
# Ultra-DenseNet for Low-Dose X-Ray Image Denoising in Cardiac Catheter-Based Procedures

Yimin Luo[1(✉)], Daniel Toth[1,2], Kui Jiang[3], Kuberan Pushparajah[1], and Kawal Rhode[1]

[1] School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK
yimin.luo@kcl.ac.uk
[2] Siemens Healthineers, Frimley, UK
[3] School of Computer Science, Wuhan University, Wuhan, China

**Abstract.** The continuous development and prolonged use of X-ray fluoroscopic imaging in cardiac catheter-based procedures is associated with increasing radiation dose to both patients and clinicians. Reducing the radiation dose leads to increased image noise and artifacts, which may reduce discernable image information. Therefore, advanced denoising methods for low-dose X-ray images are needed to improve safety and reliability. Previous X-ray imaging denoising methods mainly rely on domain filtration and iterative reconstruction algorithms and some remaining artifacts still appear in the denoised X-ray images. Inspired by recent achievements of convolutional neural networks (CNNs) on feature representation in the medical image analysis field, this paper introduces an ultra-dense denoising network (UDDN) within the CNN framework for X-ray image denoising in cardiac catheter-based procedures. After patch-based iterative training, the proposed UDDN achieves a competitive performance in both simulated and clinical cases by achieving higher peak signal-to-noise ratio (PSNR) and signal-to-noise ratio (SNR) when compared to previous CNN architectures.

## 1 Introduction

Image-guided interventions which require navigating medical therapeutic devices through a patient's cardiovascular system using X-ray imaging have seen growing use [1]. Such procedures play an important role in cardiac catheter-based intervention, a type of minimally invasive surgery for treating cardiovascular diseases, such as arrhythmias and stenoses. Despite a much quicker recovery and less postoperative discomfort for the patient, these procedures can result in significant X-ray exposure to the patient as well as to the medical staff, which is a major concern especially in paediatric patients. Given the potential risk of X-ray radiation, low-dose X-ray fluoroscopic imaging is of great value. The most common way to lower the radiation dose is to reduce the X-ray flux by decreasing the tube operating current, shortening the exposure time and decreasing the frame rate. However, this increases the noise and artifacts in the obtained images, which can reduce discernible information during the


M. Pop et al. (Eds.): STACOM 2019, LNCS 12009, pp. 31–42, 2020.
https://doi.org/10.1007/978-3-030-39074-7_4

procedures. Therefore, it is worthwhile to develop new efficient denoising methodologies that allow significant X-ray dose reduction without loss of discernible information.

Traditional image denoising approaches mainly utilise non-local statistics and image self-similarity [4–6], and they usually suffer from complex parameter selection. To overcome this drawback, some discriminative learning methods have been developed to learn image prior models in the context of truncated inference [7]. However, they train a specific model for a certain noise level, which results in poor performance when facing complex noise. Inspired by recent success of deep convolutional neural networks (CNNs) on feature representation, several CNN-based denoising methods in various fields [3, 4, 7–12] have been proposed to achieve a better image restoration through a deep end-to-end mapping between low- and high-quality images. Unlike general images, low-dose X-ray images used in cardiac catheter-based procedures suffer from quantum noise which is often modelled by a Poisson law according to the physics of X-ray generation and imaging. This that means homogeneous regions can appear highly noisy and the contrast of heterogenous regions can be extremely low, which results in complexity of denoising. Considering the strengths of CNN frameworks in recovering high-frequency details caused by complex noise, this paper develops this framework for cardiac catheter-based procedures as an X-ray image denoiser. Recently, to improve the efficiency and effectiveness of feature extraction, CNN frameworks witnessed a rapid development on their architecture [13, 14]. However, deeper architectures are particular demanding especially in the case of X-ray images which typically have a large matrix size and frame rate. To improve feature propagation and reuse in classification tasks, Huang et al. [16] connected each layer to every other layer in a feed-forward manner and proposed dense convolutional network (DenseNet). As this new architecture provides a more tenacious way to combine the low- and high-level features, it visibly outperforms other CNN-based methods. In addition, dense skip connection can also alleviate the vanishing-gradient problem in training, as it enables short paths linking directly to every layer output. Despite the increasing utilization of information, this all-round connection increases computation burden and memory consumption to a large degree.

