

Deep learning has recently received a lot of attention as a feasible solution to a variety of artificial intelligence difficulties. Convolutional neural networks (CNNs) outperform other deep learning architectures in the application of object identification and recognition when compared to other machine learning methods. Speech recognition, pattern analysis, and image identification, all benefit from deep neural networks. When performing image operations on noisy images, such as fog removal or low light enhancement, image processing methods such as filtering or image enhancement are required. The study shows the effect of using Multi-scale deep learning Context Aggregation Network CAN on Bilateral Filtering Approximation (BFA) for de-noising noisy CCTV images. Data-store is used to manage our dataset, which is an object or collection of data that are huge to enter in memory, it allows to read, manage, and process data located in multiple files as a single entity. The CAN architecture provides integral deep learning layers such as input, convolution, back normalization, and Leaky ReLu layers to construct multi-scale. It is also possible to add custom layers like adaptor normalization (μ) and adaptive normalization (λ) to the network. The performance of the developed CAN approximation operator on the bilateral filtering noisy image is proven when improving both the noisy reference image and a CCTV foggy image. The three image evaluation metrics (SSIM, NIQE, and PSNR) evaluate the developed CAN approximation visually and quantitatively when comparing the created de-noised image over the reference image. Compared with the input noisy image, these evaluation metrics for the developed CAN de-noised image were (0.92673/0.76253, 6.18105/12.1865, and 26.786/20.3254) respectively

Keywords: convolutional neural network, residual learning, multi-scale context aggregation, CCTV images

UDC 621

DOI: 10.15587/1729-4061.2022.255789

IMPROVEMENT OF NOISY IMAGES FILTERED BY BILATERAL PROCESS USING A MULTI-SCALE CONTEXT AGGREGATION NETWORK

Zinah R. Hussein

Lecturer Assistant

College of Law

University of Baghdad

Al adhamia, Al Megreb str.,

Baghdad, Iraq, 10071

E-mail: mrs.zeena@colaw.uobaghdad.edu.iq

Received date 13.03.2022

Accepted date 22.04.2022

Published date 29.04.2022

How to Cite: Hussein, Z. R. (2022). Improvement of noisy images filtered by bilateral process using a multi-scale context aggregation network. *Eastern-European Journal of Enterprise Technologies*, 2 (9 (116)), 14–20.

doi: <https://doi.org/10.15587/1729-4061.2022.255789>

1. Introduction

Deep learning has recently gotten a lot of attention as a viable answer to a lot of challenges in artificial intelligence. When compared to other machine learning algorithms in the applications of object identification and recognition, convolutional neural networks (CNNs) outperform other deep learning architectures. When working on image processing operations on noisy images like fog removal or low light enhancement see the upper and lower portion of Fig. 1, respectively, it is essential to apply image processing algorithms like filtering or image enhancement.

Speech recognition, pattern analysis, and image identification, all benefit from deep neural networks. These kinds of deep neural networks have also been employed in medicine, where they have shown to be effective at predicting and classifying patient diagnoses. The U-Net model, for example, has shown good performance in image segmentations, a critical technique in medical imaging [1] and X-ray mammography [2]. Deep neural networks, on the other hand, are subject to adversarial models. Adversarial models are samples made by accumulating an amount of noise to an existing data sample so that they appear to be regular data to humans, but the classification model erroneously classifies them.

A potential problem in using this technique is that they must be manually engineered and designed to process an image. The deep learning network, on the other hand, learns

how to process the image using data. For example, enhancing the low-light image, using classical image processing, will first input image then apply haze removal algorithm and finally invert the image again as demonstrated in Fig. 2.

It is possible to replace these three operations into a single neural network so that its output resembles the output of the technique, which is called image processing operator approximation as shown in Fig. 3.

Previous research has shown that deep learning-based approaches for low-dose fluorodeoxyglucose (FDG) positron emission tomography can reduce noise (PET) [3]. Approximation of more generic and sophisticated procedures is possible with deep learning solutions. The work [4] firstly introduced a multi-scale Context Aggregation Network (CAN) that can imitate multi-scale tone-mapping, pencil drawing, and photographic style transfer, for instance. For improved accuracy in analyzing high-frequency features, Multi-scale CAN train on full-resolution pictures [5]. After the network has been trained, it can skip the traditional processing step and process images directly.

In summary, when working on image processing operations on noisy images, it is essential to apply image processing algorithms like fog filtering or image light enhancement. Studies addressed these issues recently but more accurate findings are required for the purposes of maximizing the performance of computer vision tasks like fog removal or low light enhancement.



