https://doi.org/10.2352/EI.2024.36.10.IPAS-244

X-ray Image Denoising for Baggage Screening Using Learning Based Methods

Vempuluru Venkata Poorna Sekhar, National Institute of Technology, Calicut, Kerala, India; Deepthi P.P, National Institute of Technology, Calicut, Kerala, India; Renu M Rameshan, Vehant Technologies Private Limited, Noida, India

Abstract

The utilization of dual-energy X-ray detection technology in security inspection plays a crucial role in ensuring public safety and preventing crimes. However, the X-ray images generated in such security checks often suffer from substantial noise due to the capture process. The noise significantly degrades the quality of the displayed image and affects the performance of the automatic threat detection pipeline. While deep learning-based image denoising methods have shown remarkable progress, most existing approaches rely on large training datasets and clean reference images, which are not readily available in security inspection scenarios. This limitation hampers the widespread application of these methods in the field of security inspection. In this paper, we addressed a denoising problem designed for X-ray images, where the noise model follows a Poisson-Gaussian distribution. Importantly, our method does not require clean reference images for training. Our denoising approach is built upon the Blindspot neural network, which effectively addresses the challenges associated with noise removal. To evaluate the effectiveness of our proposed approach, we conducted experiments on a real X-ray image dataset. The results indicate that our method achieves favorable BRISQUE scores across different baggage scenes.

Introduction

The application of X-ray detection technology is widespread, spanning various domains such as medicine, industry, and security inspection. In the realm of security, X-ray technology is extensively utilized in locations like airports, customs checkpoints, and railway stations to combat criminal activities. Security personnel leverages the power of X-ray images as a means to determine the presence of prohibited items within packages. This enables them to make prompt and informed decisions, ensuring effective security measures are in place [1], [2]. However, the raw images obtained often contain noise, which degrades the image quality. Therefore, denoising X-ray images is an important task. Image noise may be caused by different intrinsic (i.e., sensor) and extrinsic (i.e., environment) conditions which are often not possible to avoid in practical situations. The noise that we are addressing is Poisson-Gaussian. X-ray emission and absorption are random processes that can be modeled well by the Poisson distribution considering the proportional relationship between X-ray intensity and the number of photons arriving at a point. When the mean of the Poisson distribution is high it is approximated as Gaussian distribution. Noise also arises from electronic components present in imaging systems. Gaussian noise is often used to model electronic noise [3]. As far as sensing noise is considered, X-ray images have spatially varying noise. The dense objects absorb most of the photons, leading to a reduced number of photons reaching the detector. In these regions noise is predominantly Poisson. While lesser-density regions have more noise with Gaussian distribution.

The goal of denoising is to reduce the noise while preserving as much of the original image information as possible. There are two main types of image denoising algorithms: traditional algorithms and learning-based algorithms. Traditional algorithms use mathematical techniques to suppress noise. Learning-based algorithms use neural networks to learn noise models from images [4], [5].

Many traditional methods for denoising images with Poisson noise, such as VST+BM3D [6], Non-Local PCA [7], and Patch Gaussian PCA [8] work by comparing the neighborhood of a pixel to other similar regions in the image. In other words, these methods look at the pixels that are close to a given pixel and use them to estimate the value of that pixel. This is done by assuming that the pixels in a neighborhood are likely to be similar to each other. These methods can be effective in suppressing the noise but do not eliminate it, and can also introduce artifacts and reduce the quality of the image. Also, it can be computationally expensive, especially for large images.