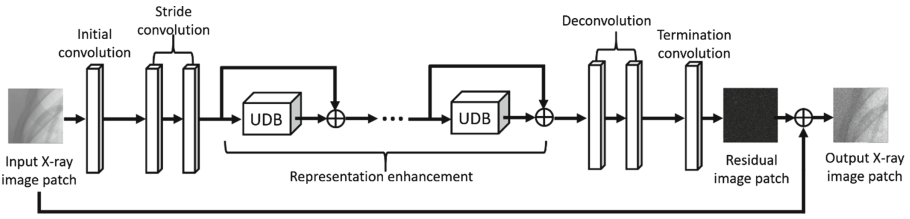


**Fig. 1.** Outline of the proposed ultra-dense denoising network (UDDN).

In this study, based on dense skip connection, a novel ultra-dense denoising network is proposed to achieve X-ray image denoising in cardiac catheter-based procedures with less memory consumption. This network effectively improves feature extraction by

establishing rich correlation between multiple-path neural units in each residual block and a better mapping between low-dose X-ray images and their full-dose ground truth can be searched. As illustrated in Fig. 1, this network contains several ultra-dense blocks (UDBs). As each UDB has three conventional dense blocks given the same convolution layers, it can greatly enhance the representational power of the network. Since the parameters between UDBs are shared with each other, this network can also release the memory burden to a large degree. Trained with high- and low-dose X-ray image pairs, a model specified in X-ray image denoising can be obtained. Experiments on both the simulated and clinical datasets validate the effectiveness of this network.

## 2 Methodology

As illustrated in Fig. 2, the ultimate goal is to learn a non-linear mapping function which can reconstruct the corresponding high-dose X-ray image from the given low-dose input. Therefore, to obtain a robust model specified in X-ray image denoising, low-dose X-ray images and their high-dose counterparts are required as samples and labels for training respectively. Firstly, to implement the training process, noise must be added to original high-dose X-ray images artificially to simulate the low-dose X-ray samples. Trained with those image pairs, a robust CNN model specified in X-ray image denoising can be obtained by iteratively minimizing the difference between the predicted high-dose X-ray image and the ground truth. Finally, with a low-dose X-ray image and the obtained network, its high-lose denoising result can be directly predicted.

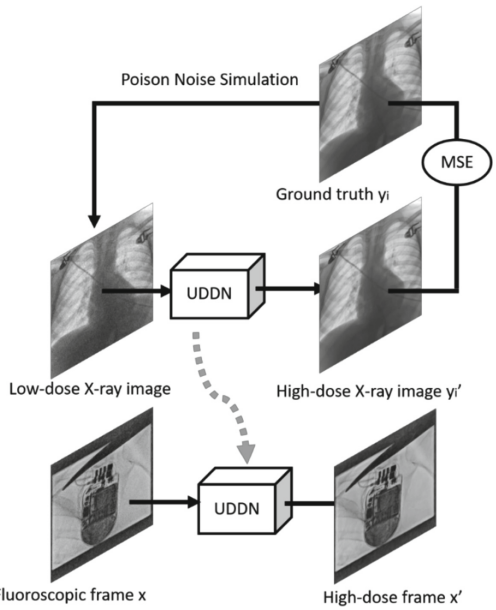


**Fig. 2.** Flowchart for X-ray image denoising via the proposed UDDN. The flowchart can be divided into two parts: training (upper) and denoising (lower).

### 2.1 Poison Noise Simulation

For X-ray imaging, quantum noise is the most dominant source of noise and is usually modelled by a Poisson law, hence, a Poisson model for the noise measurement is calculated by (1) and (2)

$$\mathbf{Noise} = \mathbf{Poisson}(\lambda) + \lambda \tag{1}$$

$$\lambda = \sigma\mu \tag{2}$$

Where σ denotes the noise level in the X-ray image and μ denotes the mean image intensity. The network is trained from normal-dose X-ray images and their corresponding low-dose images which were generated by adding Poisson noise according to this physical model. Figure 3 makes a comparison of a high-dose X-ray image and its simulated noisy counterpart with 60% poison noise which is approximately equivalent to quarter-dose acquisition.