Fig. 1. Noisy image filtering: *a* – fog removal; *b* – low-light enhancement image processing

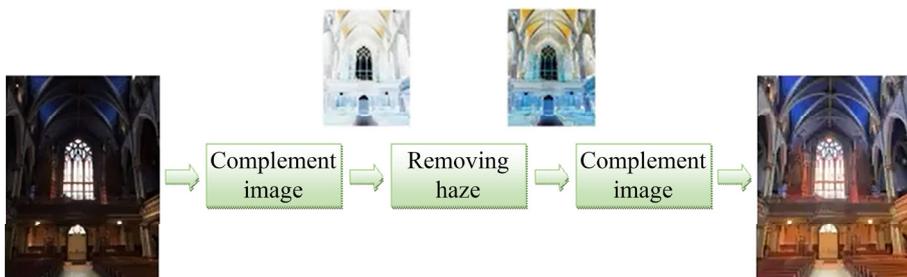


Fig. 2. The basic steps of enhancing the low-light image using classical image processing

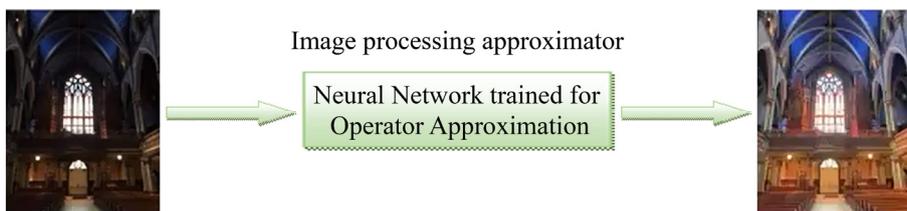


Fig. 3. Image processing approximator

2. Literature review and problem statement

A number of classical and deep network methods have been devised to achieve image operator approximations, as stated in the study [6], which suggested several traditional strategies for efficiency improvement of a given algorithm, but they can't be applied to other functions. The research tweaked the back-propagation technique to locate hidden layer targets and learn network weights efficiently while retaining performance. Another traditional method to approximate wide-range functions is to apply the approximator on a low-resolution replica of images, but due to high-frequency loss of contents, the precision of the approximation is limited. This issue is discussed by [7], in which

a particular convolutional kernel is used as filtering measured projections of computed tomography (CT) as an analytic reconstruction of images. For various reconstruction kernel selections, there is a tradeoff between noise and spatial resolution. In a clinical scenario, this frequently necessitates producing numerous pictures rebuilt with various kernels for a single CT exam, adding to the computational, reading burdens, networking, and archiving. Although this method has the potential to improve image quality, streamline, and reduce radiation exposure of the CT imaging clinical workflow it cannot process large resolution images. The paper [8] investigated the use of artificial neural networks and machine learning to de-noising a dynamic PET image by training a Deep de-noising Auto-Encoder (DAE) by using noise-free and noisy spatiotemporal image patches. However, the results provided a considerable decrease in the voxel/level noise but it required a complex algorithm for training data. The paper [9] proposed a paradigm to present deep learning on mobile strategy, where the network was trained in a decentralized method amongst thousands of edge clients. This algorithm prevents the privacy leakage of the confined model factors, which is perturbed by the Laplace-noise but cannot prove the practicality and effectiveness of their method. The study [10] used deep networks with faster algorithms to enhance the image quality of oncology FDG PET scans obtained in shorter times but the problem was with the limitation of the Full duration dataset, which is also addressed by [11]. This issue was addressed by the paper [12]

when it processed satellite images of satellite sensors with low-resolution. The study improved the performance of the network of filtering the noise but with several thousands of layers in the network.

From other applications, the research in [13] discussed the utility of motion artifact decreasing by the CNN in Multi-Arterial Phase MRI of the liver using 192 patients of a dataset. The paper presented an image filter for artifacts reduction by a deep learning network. Although the study has improved the quality of images and reduced the motion artifacts, this filter partially removes some anatomical details.

According to the above-mentioned literature, accurate image processing operations on noisy images like fog removal or low light enhancement is promising as it is

beneficial in a wide range of computer vision purposes such as robot navigation, and document analysis. Therefore, it is necessary to develop a Multi-scale CAN on Bilateral Filtering Approximation (BFA) for noisy CCTV images to obtain clear images.