Learning-based methods try to learn the noise model from a collection of noisy images. Convolutional Neural Networks(CNNs) have recently been used in many image denoising applications with successful performance. CNNs have been shown to be effective at removing noise from images by learning to identify the noise patterns [9]. There are many supervised learning techniques for denoising where clean targets are required to train the neural network. Here, the network learns to map from a noisy image to its clean target. Zhang et al. (2017) [10] proposed a deep convolutional neural network (CNN)-based denoising method called DnCNN. It achieved better denoising performance than traditional image denoising methods for images affected by Gaussian noise. Both noisy images and their clean targets are required to train the network. In many cases, clean target images are not available. This limits the application of image-denoising methods that require them. To solve this problem, researchers have developed methods that do not require clean target images. One such method is Noise2Noise (N2N) [11], which requires only independent pairs of noisy images. Another method, Noise2Void (N2V) [12], does not require any pairs of images, clean or noisy. Noise2Void [12] is a self-supervised denoising method that learns to predict the clean value of a pixel from its noisy neighbors. A blindspot neural network is also a self-supervised denoising method and uses the same idea to predict clean pixel values. In [14], the authors used the concept of blindspot neural network and introduced a loss function appropriate for Poisson-Gaussian noise. Blindspot neural network approach which is mentioned in [14] is used in this work for denoising X-ray images. There is another method where the Poisson-Gaussian noise model was created for microscopy images [18] using pix2pix GAN [15], [16]. GAN trained on pairs of real images and their binary masks. It generated realistic noisy images. During testing, GAN produced noise on black masks, forming a synthetic background. Foreground was made by adding noise to signal regions in masks. This dataset was trained on another pix2pix GAN, performing well on real microscopy images [17]. However, this method could not be directly applied to security X-ray images, as they have a white background, and predominant is Gaussian and the Poisson noise is only present in the darker regions of the image.

The remaining sections of this paper are structured as follows: **Proposed Method** section provides a description of the proposed method, including its key details and the approach adopted for training and testing. In **Experiments and Results** section, the dataset used for evaluation, and the experimental results obtained from the implementation of the proposed method are presented and discussed in detail. Finally, **Conclusion** section concludes the paper by giving the insights gained.

Proposed Method

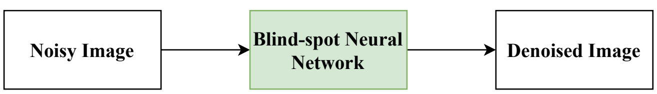

Figure 1. *Methodology followed.*

Figure 1 gives the block diagram of the proposed denoising algorithm. This method involves denoising the image using Blindspot neural network. A brief description of the Blind-spot neural network is given below for ease of understanding.

Blind-spot Neural Network

N2V [12] is a deep learning-based image denoising method that can be used with only noisy inputs. It uses a blind spot network to overcome the problem of degenerate learning, which occurs when a normal network is fed with the same noisy image as both input and label. The blind spot network works by not using the center pixel as input, but rather as the target. This encourages the network to learn denoising by leveraging information from neighboring pixels, thereby restoring the true values of the blind spot area. The network architecture is a slightly modified version of the five-level U-Net architecture [13], as utilized by Lehtinen et al. [11]. Three additional 1×1 convolution layers are included at the end of the network. Throughout the network architecture, all convolution layers utilize the leaky ReLU activation function, except for the final 1×1 convolution layer which employs a linear activation function [20].

In the self-supervised technique of denoising, the goal is to predict the values of a "clean" image $x = (x_1 \ldots x_n)$ given a "noisy" image $y = (y_1 \ldots y_n)$ by observing the neighborhood of the pixel y_i. The noise model considered in this work is the Poisson-Gaussian model. In the case of Poisson-Gaussian noise, a noisy observation is created by first adding Poisson noise to a clean observation, and then adding Gaussian noise that is independent of the clean observation. Further, since the mean of Poisson distribution is high, it is approximated as Gaussian with equal mean and variance [14]. Khademi et al. (2021) [14], proposed a loss function that is appropriate to address the Poisson-Gaussian model and used the same network architecture used in [20]. This loss function is used to train the blindspot neural network (BNN).