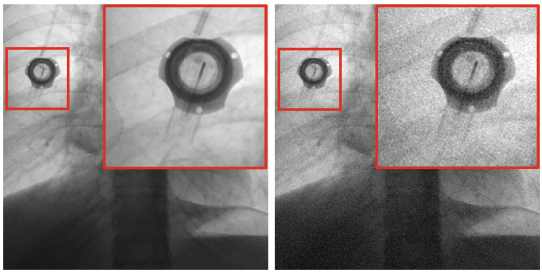

**Fig. 3.** An example of adding simulated noise to an X-ray image (Left: Image acquired at a high dose. Right: Image with 60% simulated added noise)

### 2.2 Network Architecture

In this section, we present the design of each key module under the UDDN framework in detail. As illustrated in Fig. 1, the first convolution layer is an initial layer for shallow feature extraction of the input X-ray image patches. Then, two stride convolution layers are added to map the extracted features to low-dimensional domain and this operation aims at reducing the amount of calculation. The main part of our network is stacked with multiple residual blocks and this design enhances the representation of the obtained low-dimensional features to a large degree. As shown in Fig. 4, on the basis of the dense connection, we propose a triple-path residual block called UDB, the black lines in it represent flat and common used skip connections and the blue lines represent cross connections between paths which enable sharing of information. Compared to previous dense networks, UDB contains three times as many richer paths with the same convolution layers. For the sake of these triple-path units and transition layer, the feature channels become shallower and the parameters become less, which visibly decreases the computational burden and memory consumption. In particular, the

UDBs in our network interact with each other by skip connection, so they can utilize the feature information of their predecessor, which facilitates the reuse of features. Subsequently, two deconvolution layers are added to map the obtained features back to high-dimensional domain. Finally, other convolution layers are added in our network as a termination layer to output a residual noisy image which is as the same size as the input one.

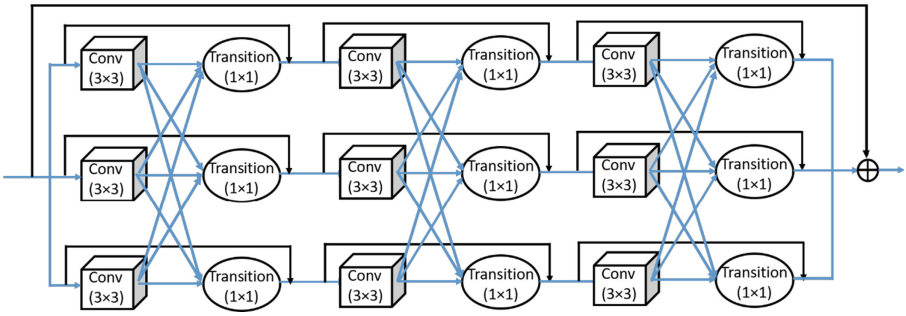


**Fig. 4.** Outline of the proposed ultra-dense blocks (UDBs) in UDDN. (Color figure online)

### 2.3 Loss Function

According to various CNN-based image processing methods [6, 7, 9–14], the loss function is commonly used to fit the target image by minimizing the distance between the output image and the ground truth based on feature level and Euclidean distance and cosine distance are the most commonly used similarity measurements. In terms of denoising, to obtain a substantially improved CNN architecture, as most of the previous methods [13–15] constrain output by iteratively minimizing the mean squared error (MSE), we use this measurement during our network training. As shown in Fig. 2, for a certain low-dose X-ray image, the MSE of its real high-dose X-ray image $y_i$ and predicted high-dose X-ray $y'_i$ is calculated by (3)

$$\mathbf{MSE_i} = \frac{\mathbf{1}}{\mathbf{width} \times \mathbf{height}} \left|\left|\mathbf{y_i} - \mathbf{y'_i}\right|\right|^2 \tag{3}$$

For the whole training datasets, the loss function of our network can be calculated by (4)

$$\mathbf{Loss} = \frac{\mathbf{1}}{\mathbf{2N}} \sum_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{MSE_i} \tag{4}$$

Where *N* represents the number of X-ray image samples for training.

### 2.4 Quantitative Indicators

Besides visual perception, peak-to-signal-noise ratio (PSNR) and structural similarity (SSIM) are usually used as evaluation metrics to assess the model performance in previous image denoising methods [5–9]. PSNR is a widely used metric in image reconstruction tasks and it is calculated based on MSE by (5)

$$\mathbf{PSNR} = \mathbf{10log_{10}}\frac{(\mathbf{2^n - 1})}{\mathbf{MSE}} \tag{5}$$

Both of these two metrics need reference images for comparison, hence, they are suitable to assess the denoising results of the simulated X-ray images. However, in clinical images, we have only low-dose X-ray images with uncertain noise level to be denoised, as there is no corresponding high-dose reference image. Therefore, for clinical images, we need an effective non-reference image quality assessment method. We utilize signal-to-noise ratio (SNR) to assess the denoising performance of our models on clinical datasets, and the SNR of a single image is the average of all patches in it. In the experiment, the patch size of sub-image is usually set to 16. Since the indicator used only stands for a ratio of average pixel value and standard deviation, there are no units for it.

