±0.000 |
| | NNV | **0.840±0.053** | **0.825±0.061** | 0.998±0.000 |
| MCDnCNN | HV | 0.837±0.058 | 0.825±0.064 | 0.998±0.000 |
| | NNV | **0.841±0.054** | **0.826±0.062** | 0.998±0.000 |

TABLE III
CLASSIFICATION ACCURACY COMPARISON FOR CLASSIFICATION CONVOLUTIONAL NEURAL NETWORK (CCNN), USING BRAIN TUMOR DATASET WITH THREE DIFFERENT NOISE LEVELS $\sigma = 50, 70, 90$ GAUSSIAN WHITE NOISE ADDED. NOTE THAT TRAIN SET AND TEST SET CONTAIN SAME NOISE LEVEL IN THIS EXPERIMENT.

| | w/o Denoise | | RED-CNN | |
|---|---|---|---|---|
| Cases | TC | TD | HV | NNV |
| Gaussian white noise $\sigma = 50$ | 0.370 | 0.936 | 0.946 | **0.960** |
| Gaussian white noise $\sigma = 70$ | 0.350 | 0.923 | 0.933 | **0.940** |
| Gaussian white noise $\sigma = 90$ | 0.343 | 0.890 | 0.926 | **0.936** |

had up to 2.6× higher accuracy. However, when it comes to comparison between HV and NNV schemes, all numbers outperformed schemes without denoising network. That is, denoising network is required for testing on dirty images. Based on the observation between HV and NNV scheme, NNV scheme successfully improved the accuracy up to 0.140 in all three cases.

## V. CONCLUSION

In this paper, due to the observation that neural network applications focus on different sight from human eyes, we introduced a neural-network-vision-based denoising framework. Unlike previous human-vision-based denoising methods, our framework could perform a better result for neural network application. By evaluating the experiment through different networks, noise types, and datasets on segmentation and classification, experimental results have shown the effectiveness and the feasibility of the proposed framework.

## REFERENCES


[1] X. Xu, Y. Ding, S. X. Hu, M. Niemier, J. Cong, Y. Hu, and Y. Shi, "Scaling for edge inference of deep neural networks," *Nature Electronics*, vol. 1, no. 4, pp. 216–222, 2018.

[2] Y. Ding, J. Liu, X. Xu, M. Huang, J. Zhuang, J. Xiong, and Y. Shi, "Uncertainty-aware training of neural networks for selective medical image segmentation," in *Medical Imaging with Deep Learning*, 2020.

[3] X. Xu, Q. Lu, L. Yang, S. Hu, D. Chen, Y. Hu, and Y. Shi, "Quantization of fully convolutional networks for accurate biomedical image segmentation," in *CVPR*, 2018, pp. 8300–8308.

[4] T. Wang, X. Xu, J. Xiong, Q. Jia, H. Yuan, M. Huang, J. Zhuang, and Y. Shi, "Ica-unet: Ica inspired statistical unet for real-time 3d cardiac cine mri segmentation," in *MICCAI*, 2020, pp. 447–457.

[5] X. Xu, T. Wang, Y. Shi, H. Yuan, Q. Jia, M. Huang, and J. Zhuang, "Whole heart and great vessel segmentation in congenital heart disease using deep neural networks and graph matching," in *MICCAI*, 2019, pp. 477–485.

[6] X. Xu, T. Wang, J. Zhuang, H. Yuan, M. Huang, J. Cen, Q. Jia, Y. Dong, and Y. Shi, "Imagechd: A 3d computed tomography image dataset for classification of congenital heart disease," in *MICCAI*, 2020, pp. 77–87.

[7] F. E. Boas, and D. Fleischmann, "Ct artifacts: causes and reduction techniques," *Imaging Med*, vol. 4, no. 2, pp. 229–240, 2012.

[8] K. Krupa, and M. Bekiesińska-Figatowska, "Artifacts in magnetic resonance imaging," *Polish journal of radiology*, vol. 80, p. 93, 2015.

[9] J. Liu, Y. Ding, J. Xiong, Q. Jia, M. Huang, J. Zhuang, B. Xie, C.-C. Liu, and Y. Shi, "Multi-cycle-consistent adversarial networks for ct image denoising," in *ISBI*, 2020, pp. 614–618.

[10] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "No new-net," in *International MICCAI Brainlesion Workshop*, 2018, pp. 234-244.

[11] H. H. Sultan, N. M. Salem, and W. Al-Atabany, "Multi-classification of brain tumor images using deep neural network," *IEEE Access*, vol. 7, pp. 69215–69225, 2019.

[12] Y.-J. Chen, Y.-J. Chang, S.-C. Wen, Y. Shi, X. Xu, T.-Y. Ho, Q. Jia, M. Huang, and J. Zhuang, "Zero-Shot Medical Image Artifact Reduction," in *ISBI*, 2020, pp. 862–866.

[13] H. Chen , Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose ct with a residual encoder-decoder convolutional neural network," *IEEE transactions on medical imaging*, vol. 36, no. 12, pp. 2524-2535, 2017.

[14] D. Jiang, W. Dou, L. Vosters, X. Xu, Y. Sun, and T. Tan, "Denoising of 3d magnetic resonance images with multi-channel residual learning of convolutional neural network," *Japanese journal of radiology*, vol. 36, no. 9, pp. 566-574, 2018.

[15] D. Liu, B. Wen, J. Jiao, X. Liu, Z. Wang and T. S. Huang, "Connecting Image Denoising and High-Level Vision Tasks via Deep Learning," *IEEE Transactions on Image Processing*, vol. 29, pp. 3695–3706, 2020.

[16] Z. Fan, L. Sun, X. Ding, Y. Huang, C. Cai, and J. Paisley, "A Segmentation-Aware Deep Fusion Network for Compressed Sensing MRI," in *ECCV*, 2018, pp. 55–70.

[17] X. Zhuang, "Challenges and methodologies of fully automatic whole heart segmentation: a review," *Journal of healthcare engineering*, vol. 4, no. 3, pp. 371-407, 2013.

[18] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, et al., "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993-2024, 2014.

[19] J. Cheng, "brain tumor dataset," 2016.

[20] A. Mortazi, J. Burt, and U. Bagci, "Multi-planar deep segmentation networks for cardiac substructures from mri and ct," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2017, pp. 199-206.

[21] A. Zhao, G. Balakrishnan, F. Durand, J. V. Guttag, and A. V. Dalca, "Data augmentation using learned transformations for one-shot medical image segmentation," in *CVPR*, 2019, pp. 8543-8553.