This network is trained in such a way that it takes inputs as noisy images and estimates mean μ_i and total variance $\hat{\sigma_i}^2$ at each noisy pixel by observing its neighbourhood. The loss function used to train this network is as follows:

$$\mathscr{L} = \sum_i \left(\frac{(y_i - \mu_i)^2}{\hat{\sigma_i}^2} + \log(\hat{\sigma_i}^2) \right) \tag{1}$$

The mean μ_i and total variance $\hat{\sigma_i}^2$ are estimates of the noisy pixel y_i. The clean pixel value is estimated using noisy pixel y_i, Poisson-Gaussian noise parameters a and b are obtained as shown below:

$$\hat{x_i} = \frac{y_i \sigma_i^2 + (a\mu_i + b)\mu_i}{(a\mu_i + b) + \sigma_i^2} \tag{2}$$

where a and b are Poisson-Gaussian noise parameters, $\sigma_i^2 = \hat{\sigma_i}^2 - (a\mu_i + b)$ and y_i is noisy pixel value [14]. These noise parameters a and b are estimated using Nelder-Mead optimization [19] with objective function as

$$a, b = \arg\min_{a,b} \sum_i \left(\frac{(y_i - \mu_i)^2}{a x_i + b} + \log(a x_i + b) \right) \tag{3}$$

where x_i is the clean pixel value. since the network does not have access to clean data μ_i is used instead of x_i [14].

Training strategy

The network architecture follows modified U-net architecture [20], trained for 300 epochs with a learning rate of 0.0003, each epoch consisting of 50 batches of 128×128 crops from random images from the training set and batch size of 4. In our dataset, each image is a 16-bit grayscale image and of different shapes. But the network can be trained only on 8-bit grayscale images of shape 512×512. So, the original image is mirror padded to 1024×1024 and that image is split into four 512×512 images. Khademi et al. (2021) [14] used this blindspot network for denoising microscopy images. So first we trained the network with microscopy images which are available, to make the network learn the noise model. When the network is directly tested with the split 512×512 x-ray images, the denoised images got blurred and visibility became poor. This is because most of the pixels in microscopy images are black and X-ray images have white background.

So, the Blind-spot network was re-trained with those split 16-bit 512×512 images and tested. During testing, the four parts of the mirror padded image are tested separately, then merged those denoised parts to form 1024×1024 and cropped out the mirror padded part to bring the denoised image back to its original shape as the test image.

Experiment and results

In this section, we begin by providing an introduction to our X-ray image dataset. The experiments in this study were conducted on a GPU system equipped with the following specifications: Intel(R) Core(TM) i7-10700 CPU, 32GB RAM, and NVIDIA GeForce RTX 3060 GPU.

Dataset Overview

The dual energy X-ray image dataset is obtained from a simulated baggage inspection environment, where bags were packed with random objects and subjected to X-ray scanning using security equipment. The dual-energy X-ray images used in our study were provided by Vehant Technologies Private Limited, India. The grayscale images vary in size and are captured in 16-bit format. The dataset comprises a total of 1217 X-ray grayscale images, and they were divided into a training set consisting of 1085 images and a test set containing 132 images. This division allows us to train and evaluate our proposed method effectively on diverse X-ray scenes and configurations present in the dataset.

Evaluation Criteria

Due to the unavailability of clean and noise-free X-ray images, traditional evaluation metrics like peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) cannot be utilized. To address this challenge, we opted for a non-reference image quality assessment method called Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) to objectively evaluate the performance of our method. BRISQUE leverages scene statistics of locally normalized luminance coefficients to quantify potential quality losses caused by distortions, enabling the evaluation of image quality. A lower BRISQUE value indicates higher image quality, providing a reliable metric for assessing the effectiveness of our approach [4].

BRISQUE algorithm involves normalizing the pixel intensities using Mean Subtracted Contrast Normalization (MSCN), MSCN images are multiplied by shifted versions of themselves in four orientations (horizontal, vertical, off-diagonal, and on diagonal). This captures the neighborhood relationships between pixels. Features are extracted at 2 scales - the original image scale, and at a reduced resolution (low pass filtered and downsampled by a factor of 2). From the normalized and pairwise product images, a feature vector of size 36×1(18 at each scale) is computed. Now, these feature vector is fed to a pre-trained support vector machine regressor to calculate the final quality score of the image [21].