## 3 Experiments

In our experiments, we used a desktop computer with an NVIDIA GTX1060Ti GPU with 6.0 Gb RAM, an Intel I7-8700K CPU @ 3.20 GHz with 16.0 Gb RAM for training and testing. Our model was implemented on TensorFlow with Python3.6 under Windows10, CUDA9.0 and CUDNN5.1.

### 3.1 Datasets

Table 1 shows a summary of all data used for experiments. Data sources were a publically available data of plain chest X-ray images [18] and clinical catheter laboratory images acquired at St. Thomas' hospital during cardiac catheter procedures.

**Table 1.** Summary of all data used in experiments

| Dataset | Source | Training | Testing |
|---|---|---|---|
| Chest X-ray (CXR) [18] | Standard plain chest X-rays<br>108,948 frontal view X-ray images from 32,717 patients | 5,000 X-rays images (1024 × 1024 pixels)<br>30,443 patches (96 × 96 pixels)<br>+ synthetic noise | 300 images, 300 central patches 576 × 576 pixels<br>+ synthetic noise |

(*continued*)

**Table 1.** (*continued*)

| Dataset | Source | Training | Testing |
|---|---|---|---|
| Catheter Laboratory Data 1 (CL1) | 1,080 X-rays images (96 × 96 pixels) Procedures at St. Thomas' Hospital, London | 800 X-rays images (512 × 512 pixels) 10,554 patches (96 × 96 pixels) + synthetic noise | |
| Catheter Laboratory Data 2 (CL2) | 623 X-ray sequences from 20 patients Procedures at St. Thomas' Hospital, London | | 100 low-dose X-ray sequences with 3,262 images (400 × 400 pixels) 72 high-dose X-ray sequences with 2,166 images (400 × 400 pixels) |

### 3.2 Training

The training data were created by adding 60% Poisson noise to the original training images. This represents a significant simulated reduction in radiation dose as mentioned previously and is comparative to what can be achieved on a clinical X-ray system. Based on the settings presented in [18], we inputted one batch consisting of 16 patches with the size of 96 × 96 from the training datasets (CXR & CL1) to our network each time. The learning rate was initialized to $10^{-3}$ for all layers and halved for every $10^4$ steps up to $10^{-5}$ and we selected PReLU as our activation following each convolution layer which contains 8 × 8 filters. The depth of UDB in our network was 6 and training a denoising model took approximately 20 h. We compared our method with other image denoising methods, including DnCNN [7] and DenseNet [15] (8 dense blocks) and trained these networks with the same training data.

### 3.3 Validation

We firstly examined the effectiveness of the proposed UDDN on both training and testing datasets to check whether the obtained model was overfitting or not. Figure 5 displays the comparison denoising results according to the iterations of UDDN on these two datasets. For the training dataset we used 10% of the complete training data (CXR & CL1) and for the testing data set we used all the data (CXR). Comparatively, the denoising results of training dataset exhibit faster convergence and visibly higher PSNR and SSIM than that of the testing dataset, and this superiority shows that the obtained denoising model performs better on training dataset than testing dataset. Accordingly, there is no overfitting for the proposed method and our UDDN is reliable for X-ray image denoising.

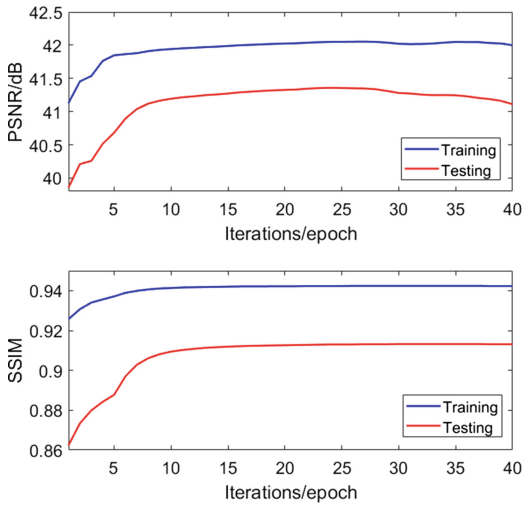


**Fig. 5.** Training process for UDDN with the noise level of 60%. The blue lines denote the convergence process of the training dataset and the red lines at the bottom refer to that of the testing dataset. (Color figure online)

