JOINT LOW DOSE CT DENOISING AND KIDNEY SEGMENTATION

Mohammad Eslami, Solale Tabarestani, Malek Adjouadi

Center for Advanced Technology and Education (CATE),
Department of Electrical and Computer Engineering
Florida International University, Miami, USA.
https://cate.fiu.edu/ e-mail: meslami@fiu.edu

ABSTRACT

In this research, both image denoising and kidney segmentation tasks are addressed jointly via one multitask deep convolutional network. This multitasking scheme yields better results for both tasks compared to separate single-task methods. Also, to the best of our knowledge, this is a first time attempt at addressing these joint tasks in low-dose CT scans (LDCT). This new network is a conditional generative adversarial network (C-GAN) and is an extension of the *image-to-image* translation network. To investigate the generalized nature of the network, two other conventional single task networks are also exploited, including the well-known 2D *UNet* method for segmentation and the more recently proposed method *WGAN* for LDCT denoising. Implementation results proved that the proposed method outperforms *UNet* and *WGAN* for both tasks.

Index Terms— Low Dose CT, Denoising, Segmentation, Deep Convolutional Network, Conditional GAN, image-to-images translation.

1. INTRODUCTION

Chronic kidney disease (CKD) affects over 37 American adults and leads to serious health complications, making it a heavy burden to the health system here in the US and across the globe [1]. Computerized Tomography (CT) is one of the most common imaging modalities in evaluating kidney problems due to its non-invasive nature, high diagnostic accuracy, specificity and sensitivity. Low Dose Computed Tomography (LDCT) is an imaging modality that gained greater attention recently because of the lower dose of radiation it necessitates. With its wider use, cost-effectiveness and faster scanning time, LDCT has become most suitable for screening, diagnosis and follow up studies [2].

Kidney segmentation is an initial preprocessing step that is of great importance in medical diagnosis, and for treatment planning and image-guided interventions. A potential for accurate and robust segmentation algorithms could reside in the use of computer assisted technologies [3]. Image segmentation remains a subtle task prone to error due to the ubiquitous presence of noise and the subjective use of thresholds that define the properties of contours and boundaries of objects in an image. While segmentation is in itself a complex problem in imaging, image enhancement in LDCT is also a contentious issue [4]. Moreover, the convenience of the low dose radiation exposure makes the delineation and assessment of organs such as kidney, ureter, and bladder non-trivial [5].

Recently, deep learning methods , mostly relying on Convolutional Neural Networks (CNN), have gained resurgence in medical image analysis to involve the important tasks of segmentation, denoising and etc [6, 3, 7, 4, 8, 9]. Unlike image denoising which has been well reported in the literature [4], it is worth mentioning that kidney segmentation on LDCT has not been extensively studied [7]. In this paper, and for the first time, we consider both tasks of image segmentation and image enhancement on LDCT slices jointly. For this reason, the developed method named *image-to-images translation* is exploited, proving that multitask learning augments the performance of the system by considering the interplay between tasks that are inherently inter-related [10]. Previously, the proposed network has been successfully exploited for bone suppression and segmentation in chest x-ray images, and related source codes has been publicly shared online on GitHub by us (https://github.com/mohaEs/image-to-images-translation).

The paper is organized as follows: Section 2 introduces the methodology and the data used for this study. The experimental results are reported and discussed in section 3. Finally, the paper provides concluding remarks on the deployment of this new multitask method in 4.

2. DATA AND METHODOLOGY

The proposed model is an extension of the *pix2pix* network. This modified conditional generative adversarial network *(C-GAN)* extends the *image-to-image* network into *image-to-images* architecture in order to perform multitask learning. In this research, the two tasks that will be performed simultane-

This research is supported by the National Science Foundation (NSF) under grants CNS-1920182, CNS-1532061, CNS-1338922 and CNS-1551221.

978-1-7281-7401-3/20/$31.00 ©2020 IEEE

ously are image denoising/enhancement and kidney segmentation in low-dose CT imaging modality. Other application models of this architecture such as neuroimaging translation for cross-modality generation and chest x-ray image analysis has been introduced in [8]. The proposed architecture contains a generator and a discriminator. The discriminator is a CNN network which aim to assess the meaningfulness of its input tensors. The discriminator network produces two output matrices, D_R and D_F corresponding to the real and fake input tensors (R and F). The fake input tensor (F) is a concatenation of input image and outputs ($F : X \parallel \hat{Y}_1 \parallel \hat{Y}_2$), and the real input tensor (R) is the concatenation of the input image and targets ($R : X \parallel Y_1 \parallel Y_2$). The generator network should produce outputs $\hat{Y}_1$ and $\hat{Y}_2$ similar to targets Y_1 and Y_2. The generator consists of 8 layers with skip connections, in which dilation convolutions with rate of 2 has been employed in layers 2 through 7 to generate more efficient receptive fields. The loss function for training the generator and discriminator are $\mathcal{L}_G = \mathbb{E}[-log(D_F + \epsilon)] + \lambda\, \mathbb{E}[\mid Y - \hat{Y} \mid_1]$ and $\mathcal{L}_D = \mathbb{E}[-(\, log(D_R + \epsilon) + log(1 - D_F + \epsilon)\,)]$, respectively where $|\ |_1$ represents the $L1$ distance and $\mathbb{E}$ is the expectation measure. More details can be found in [8].