Results and Discussion

The results of denoising for different baggage scenes are shown in Figures 2-4. In the context of 16-bit grayscale security images with a white background, identifying differences between noisy and denoised images can be visually challenging. So, qualitative results are shown in Tables 1, 2 for high and low energy images. From those results, it is observed that after denoising the quality of the image has been improved as denoised images have low BRISQUE scores.

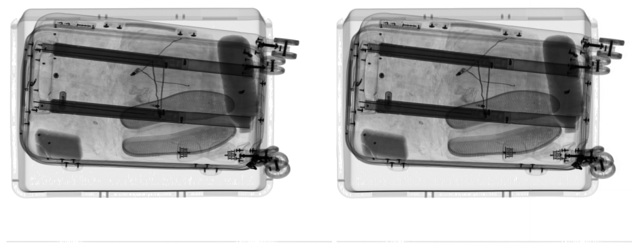

Figure 2. *BAG 1, Left to Right: Noisy version, Denoised Version*

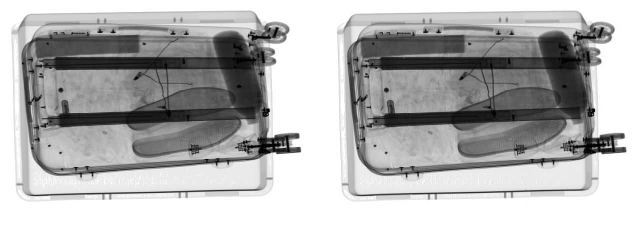

Figure 3. *BAG 2, Left to Right: Noisy version, Denoised Version*

Figure 4. *Left to Right: Noisy CTP image, Denoised CTP image*

Comparison Results

The deep learning method was compared against conventional techniques like VST+BM3D and non-local PCA for image denoising using the BRISQUE score metric. The results consistently favored the deep learning method, indicating that it outperforms VST+BM3D and non-local PCA in terms of image quality assessment (the visual comparison is shown in the Figure 5). The deep learning method's ability to capture and learn complex features from the data allows it to achieve higher perceptual quality, resulting in lower BRISQUE scores compared to the traditional methods. The results are shown in Tables 1 and 2. It can be observed that the BNN based solution has the lowest BRISQUE scores, showing the effectiveness of the method.

Though the results are good, there are some discrepancies that were observed. We tested the method for 132 images and the method failed for 16 images. Also, it is well known that the high energy images are more noisy, still, we observe that the BRISQUE

Table 1. BRISQUE Scores of different high energy baggage scenes before and after applying denoising methods

		Denoising methods			
Baggage Scene	Before Denoising	VST+BM3D	NLPCA	PGPCA	BNN
BAG3	35.5	36.72	47.43	36.14	**33.22**
BAG4	36.16	35.76	47.38	35.70	**33.61**
CTP	41.87	43.65	47.08	41.90	**35.94**

Table 2. BRISQUE Scores of different low energy baggage scenes before and after applying denoising methods

		Denoising methods			
Baggage Scene	Before Denoising	VST+BM3D	NLPCA	PGPCA	BNN
BAG3	36.5	41.29	46.99	36.35	**33.64**
BAG4	36.8	38.02	47.28	36.09	**33.65**
CTP	43.89	45.9	47.67	43.99	**39.44**

scores are lesser for high energy compared with the low energy. This could be due to the fact that BRISQUE scores are calculated based on perception and also, it was not trained for Poisson noise. We are working on these aspects to improve the method and the metric.

Conclusion

In this work, we have addressed the challenges posed by the Poisson-Gaussian noise model in security imaging by proposing a deep learning-based denoising method tailored to this specific noise model. Our experimental results have shown that the deep learning method outperforms traditional denoising methods, such as VST+BM3D and Non-Local PCA, in terms of denoising performance, as evaluated by the BRISQUE score.

References

[1] Riffo, Vladimir & Flores, Sebastian & Mery, Domingo. (2017). Threat Objects Detection in X-ray Images Using an Active Vision Approach. Journal of Nondestructive Evaluation. 36. 10.1007/s10921-017-0419-3.