### 3.4 Testing on CXR Dataset

The denoising results on the test CXR data with the noise level of 60% for the proposed approach and the comparison methods are shown in Fig. 6. We selected several but representative images with different contained structures (i.e., rib, spine, electrode and catheter). Subjectively it can be seen that our UDDN outperforms the other methods by visual inspection. In terms of the evaluation results, UDDN achieves a better PSNR and SSIM (41.4 dB and 0.914), which are about 0.4 dB and 0.015 higher than those of DenseNet and 4 dB and 0.12 higher than those of DnCNN, respectively.

Furthermore, to validate the more general ability of the CNN-based algorithms, we tested them with a range of noise levels from 0% to 100%. Figure 7 tabulates the results in terms of PSNR and SSIM. From these results, we can see that both the DenseNet and DnCNN denoising methods exhibit lower scores than our new network. DenseNet and UDDN are less sensitive to changes in noise level than DnCNN and UDDN shows better denoising capability when facing higher noise levels. Amongst the methods, UDDN shows the best performance because of its ultra-dense-connection-based effective framework for local spatial information extraction.

### 3.5 Testing on Clinical Data

Testing was also carried out using the CL2 clinical dataset. For these data there was no ground truth and so we used SNR as a measure of denoising performance. The CL2 dataset had both low-dose fluoroscopic images and high-dose acquisition images as detailed in Table 1. The results are shown in Table 2. In terms of fluoroscopic sequences, the proposed UDDN achieves the highest SNR (25.8), which is about