3. The aim and objectives of the study

The main aim of the study is to show the effect of using Multi-scale CAN on Bilateral Filtering Approximation (BFA) of noisy CCTV images. This is achieved by conducting the following objectives:

- to create bilateral filtering approximation for a noisy input image;
- to perform CAN operator on the bilateral filtering noisy image;
- to evaluate the developed CAN approximation visually and quantitatively by comparing the produced de-noised image against a reference image using three image evaluation metrics (the Structural Similarity Index (SSIM), Naturalness Image Quality Evaluator (NIQE), and the Peak Signal-To-Noise Ratio (PSNR)).

4. Methods and Materials

This paper demonstrates how to use a multi-scale CAN to simulate the filtering process on CCTV images. The operator approximation method seeks different methods for processing CCTV images with the purpose that the result matches that of a traditional image pipeline or processing operations. The approximation of the CAN operator is frequently used to decrease the time amount consumed to process images. MATLAB-based Deep network platform is used to perform the developed image processing approximate for fog-removal operation on CCTV road images as depicted in Fig. 4.

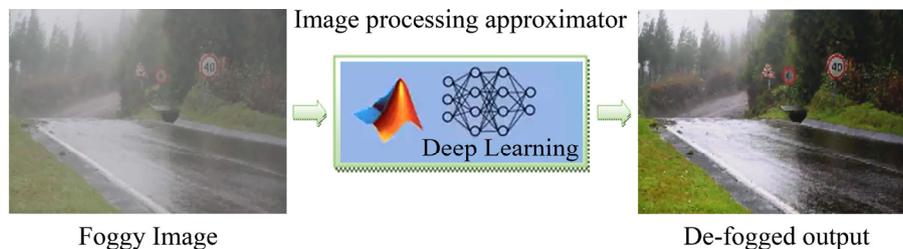


Fig. 4. The developed image processing approximate for fog-removal operation on CCTV road images

This research aims into training multi-scale deep learning CAN approximate image filtering by a bilateral operation that minimizes picture noise whilst maintaining image boundary sharpness. The considered application of CCTV images processing shows the entire inference and training operations, which includes setting the training settings, creating a trained data store, training the model, and using this model to analyze tested images.

To start with, let's set the foggy input images and corresponding the defogged label images to a different database. Let's use Data-store to manage our dataset, which is an object or collection of data that is huge to be processed in system memory, it allows to read, manage, and process data located in multiple files as a single entity. One of the most popular architectures used for image processing approximation is the Multi-scale Context Aggregation Network (CAN). This architecture includes the input layer, the middle layers, and the final layer as shown in Fig. 5.

The CAN architecture provides inherent deep network layers such as the input layer, back normalization layer, convolution Layer, and Leaky ReLU layer to construct multi-scale. It is also possible to add custom layers like adaptor normalization (μ) and adaptive normalization (Lambda) to the network.

After processing the input image via multi-scale context aggregation, the multi-scale CAN network is employed for minimizing the l_2 losses between the traditional yield of the image processing operations and the network responses. Rather than limiting the search to a narrow neighborhood surrounding the pixel, multi-scale CAN seeks information concerning every pixel throughout the whole image. The control algorithm of the CAN network is shown in Fig. 6.

The multi-scale CAN architecture has a huge receptive field to aid the network in learning global image attributes. Since the CAN operator be supposed to not modify the image's dimension, the first layer and last layer are the same sizes. Exponentially increasing scale factors widen successive intermediate levels, hence the "multi-scale" nature of the CAN. The dilation allows the architecture to search for spatially divided features at a different spatial frequency without lowering the resolution of input images. The network employs adaptive normalization after each convolution layer to stabilize the influence of identity mapping and batch normalization on the approximation operator.

