2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)
April 13-16, 2021, Nice, France

CT IMAGE DENOISING WITH ENCODER-DECODER BASED GRAPH CONVOLUTIONAL NETWORKS

Yu-Jen Chen[1] *Cheng-Yen Tsai*[1] *Xiaowei Xu*[3] *Yiyu Shi*[2]
Tsung-Yi Ho[1] *Meiping Huang*[3] *Haiyun Yuan*[3] *Jian Zhuang*[3]

[1]Department of Computer Science, National Tsing Hua University, Taiwan
[2]Department of Computer Science and Engineering, University of Notre Dame, USA
[3]Guangdong Provincial People's Hospital, Guangdong Acadamic of Medical Science, Guangzhou, China

2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI) | 978-1-6654-1246-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/ISBI48211.2021.9433900

ABSTRACT

Image denoising of low-dose CT images is a key problem in modern medical practice. Recently, several works adopted Convolutional Neural Network (CNN) to precisely capture the similarity between local features resulting in significant improvements. However, we discovered that the main drawback of existing works is the lack of non-local feature processing. On the other hand, currently, graph convolutional networks (GCN) have been widely used to process non-Euclidean geometry data considering both local and non-local features. Motivated by the property of GCN, in this paper, we propose an encoder-decoder-based graph convolutional network (ED-GCN) for CT image denoising. Particularly, we combine local convolutions and graph convolutions to process both local and non-local features. We collected seven CT volumes with Gaussian noise and Poisson noise in the experiment. Experimental results show that the proposed method outperforms existing CNN-based approaches significantly.

Index Terms— Image denoising, Deep learning, Graph convolution network

1. INTRODUCTION

Deep learning [1, 2, 3] has demonstrated its great power in medical image denoising [4, 5], a fundamental task in medical image analysis to produce clean images for clinical diagnosis, decision making, and accurate quantitative image analysis. Recent works [6, 7, 8, 9] have shown that deep-learning-based methods outperform the non-learning-based methods, such as Block Matching 3D (BM3D) [10]. Among all deep-learning-based methods, Convolutional Neural Network (CNN) is the most popular method, which still has a large room for performance improvement. Yang et al. [9] proposed to use Wasserstein distance on Generative Adversarial Network (CNN for generator) with pre-trained VGG perceptual loss (WGAN-VGG) for artifact reduction.

As noise-free images are expensive to obtain in clinical, Kang et al. [7] proposed Cycle Consistent Adversarial Denoising Network (CCADN), which tried to remove noises from two different noise-affected datasets with cycleGAN. Liao et al. [8] proposed Artifact Disentanglement Network (ADN) to extract metal artifacts from input images and reconstruct noise-free images at the same time.

We discovered that the main drawback of existing CNN-based methods is due to the lack of non-local feature processing on medical images, e.g., a single tissue will scatter throughout different regions located in an image. Though CT image denoising can be solved by CNN, non-local pixels with similar features also support the pixel recovery. As shown in Fig. 1(a), the regions marked with yellow are the same tissue as ribs. For the pixel at the center of the red circle, both the pixels marked with red and yellow can be beneficial to pixel recovery. Fig. 1(b) and (c) show the difference of the receptive field of convolution and graph convolution layer, respectively.

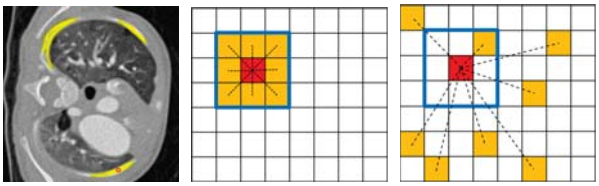
Different from CNN, Graph Convolutional Network (GCN) [11, 12] followed the pre-defined graph for aggregation. There are existing researches [13, 14] that adopted GCN on image denoising. However, none of them are used for CT images.