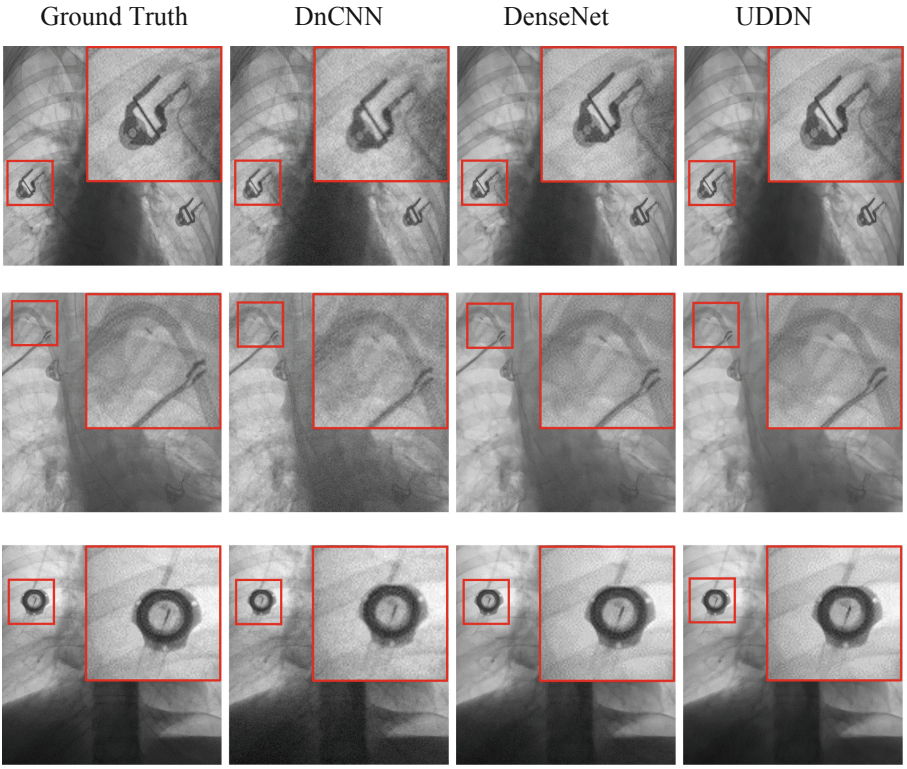


**Fig. 6.** Examples of the denoising results on the CXR test data using the noise level of 60%.

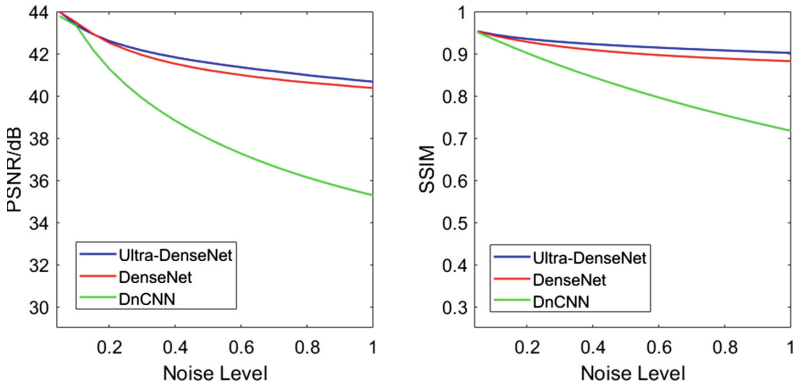


**Fig. 7.** A comparison of denoising results on the CXR test data for carrying noise levels using three CNN-based methods: DnCNN, DenseNet and our UDDN.

1.5 dB higher than that of DenseNet and 6.5 dB higher than that of DnCNN, respectively. However, both DenseNet and UDDN show less effectiveness on acquiration X-ray images than DnCNN, this may because they are insensitive to high-dose X-ray images.

**Table 2.** Average denoising results on the clinical dataset CL2

| | Fluoroscopic | Acquisition |
|---|---|---|
| UDDN | 25.8 | 25.7 |
| DenseNet | 24.3 | 25.4 |
| DnCNN | 19.4 | 27.6 |
| **Input** | 11.5 | 19.0 |

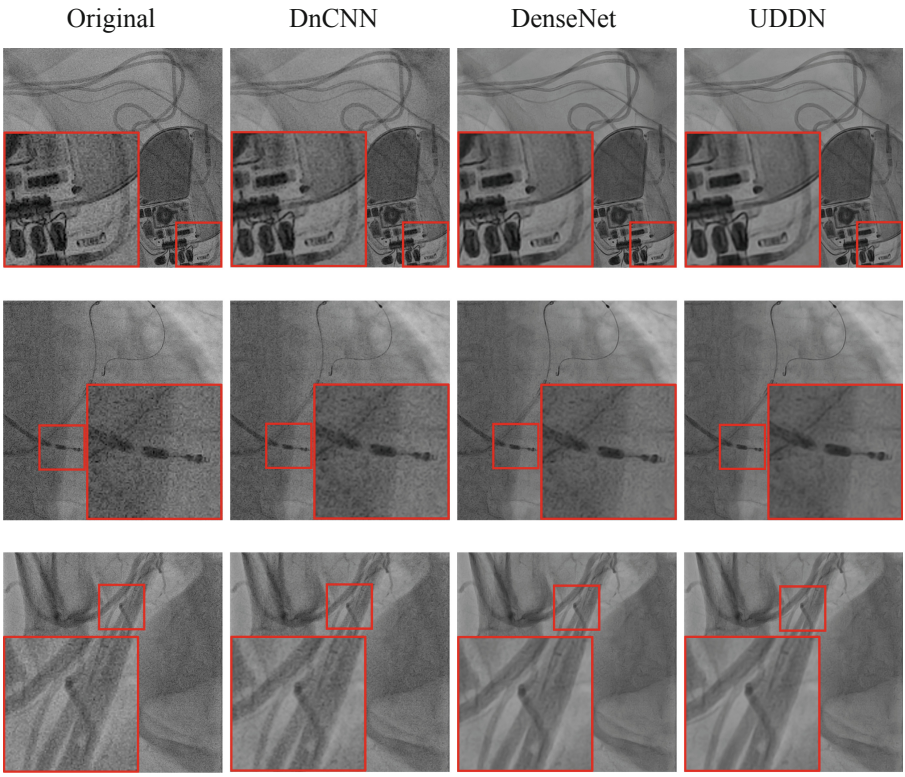


**Fig. 8.** The denoising results on the clinical dataset.

In terms of visual perception, we selected several different but representative structures from fluoroscopic images, i.e. vessels, electrodes and catheters to make comparisons. Notably, the proposed UDDN and previous DenseNet surpass DnCNN

significantly and UDDN is subjectively best under visual inspection. As illustrated in Fig. 8, compared to UDDN, the other two CNN-based methods produce more noticeable artifacts. These observations remained to be proved by future observational study using clinical experts as observers.