The dataset used in this study is from *'Multi-Atlas labeling beyond the cranial vault'* challenge, containing CT scans and segmentation labels of 13 abdominal organs (including kidneys) of 50 subjects [11]. Labels of 30 subjects, which are shared publicly, are used in this paper. To simulate the LDCT scans from CT scans, the method based on additive Poisson noise on sonograms of CT scans is used [12].

3. SIMULATION AND RESULTS

Since our method is based on 2D slices and 2D convolutional network, 2D *UNet* [9] is used as the competitor for segmentation task. The GAN based method with *Wasserstein* distance (*WGAN-MSE*) proposed in [4] is used as the competitor for the denoising task which is also based on 2D slice processing. The original *pix2pix* single task method is referred to as *p2p ST* and the proposed extension method for multitasking is referred to as *p2p MT*. The 7-fold cross validation on subjects is used in this experiment for evaluation purposes. Indeed for each fold, all the axial slices of 4 subjects are addressed as test images and all the rest axial slices of 26 subjects are used for training. All the hyper parameters and implementations are the same as in [8] (*https://github.com/mohaEs/image-to-images-translation*), [9] and [4] (*https://github.com/yyqqss09/ldct_denoising*). The batch size for all experiments are 20 slices. For evaluating the results, root mean square error (RMSE) and structural similarity index (SSIM) are considered for the denoising task, and Dice score and false positive rate (FPR) are used for the segmentation task [8].

The first interesting observation of this experiment is in the convergence ability of the different networks. While the denoising problem is manageable and all the methods for all folds converged properly, the segmentation problem is found to be very challenging and makes the network converge to improper points. Table 1 shows the number of successful folds of different methods for both tasks (Seg. and Den.).

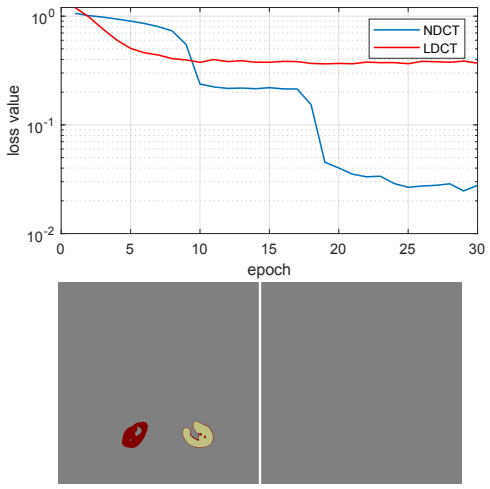

Fig. 1. Top) Loss curves of training by *UNet* architecture on NDCT and LDCT sets for the same fold. Bottom) Segmentation result by *UNet* on NDCT (left) and LDCT (right) for the same test slice.

Table 1. The number of successful converged training fold by different methods of both tasks among all 7 folds.

LDCT Seg. ST	LDCT Den. ST	LDCT Seg.+Den. MT	LDCT Seg. UNet	NDCT Seg. UNet	LDCT Den. WGAN
3	7	5	0	7	7

For the segmentation task, while 2D *UNet* was successful for all folds of training and segmenting on nomral-dose CT (NDCT) set, it was completely unsuccessful for training on LDCT set and was often stuck on local minimum points. As far as authors know, there is no reported literature about using 2D *UNet* on LDCT. Figure 1 (top) shows the loss curves of two same folds of *UNet* training on NDCT and LDCT. As shown in Fig. 1 (bottom), the training on LDCT converged to background class only, while the NDCT result is great. It is worth mentioning that, in addition to the noisy images of LDCT, the target classes of (background, left and right kidney) are highly unbalanced, which makes the problem even more challenging. *UNet* architecture with 3 more convolutional layers was also tested to make sure about the proper size of the receptive fields and verify the convergence issue.