[2] Wangzhe, Du & Shen, Hongyao & Jianzhong, Fu & Zhang, Ge & He, Quan. (2019). Approaches for Accuracy Improvement of the X-ray Image Defect Detection of Automobile Casting Aluminum Parts Based on Deep Learning. NDT & E International. 107. 102144. 10.1016/j.ndteint.2019.102144.

[3] Lee, Sangyoon, Min Seok Lee, and Moon Gi Kang. 2018. "Poisson–Gaussian Noise Analysis and Estimation for Low-Dose X-ray Images in the NSCT Domain" Sensors 18, no. 4: 1019. https://doi.org/10.3390/s18041019

[4] D. Liu, J. Liu, P. Yuan and F. Yu, "A Lightweight Denoising Method Based on Noise2Void for X-ray Pseudo-color Images in X-ray Security Inspection," 2022 4th International Conference on Industrial Artificial Intelligence (IAI), Shenyang, China, 2022, pp. 1-6, doi: 10.1109/IAI55780.2022.9976566.

[5] Girdher, Amina and Goyal, Bhawna and Dogra, Ayush and Dhindsa, Anaahat and Agrawal, Sunil, Image Denoising: Issues and Challenges (September 2, 2019). Proceedings of International Conference on Advancements in Computing & Management (ICACM) 2019.

[6] K. Dabov, A. Foi, V. Katkovnik and K. Egiazarian, "Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering," in IEEE Transactions on Image Processing, vol. 16, no. 8, pp. 2080-2095, Aug. 2007, doi: 10.1109/TIP.2007.901238.

[7] J. Salmon, C. -A. Deledalle, R. Willett and Z. Harmany, "Poisson noise reduction with non-local PCA," 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 2012, pp. 1109-1112, doi: 10.1109/ICASSP.2012.6288081.

[8] Deledalle, Charles & LTCI, CNRS & Salmon, Joseph & Dalalyan, Arnak. (2011). Image denoising with patch based PCA: local versus global. 10. 10.5244/C.25.25.

[9] Ilesanmi, A.E., Ilesanmi, T.O. Methods for image denoising using convolutional neural network: a review. Complex Intell. Syst. 7, 2179–2198 (2021). https://doi.org/10.1007/s40747-021-00428-4

[10] K. Zhang, W. Zuo, Y. Chen, D. Meng and L. Zhang, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," in IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3142-3155, July 2017, doi: 10.1109/TIP.2017.2662206.

[11] Lehtinen, Jaakko & Munkberg, Jacob & Hasselgren, Jon & Laine, Samuli & Karras, Tero & Aittala, Miika & Aila, Timo. (2018). Noise2Noise: Learning Image Restoration without Clean Data.

[12] Krull, Alexander & Buchholz, Tim-Oliver & Jug, Florian. (2019). Noise2Void - Learning Denoising From Single Noisy Images. 2124-2132. 10.1109/CVPR.2019.00223.

[13] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

[14] Khademi, Wesley & Rao, Sonia & Minnerath, Clare & Hagen, Guy & Ventura, Jonathan. (2021). Self-Supervised Poisson-Gaussian Denoising. 2130-2138. 10.1109/WACV48630.2021.00218.

[15] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. WardeFarley, S. Ozair, et al., "Generative adversarial nets", Advances in Neural Information Processing Systems, pp. 2672-2680, 2014.

[16] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou and Alexei A. Efros, "Image-to-Image Translation with Conditional Adversarial Berkeley rXiv:1611.07004v3 [cs.CV] 26 Nov 2018. Networks", Berkeley AI Research (BAIR) Laboratory UC.