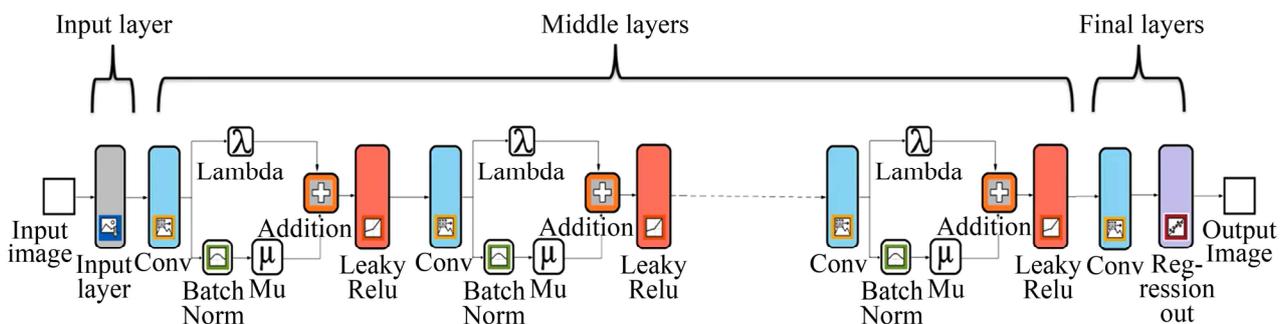


Fig. 5. Multi-scale Context Aggregation Network (CAN) architecture

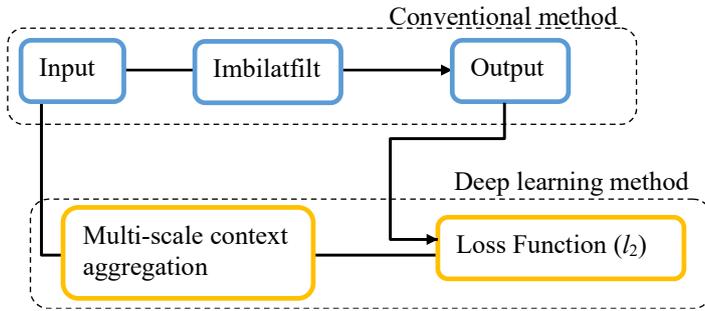


Fig. 6. The control algorithm of the CAN network

4. 1. Dataset

This dataset contains photographs and weather data obtained from the Polish General Directorate of National Roads and Motorways' network of measurement stations. These gadgets, which are installed along the country's main roadways, are outfitted with a CCTV camera and a collection of weather sensors. The generated dataset contains around 3 300 000 records gathered between November 2018 and March 2019 [14].

4. 2. Methodology steps

1. Preparing data for training, which includes training the network with a small subset of the downloaded dataset. Let's read in pristine images and write out bilaterally filtered images to construct a training data set. The filtered images are saved into a specified directory.

2. To train the network, let's create a Random Patch Extraction Data-store. The desired network responses and network inputs are stored in two image datastores, and this datastore selects random related patches from them. The network inputs in this study are the immaculate photos in pristine images. The processed images after bilateral filtering are the desired network responses.

3. An adaptive batch normalization layer is implemented using two custom scale layers. One scale layer controls the batch-normalization branch's strengths, while the other modifies the identity branch's strengths. Image patches are used by the first layer. The patch size is determined by the network receptive field or the spatial image region that influences the network's top-most layer's response. The network receptive field should ideally be the same size as the image so that it can perceive all of the image's high-level elements. The approximation picture patch size for a bilateral filter is set at 256 by 256 pixels. After the image input layer, a 2-D convolution layer with 32 3-by-3 filters is applied. Zero-pad the inputs to each

convolution layer so that feature maps after each convolution are the same size as the input. Let's set the weights to the identity matrix as a starting point.

4. An adaptive normalization scale-layer and a batch normalization layer follow each convolution layer, adjusting the batch-normalization branch strengths. The adaptive normalization scale-layer, which modifies the strength of the identity division, is then created.

5. Specify the network's middle levels using the same pattern. The factor of dilation of successive convolutional layers rises in exponential profile with the network depth.

6. In the convolution layer (from second to last), a dilation factor is employed. The final convolution layer reconstructs the image using only one filter of 1x1x32x3 size as an alternative to leaky ReLU layers. A regression layer is the last layer in the network. The mean-squared error for the network prediction and the bilaterally filtering image is computed by the regression layer. Concatenate all of the layers after that.

7. Making skip connections, which serve as the adaptive normalization equation's identity branch. Connect the additional layers to the skip connections.

8. Plot the layer graph.

9. The Adam optimizer is used to train the network, and the *trainingOptions* (Deep Learning Toolbox) function is used to specify the hyperparameter parameters. Let's employ the defaulting values of 0.8 for 'Momentum' and 0.0002 for the weight decay of the network, and train for 181 epochs on a fixed learning-rate of 0.0002.

10. Using Multi-scale CAN for Bilateral Filtering Approximation, which includes the following steps:

- using a reference image to produce a sample input noisy image;
- using the imbilatfilt function, do traditional bilateral filtering on noisy images;
- using the CAN, execute an approximator operation of bilateral filtering on noisy images;
- visually comparing the de-noised images produced by operator approximation to traditional bilateral filtering;
- measuring the similarity of the produced de-noised image quantitatively with respect to the pristine reference image to assess image quality.