As shown in Table 1, for the segmentation task, the *p2p ST* and proposed *p2p MT* both converged properly using 3 folds and 5 folds, respectively. This demonstrates that *p2p ST* is better than *UNet* because of its conditional generative adversarial nature which makes the network more amenable to

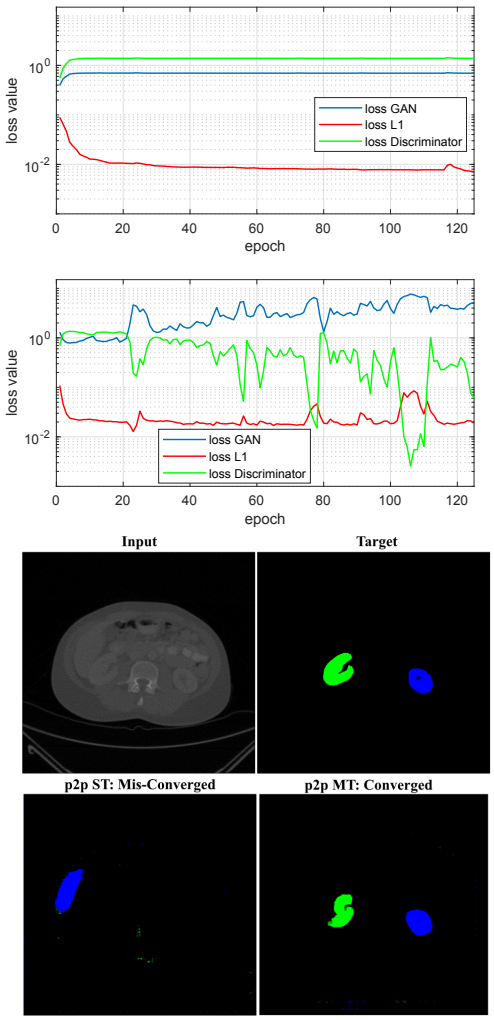

Fig. 2. Top) Loss curves of a successful training for the *p2p* GAN architecture on a LDCT fold. Middle) Same as Top subfigure but an unsuccessful fold. Bottom) Segmentation result on a LDCT test slice with successful and unsuccessful trained *p2p*-based networks.

overcome local minima. On the other hand, *p2p MT* is better than *p2p ST* because the second task (denoising) is augmenting the whole system through the regularization mechanism and in finding a solution on a smaller area of representations at the intersection of both tasks. Notice that, if the *p2p*-based method is not successful in converging, it means that the discriminator wins the challenge and the generator can not produce enough accurate outputs to fool the discriminator. Figure 2 shows the loss curves of training for a converged fold (top: loss of GAN (generator) smaller than loss of discriminator) and non-converged fold (middle) via *p2p*-based method. The figure also shows an example of the achieved segmentation results (bottom).

In addition to increasing the chance for convergence via multitasking (*p2p MT*), the feature selection and filter values could be reassessed to improve the understanding of the system from the nature of the inputs that improve the accuracy of the results. Figure 3 shows the qualitative and quantitative results for one test slice for both tasks, comparing the original *p2p ST* and proposed multitask network *p2p MT*. For both tasks, the multitask method outperforms the single-task method while having only half of the network weights. Figure 4 shows the qualitative results of different methods for a test slice and for denoising task only to demonstrate the superiority of the multitask method *p2p MT*. The magnified area is shown to emphasize noticeable differences. The overall achieved results by different methods are summarized in Tables 2 and 3 demonstrating the efficiency of *p2p MT*. To make a fair comparison, the folds considered are those where all the methods proved successful during training.

Table 2. Average Dice and FPR results on 3 overlapped successful folds for segmentation task.

	p2p ST	*UNet* (LDCT)	*UNet* (NDCT)	*p2p MT*
Dice	72.5 ± 8.0	0.0 ± 0.0	82.4 ± 6.8	79.4 ± 3.1
FPR	0.36 ± 0.07	0.0 ± 0.0	0.04 ± 0.09	0.24 ± 0.04

Table 3. Average RMSE and SSIM results on 5 overlapped successful folds for denoising task.

	p2p ST	*WGAN-MSE*	*p2p MT*
RMSE	0.94 ± 0.4	0.72 ± 0.03	0.69 ± 0.17
SSIM	0.74 ± 0.24	0.81 ± 0.09	0.84 ± 0.07

4. CONCLUSIONS

In this study, we presented a multitask approach based on *pix2pix MT* to perform simultaneously the two tasks of image enhancement and kidney segmentation in low-dose CT imaging. We investigated the performance of the proposed architecture through 7-fold cross validation on subjects from the 'Multi-Atlas labeling beyond the cranial vault' dataset challenge. Experimental results proved the efficiency of the multitask learning approach in comparison to single task learning in terms of higher chance for convergence, higher accuracy and lower network parameters which also means faster processing time. These results have been compared to related state-of-the art methods relevant to this research endeavor. In the future, the authors intend to extend the exploited 2D CNNs in the proposed method to 3D CNNs, which are expected to bring improvements because of the 3D nature of the organs in CT imaging modality.