Both DenseNet and the proposed UDDN show effectiveness on low-dose X-ray image denoising, which validates the advantages of dense connection on network design. And UDDN slightly outperforms previous DenseNet by achieving higher quantitative Indicators but limited improvements on visual perception.

## 4 Conclusions and Future Works

In this study, we propose a simple but effective technique for X-ray image denoising in cardiac catheter-based procedures. In particular, we present a multiple-path UDB for local feature extraction. Unlike the previous DenseNet, more flexible dense connections between layers and units in our network promote information interaction and improve reutilization. Extensive experiments on both the simulated and clinical datasets indicate that UDDN outperforms existing CNN-based denoising techniques. This technique may prove valuable in dose reduction in the setting of real-time X-ray imaging for guiding interventions, especially paediatric interventions. Future work will focus on more extensive clinical testing (especially visual scoring of denoised imaged by expert clinicians), real-time implementation (current frame rate is 4 frames per second on the mentioned hardware) and real-time testing in the clinical setting. Moreover, the noise level of X-ray images obtained in actual clinical cases is uncertain and unpredictable. Therefore, a more flexible training method instead of using a fixed noise level should be proposed to promote network effectiveness.

**Acknowledgements.** This work was funded by the KCL NIHR Healthcare Technology Centre and KCL-China Scholarship Scheme. This research was also supported by the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's and St. Thomas' NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

## References

1. Cleary, K., Peters, T.M.: Image-guided interventions: technology review and clinical applications. Ann. Rev. Biomed. Eng. **12**, 119–142 (2010)
2. Wang, S., Housden, J., Zar, A., Gandecha, R., Singh, D., Rhode, K.: Strategy for monitoring cardiac interventions with an intelligent robotic ultrasound device. Micromachines **9**(2), 65 (2018)
3. Chen, H., et al.: Low-dose CT with a residual encoder-decoder convolutional neural network. IEEE Trans. Med. Imaging **36**(12), 2524–2535 (2017)
4. Kang, E., Min, J., Ye, J.C.: A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. Med. Phys. **44**(10), 360–375 (2017)
5. Crouse, M.S., Nowak, R.D., Baraniuk, R.G.: Wavelet-based statistical signal processing using hidden Markov models. IEEE Trans. Sig. Process. **46**(4), 886–902 (1998)

6. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans. Image Process. **16**(8), 2080–2095 (2007)
7. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Trans. Image Process. **26**(7), 3142–3155 (2017)
8. Wang, R., Tao, D.: Non-local auto-encoder with collaborative stabilization for image restoration. IEEE Trans. Image Process. **25**(5), 2117–2129 (2016)
9. Zhang, K., Zuo, W., Zhang, L.: FFDNet: toward a fast and flexible solution for CNN-based image denoising. IEEE Trans. Image Process. **27**(9), 4608–4622 (2018)
10. Kang, E., Chang, W., Yoo, J., Ye, J.C.: Deep convolutional framelet denosing for low-dose CT via wavelet residual network. IEEE Trans. Med. Imaging **37**(6), 1358–1369 (2018)
11. Cho, S.I., Kang, S.: Gradient prior-aided CNN denoiser with separable convolution-based optimization of feature dimension. IEEE Trans. Multimedia **21**(2), 484–493 (2019)
12. Yuan, Q., Zhang, Q., Li, J., Shen, H., Zhang, L.: Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network. IEEE Trans. Geosci. Remote Sens. **57**(2), 1205–1218 (2019)
13. Kim, J., Lee, J.K., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1646–1654 (2016)
14. Lai, W., Huang, J., Ahuja, N., Yang, M.: Deep Laplacian pyramid networks for fast and accurate super-resolution. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 5835–5843 (2017)
15. Tong, T., Li, G., Liu, X., Gao, Q.: Image super-resolution using dense skip connections. In: IEEE Conference on International Conference on Computer Vision, pp. 4809–4817 (2017)
16. Huang, G., Liu, Z., Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2261–2269 (2017)
17. Tao, X., et al.: Detail-revealing deep video super-resolution. In: IEEE International Conference on Computer Vision, pp. 4482–4490 (2017)
18. Wang, X., et al.: ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2097–2106 (2017)