In this paper, we propose an encoder-decoder-based Graph Convolutional Network (ED-GCN) by utilizing the power of GCN to denoise low-dose CT images. We compare the denoising results with four widely adopted methods, including non-learning-based and learning-based methods on Gaussian noise and Poisson noise. To the best of the authors' knowledge, this is the first paper that applies Graph Convolutional Network to CT image denoising.

2. METHOD

2.1. Overall Network Architecture

As shown in Fig. 2, our network architecture has two modules, an encoder and a decoder. In the encoder, every resolution level is formed with two graph convolution layers, and one activation layer with batch normalization. Moreover, two graph convolution layers in the same resolution level share



(a) (b) (c)

Fig. 1. (a) Example of a single tissue scatters throughout regions in a CT image. The regions marked with yellow are the ribs, which are non-local and pixels marked with red are the local neighbors. Receptive field demonstration for red pixel through (b) convolution layer and (c) graph convolution layer. The blue square region represents the local neighbor of the red pixel in the center.

the same graph, which is generated by the graph construction block in real-time. A convolution layer with stride 2 downscales the image size by half. The decoder is similar to the encoder and uses only one graph convolution layer instead of two. Deconvolution layers are also included to upsample the image size into the size of the original input. Finally, the last layer is a classical 3x3 convolution layer which generates the denoised result.

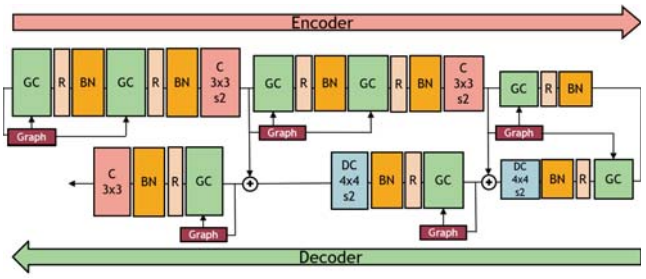


Fig. 2. The overall network architecture of the proposed encoder-decoder-based graph convolutional network (ED-GCN). (GC: graph convolution, C: 2D convolution, R: ReLU, BN: batch normalization, DC: 2D deconvolution, Graph: graph construction)

2.2. Graph Construction

We convert the unstructured image data into graphs as follows. Assume pixel p_i as an independent node and n is the number of nodes. The distance d_{ij} between node p_i and p_j in the feature space is defined by:

$$d_{ij} = \|p_i - p_j\|_2 \tag{1}$$

We then follow k-nearest neighbor (k-NN) to find the nearest k nodes in the feature domain for each node. Note that these k nodes could be either spatially surrounded or scattered for the node of interest. Finally, by connecting node p_i with its nearest k nodes (except itself), a graph $G(N, \mathcal{E})$ is obtained. N and $\mathcal{E} \subseteq N \times N$ are the number of nodes and the number of the corresponded edges of each node, respectively. We use an adjacency matrix $A \in \mathbb{R}^{n \times n}$ to represent the graph. A common adjacency matrix contains only 1 and 0, where $A_{ij} = 1$ means there is an edge between node p_i and p_j, and all edges have equal weight. In this paper, we propose a weighted adjacency matrix $A' \in \mathbb{R}^{n \times n}$ from the original adjacency matrix A with the formula:

$$A'_{ij} = \begin{cases} \frac{d_{ij}}{k}, & \text{if } A_{ij} = 1 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

All edges of the modified adjacency matrix A' are weighted to show the difference of significance of each edge.