[17] L. Zhong, G. Liu and G. Yang, "Blind Denoising of Fluores-

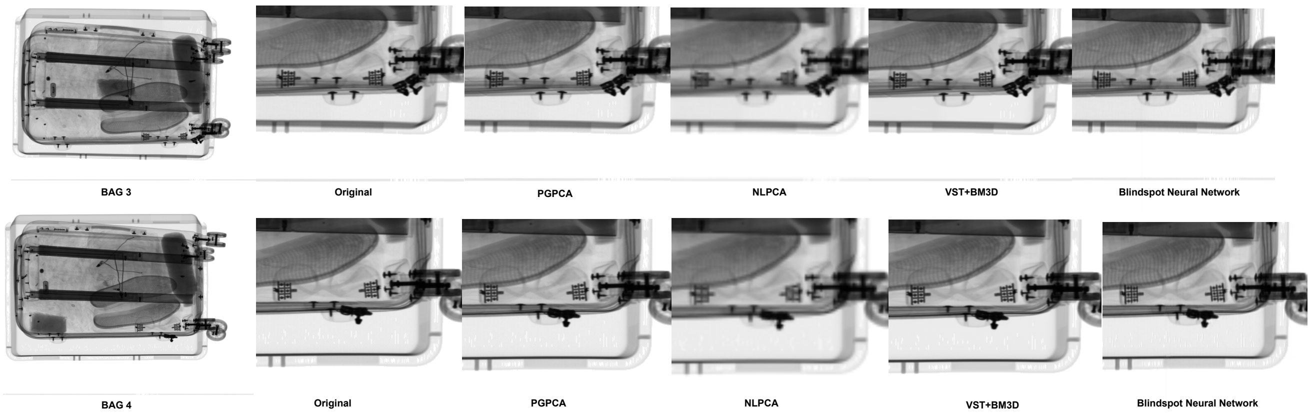

Figure 5. *Comparison between Traditional Methods and Deep Learning Method*

cence Microscopy Images Using GAN-Based Global Noise Modeling," 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France, 2021, pp. 863-867, doi: 10.1109/ISBI48211.2021.9434150.

[18] Yide Zhang, Yinhao Zhu, Evan Nichols, Qingfei Wang, Siyuan Zhang, Cody Smith, and Scott Howard. A poisson-gaussian denoising dataset with real fluorescence microscopy images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 11710–11718, 2019.

[19] John A Nelder and Roger Mead. A simplex method for function minimization. The computer journal, 7(4):308–313, 1965.

[20] Samuli Laine, Tero Karras, Jaakko Lehtinen, and Timo Aila. High-quality self-supervised deep image denoising. In Advances in Neural Information Processing Systems, pages 6968–6978, 2019.

[21] A. Mittal, A. K. Moorthy and A. C. Bovik, "No-Reference Image Quality Assessment in the Spatial Domain," in IEEE Transactions on Image Processing, vol. 21, no. 12, pp. 4695-4708, Dec. 2012, doi: 10.1109/TIP.2012.2214050.

Author Biography

Vempuluru Venkata Poorna Sekhar *is currently pursuing M.Tech in Signal Processing at the National Institute of Technology, Calicut, and received a B.Tech degree in Electrical Engineering from IIT Goa in 2020.*

Deepthi P.P *received the B.Tech. degree in electronics and communication engineering from the N. S. S. College of Engineering, Palakkad (University of Calicut), in 1991, the M.Tech. degree in instrumentation from the Indian Institute of Science, Bengaluru, in 1997, and the Ph.D. degree in secure communication from the National Institute of Technology Calicut, in 2009. She has been working as a Faculty in institutions under IHRD, Thiruvananthapuram, from 1992 to 2001, and the Department of Electronics and Communication Engineering, National Institute of Technology Calicut, since 2001. Her current research interests include signal processing with security applications, cryptographic system implementations, information theory, and coding theory.*

Renu M Rameshan *received the M.Sc (Engg.) in electrical communication engineering from the Indian Institute of Science, Bengaluru, India, in 1997, and the Ph.D. degree from IIT Bombay, Mumbai, India, in 2013., She worked as an Assistant Professor with the School of Computing and Electrical Engineering (SCEE), IIT Mandi, Suran, India. She is currently working as a Lead Research Scientist at Vehant Technologies Private Limited. Her research interests lie in the fields of image processing, computer vision, and ill-posed problems.*