5. REFERENCES

[1] Boris Bikbov and et al., "Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the global burden of disease study

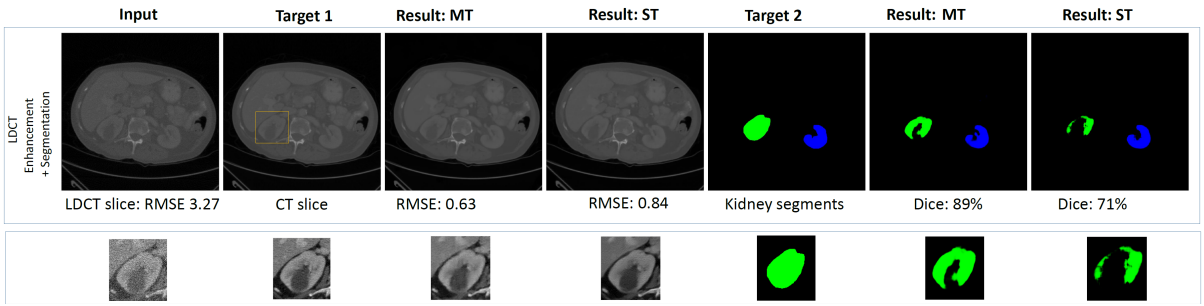

Fig. 3. Illustration of the results for two tasks by *p2p ST* and *MT* for a LDCT test slice.

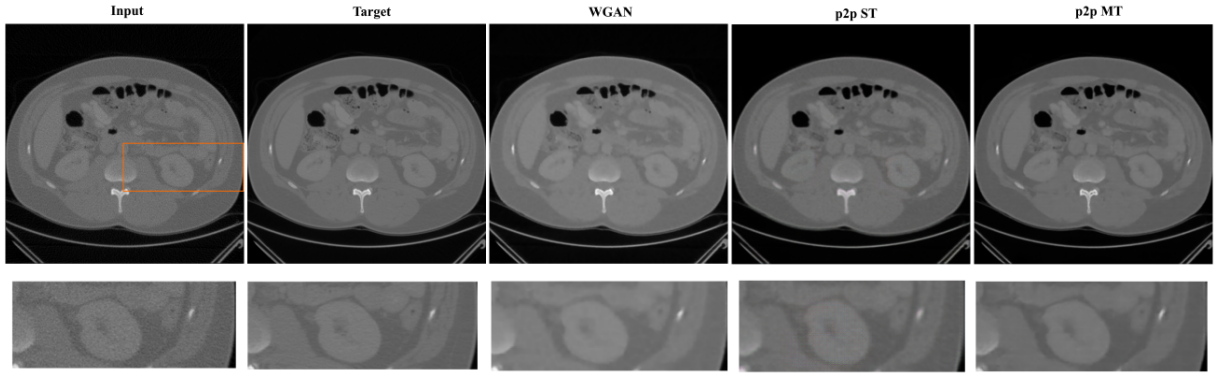

Fig. 4. Illustration of the results for denoising task by different methods for a LDCT test slice.

2017," *The Lancet*, vol. 395, no. 10225, pp. 709–733, 2020.

[2] Diego Ardila and et al., "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," *Nature medicine*, vol. 25, no. 6, pp. 954, 2019.

[3] Kai-jian Xia and et al., "Deep semantic segmentation of kidney and space-occupying lesion area based on scnn and resnet models combined with sift-flow algorithm," *Journal of medical systems*, vol. 43, no. 1, pp. 2, 2019.

[4] Qingsong Yang and et al., "Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1348–1357, 2018.

[5] Joseph Kuebker and et al., "Radiation from kidney-ureter-bladder radiographs is not trivial," *Urology*, vol. 125, pp. 46–49, 2019.

[6] Mohammad Eslami and et al., "Automatic vocal tract landmark localization from midsagittal mri data," *Scientific Reports*, vol. 10, no. 1, pp. 1–13, 2020.

[7] Price Jackson and et al., "Deep learning renal segmentation for fully automated radiation dose estimation in unsealed source therapy," *Frontiers in oncology*, vol. 8, pp. 215, 2018.

[8] Mohammad Eslami and et al., "Image-to-images translation for multi-task organ segmentation and bone suppression in chest x-ray radiography," *IEEE Transactions on Medical Imaging*, 2020, doi:10.1109/TMI.2020.2974159.

[9] Olaf Ronneberger and et al., "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[10] Solale Tabarestani and et al., "A distributed multitask multimodal approach for the prediction of alzheimer's disease in a longitudinal study," *NeuroImage*, vol. 206, pp. 116317, 2020.

[11] "Multi-atlas labeling beyond the cranial vault – workshop and challenge at miccai 2015," https://www.synapse.org/#!Synapse:syn3193805/wiki/217752, Accessed: 2019-12-30.

[12] Hu Chen and et al., "Low-dose ct with a residual encoder-decoder convolutional neural network," *IEEE transactions on medical imaging*, vol. 36, no. 12, pp. 2524–2535, 2017.