2.3. Graph Convolution Layer

As shown in Fig. 3(a), compared with classical graph convolution layers (Fig. 1(c)), we combine local convolutions with graph convolutions (graph aggregation) together to extend the receptive field. In graph convolution layers, the graph information from graph construction blocks is required for non-local feature aggregation. The operation of graph aggregation $g : \mathbb{R}^{n \times l_1} \rightarrow \mathbb{R}^{n \times l_2}$ (l_1 and l_2 are dimensions for input and output nodes, respectively) is defined as:

$$g(I) = \sigma(A'IW + \beta), \tag{3}$$

where $A' \in \mathbb{R}^{n \times n}$ is the modified adjacency matrix, and $I \in \mathbb{R}^{n \times l_1}$ is the input feature. $W \in \mathbb{R}^{l_1 \times l_2}$ and $\beta \in \mathbb{R}^{n \times l_2}$ are the trainable weight and bias, respectively. $\sigma(\cdot)$ represents the activation function, e.g., ReLU. Eventually, the output feature of the graph convolution layer is the average of the outputs of local 3x3 convolution operation and a non-local graph aggregation operation.

An illustration of the receptive field of the proposed graph convolution layer is shown in Fig. 3(b). Local CNN can only take neighbors for feature extraction, while graph aggregation can take distant pixels into consideration. Thus, the proposed graph convolution layer combines the above two can consider both local and non-local features. In this way, the proposed layer can achieve greater diversity in the feature space as shown in Fig. 3(c), which can be potentially beneficial for image denoising.

3. EXPERIMENTS

3.1. Experimental Setup

All the experiments run on an Nvidia GTX 1080Ti GPU with 11GB memory. The proposed method, ED-GCN is implemented with the number of filters in graph convolution layers is set to 32, 64, and 128, for original image size, half size, and one-fourth size, respectively. Due to the limitation of the

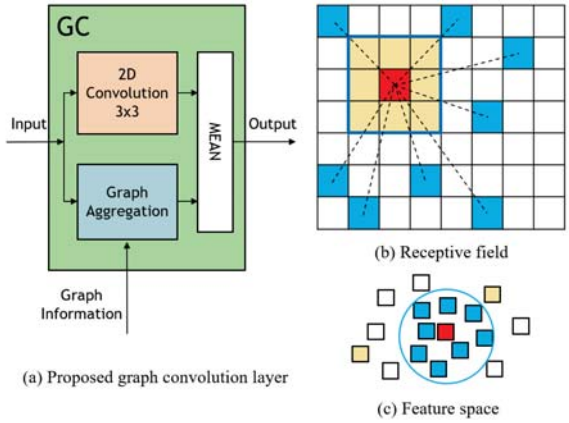


Fig. 3. An illustration of (a) the proposed graph convolution layer and (b) the corresponding receptive field in the 2D image and (c) the feature space. The pixels marked with red, brown, and blue are the target pixel, local neighbors, and the non-local neighbors in spatial space.

memory, our model is trained with patches using the size of 32×32 and batch size at 1. Mean Square Error (MSE) loss function is minimized by Adam optimizer [15] with a learning rate of 0.0001 and the number of epochs is 30. Hyperparameter k in Section 2.2 is set to 8.

We adopted structural similarity index (SSIM), root mean square error (RMSE), and peak signal-to-noise ratio (PSNR) for evaluation [6, 9]. Note that for SSIM and PSNR, a larger value represents a better result. On the contrary, for RMSE, a smaller value is desired.

3.2. Dataset

For evaluation, we collected seven 3D cardiac CT volumes including 2,105 gray-scale 2D images, which were then split into training and test set. Training set contained five 3D CT volumes, we randomly cropped 100,000 patches of size 32×32 from original images of size 512×512. For test set, the remaining two 3D CT volumes, which contained 424 CT images with size 512×512 were applied to examine the proposed method. In our experiment, similar to [6, 16], either Gaussian white noise with $\sigma = 70$ or Poisson noise was superposed to form paired noise-affected and noise-free images for both training and test set.

3.3. Post-processing

In our experiment, we divided the image into patches for image denoising. We then concatenated the result back to the original size. However, sometimes the adjacent patches showed subtle color inconsistencies. To alleviate such effect, we introduced Poisson image editing [17] as our post-processor.

Poisson image editing is an interpolation machinery based on solving Poisson equations, which is used to blend images with different image gradients. We denote S, a closed subset of $\mathbb{R}^2$, as the image domain, Ω as a closed subset of S with boundary $\partial\Omega$. f^* is the destination function over $S - \Omega$, and f as an unknown function over Ω. Given a vector field v defined over area Ω, which is the gradient of each pixel in Ω, describes the change of brightness of the source image. Then the process of interpolation can be formulated as an optimization problem below:

$$\min_f \iint_\Omega |\nabla f - v|^2 \text{ with } f|_{\partial\Omega} = f^*|_{\partial\Omega}, \tag{4}$$

which means we want to find an f that has a similar gradient with the vector field v, while setting the pixels in $\partial\Omega$ exactly the same as the destination image.

In [17], Poisson image editing is capable to integrate tileable rectangular images by setting periodic boundary conditions. In this paper, we modify the boundary conditions to $p^*_{\text{top}} = 0.5(p_{\text{top}} + g_{\text{bottom}})$, where p is the target patch to be interpolated, and g is the top-adjacent patch (similarly for the right, left, and bottom borders). Also, for corners of patches, we compute the pixels with $p^*_{\text{top-right}} = 0.25(2 \times p_{\text{top-right}} + g_{\text{top-left}} + h_{\text{bottom-right}})$, where p is the target patch to be interpolated, g is the right-adjacent patch, and h is the top-adjacent patch (similarly for the other corners). By iterating these conditions, each patch will blend with neighboring patches.

4. RESULTS

We compared the proposed ED-GCN with four existing methods: BM3D [10], WGAN-VGG [9], CCADN [7], and ADN [8] for evaluation. To compare with the state-of-the-art, we selected two CT images in the test set for both qualitative and quantitative comparison.

4.1. Quantitative Results

Quantitative comparison in mean and standard deviation for Gaussian noise is shown in Table 1. BM3D can achieve about 19.2% improvement on SSIM but only 3.2% and 1.4% on RMSE and PSNR. The other three methods, WGAN-VGG, CCADN, and ADN can achieve significant improvement over BM3D with over 50% on all three evaluation metrics. The proposed method, ED-GCN achieved optimal performance in all three metrics which yielded an improvement of 2.8%, 4.5%, and 11.9% on SSIM, RMSE, and PSNR, respectively over the optimal ones of the other four works. In addition, ED-GCN also achieved the smallest standard deviations on SSIM. Though WGAN-VGG has the smallest standard deviation on RMSE, its performance on RMSE is much lower than ED-GCN. Standard deviations of ED-GCN on all metrics were only 24%-66% of that of CCADN and ADN, which

Table 1. Quantitative comparison between ED-GCN and four existing methods in mean and standard deviation for Gaussian white noise with $\sigma = 70$.

Case	Overall (mean $\pm$ SD)			Improve		
	SSIM	RMSE	PSNR	SSIM	RMSE	PSNR
Input	0.510 ± 0.128	26.140 ± 5.571	19.992 ± 1.926	-	-	-
BM3D [10]	0.608 ± 0.070	25.293 ± 5.308	20.270 ± 1.882	19.2%	3.2%	1.4%
WGAN-VGG [9]	0.843 ± 0.048	5.269 ± 1.346	33.963 ± 2.130	65.3%	79.8%	69.9%
CCADN [7]	0.915 ± 0.018	6.988 ± 3.368	32.232 ± 4.163	79.6%	73.3%	61.2%
ADN [8]	0.897 ± 0.046	6.565 ± 4.483	33.261 ± 4.748	76.0%	74.9%	66.4%
ED-GCN	$\mathbf{0.930 \pm 0.015}$	$\mathbf{4.094 \pm 1.437}$	$\mathbf{36.342 \pm 2.710}$	**82.4%**	**84.3%**	**81.8%**

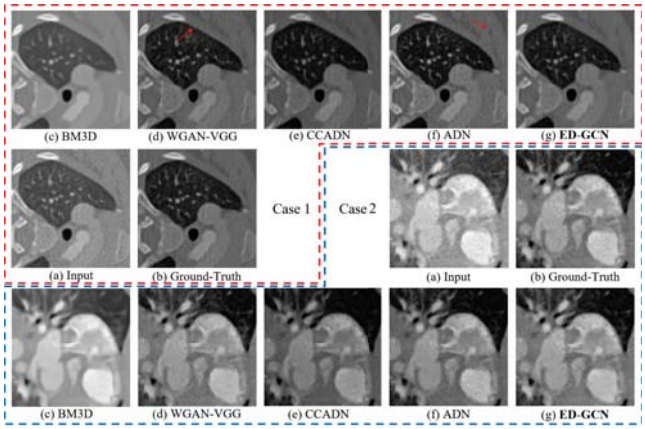


Fig. 4. Qualitative comparison using CT image on Gaussian white noise with $\sigma = 70$.

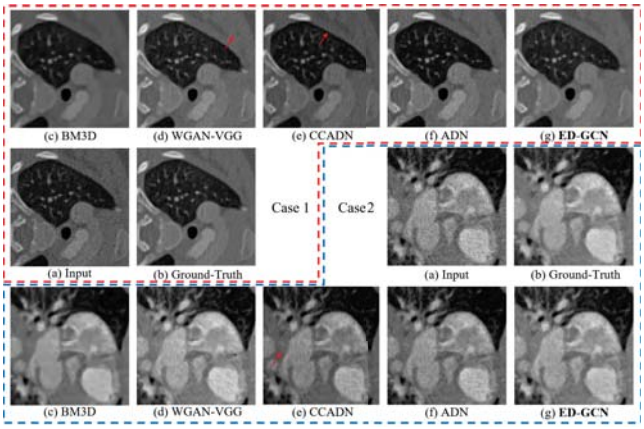


Fig. 5. Qualitative comparison using CT image on Poisson noise.

shows that ED-GCN is much more stable than existing methods.

The improvement trend of all the methods on Poisson noise is the same as that on Gaussian noise. Compared with Gaussian noise, the overall improvement of all methods on Poisson noise is relatively small. Among all the five approaches, BM3D achieved about 10.6% improvement on SSIM and only 6.6% and 2.7% on RMSE and PSNR, respectively. WGAN-VGG, CCADN, and ADN achieved improvements of over 4.5% on all the three metrics over BM3D. ED-GCN obtained the best performance with an improvement of 1.9%, 2.9%, and 3.0% on SSIM, RMSE, and PSNR, respectively over the optimal scores of existing methods.

4.2. Qualitative Results

Qualitative comparison for Gaussian noise and Poisson noise are shown in Fig. 4 and Fig. 5, respectively. Some noisy regions are further highlighted with red arrows for discussion. For all the four cases in both noises, we can notice that the results of BM3D are oversmoothed, and the edges of the tissue become blurry. There exist some slight noises in the results of WGAN-VGG and ADN, especially in case 1 of Fig. 4. For the case 1 in Fig. 5, WGAN-VGG oversmooths a tiny tissue, which should not be considered as noise. The results of CCADN are rather shadowy, which might be due to biased training. Moreover, regions marked with a red arrow in both cases of Fig. 5 for CCADN still contain slight noise. For both cases in both types of noise, ADN and ED-GCN achieved almost the same performance which is comparable with the ground-truth as shown in both Fig. 4 and Fig. 5. This is due to the fact that their scores on all three metrics are close to each other in the quantitative results.

5. CONCLUSION

In this paper, we proposed a novel encoder-decoder-based graph convolutional network for CT image denoising by combining CNN and GCN to consider both local and non-local features thus enhancing the performance. We collected seven CT images for evaluation and discussed both Gaussian and Poisson noises in the experiment by comparing our method with four popular existing methods in both quantitative and qualitative results. The results showed that our method outperforms existing methods by a large margin.

6. ACKNOWLEDGMENTS

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

7. REFERENCES

[1] Yukun Ding, Jinglan Liu, Xiaowei Xu, Meiping Huang, Jian Zhuang, Jinjun Xiong, and Yiyu Shi, "Uncertainty-aware training of neural networks for selective medical image segmentation," in *Medical Imaging with Deep Learning*, 2020.

[2] Xiaowei Xu, Qing Lu, Lin Yang, Sharon Hu, Danny Chen, Yu Hu, and Yiyu Shi, "Quantization of fully convolutional networks for accurate biomedical image segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8300–8308.

[3] Xiaowei Xu, Tianchen Wang, Jian Zhuang, Haiyun Yuan, Meiping Huang, Jianzheng Cen, Qianjun Jia, Yuhao Dong, and Yiyu Shi, "Imagechd: A 3d computed tomography image dataset for classification of congenital heart disease," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 77–87.

[4] Yu-Jen Chen, Yen-Jung Chang, Shao-Cheng Wen, Yiyu Shi, Xiaowei Xu, Tsung-Yi Ho, Qianjun Jia, Meiping Huang, and Jian Zhuang, "Zero-shot medical image artifact reduction," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 862–866.

[5] Shao-Cheng Wen, Yu-Jen Chen, Zihao Liu, Wujie Wen, Xiaowei Xu, Yiyu Shi, Tsung-Yi Ho, Qianjun Jia, Meiping Huang, and Jian Zhuang, "Do noises bother human and neural networks in the same way? a medical image analysis perspective," *arXiv preprint arXiv:2011.02155*, 2020.

[6] Hu Chen, Yi Zhang, Mannudeep K Kalra, Feng Lin, Yang Chen, Peixi Liao, Jiliu Zhou, and Ge 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.

[7] Eunhee Kang, Hyun Jung Koo, Dong Hyun Yang, Joon Bum Seo, and Jong Chul Ye, "Cycle consistent adversarial denoising network for multiphase coronary ct angiography," *arXiv preprint arXiv:1806.09748*, 2018.

[8] Haofu Liao, Wei-An Lin, S Kevin Zhou, and Jiebo Luo, "Artifact disentanglement network for unsupervised metal artifact reduction," *arXiv preprint arXiv:1906.01806*, 2019.

[9] Qingsong Yang, Pingkun Yan, Yanbo Zhang, Hengyong Yu, Yongyi Shi, Xuanqin Mou, Mannudeep K Kalra, Yi Zhang, Ling Sun, and Ge Wang, "Low dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE transactions on medical imaging*, 2018.

[10] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian, "Image denoising with block-matching and 3d filtering," in *SPIE Electronic Imaging*. International Society for Optics and Photonics, 2006, vol. 6064, p. 606414.

[11] Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodola, Jan Svoboda, and Michael M Bronstein, "Geometric deep learning on graphs and manifolds using mixture model cnns," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5115–5124.

[12] Thomas N Kipf and Max Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[13] Stamatios Lefkimmiatis, "Non-local color image denoising with convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3587–3596.

[14] Diego Valsesia, Giulia Fracastoro, and Enrico Magli, "Image denoising with graph-convolutional neural networks," in *IEEE International Conference on Image Processing*, 2019, pp. 2399–2403.

[15] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[16] Jelmer M Wolterink, Tim Leiner, Max A Viergever, and Ivana Išgum, "Generative adversarial networks for noise reduction in low-dose ct," *IEEE transactions on medical imaging*, vol. 36, no. 12, pp. 2536–2545, 2017.

[17] Patrick Pérez, Michel Gangnet, and Andrew Blake, "Poisson image editing," in *ACM SIGGRAPH 2003 Papers*, pp. 313–318. 2003.