The test data set, which consists of the test images, includes 21 clean photos that have been delivered to an image-Data-store (MATLAB-based Deep Learning functions). Arbitrary tested images are shown in Fig. 7.



Fig. 7. Arbitrary tested images

The training of the network includes specifying the training options and then using a train network function. After training, we tested the network with foggy images and pass them through the trained network. The visualization of the output image was performed using activation from the final regression layer.

5. Results of the developed CAN network

5.1. Using common bilateral filtering on a created noisy image

A noisy image is produced to compare the outcomes of operator approximation versus traditional bilateral filtering. For bilateral filtering, one image has been chosen to serve as a reference image and converted to (uint8) data type. As seen in Fig. 8, *a* reference image is displayed. The original clean and noisy images are shown as *a* and *b* respectively.

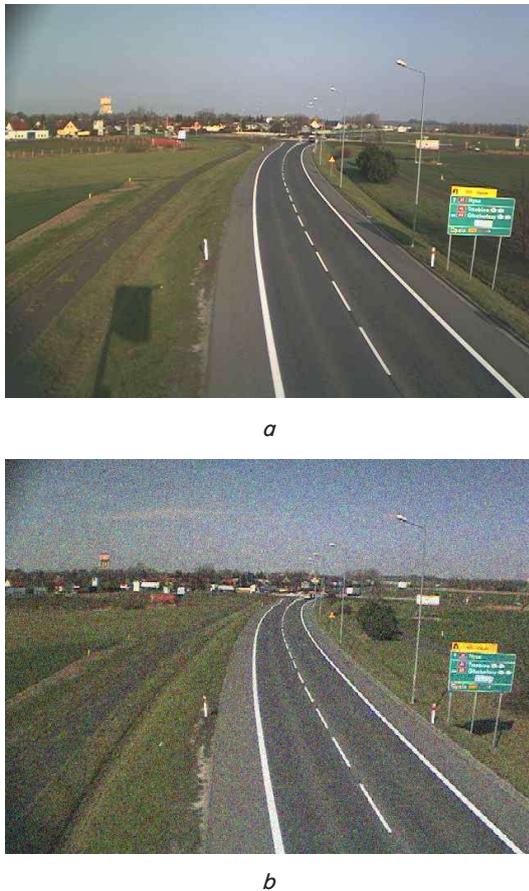


Fig. 8. Producing noisy image: *a* – reference original image; *b* – noisy image

The noisy image has been generated by adding Gaussian white-noise with zero mean and variance of 0.00002 to the reference image. The network requires an RGB test image that is at least 256 by 256 pixels in size.

The common bilateral filtering is a standard method to mitigate the noise and preserve edge sharpness, which requires specifying the smoothing degree to equal the pixel values variance. The de-noised image obtained from the common bilateral filtering is shown in Fig. 9.



Fig. 9. The de-noised image is processed by the common bilateral filtering

5.2. Performing CAN approximation operation on the bilateral filtering noisy image

After normalizing the input image, the developed CAN network output is the target de-noised image that is applied on the bilateral filtered noisy image as shown in Fig. 10.



Fig. 10. The developed network output de-noisy image using CAN

The result of the developed multi-scale deep learning CAN approximation operation, in Fig. 10, visualizes better quality, which is obtained from the regression layer (final layer).

5.3. Visually and quantitatively evaluated for the developed CAN approximation

Let's compare a cropped small region as a region of interest (ROI) for the developed CAN de-noised images obtained from CAN operator approximation with the common bilateral filtering image of the same reference image using the format (x-coordinate y-coordinate width height). The results of cropping the ROI images as a montage are shown in Fig. 11.

The developed deep learning CAN approximation of the lateral filtering method eliminates more noise than the common bilateral filtering method. Both methods maintain edge sharpness for images.

In order to quantitatively compare the images created using the conventional bilateral filtering and the developed deep learning CAN approximation, image quality metrics were computed. Four key factors to evaluate the produced image with respect to the reference image were used as described in Table 1.



Fig. 11. The cropped ROI images as a montage from: *a* –reference image; *b* – noisy image; *c* – bilateral filtering operation; *d* – developed CAN approximation

Table 1 outcome indicates that the CAN operator approximation creates better metric rates. To see the application of the developed deep learning CAN approximation operator on real CCTV Foggy input image, the processed de-noisy image is shown in Fig. 12.

As expected, the result of the developed multi-scale deep learning CAN approximation operation, in this Fig. 12, visualizes significant improvement on the input foggy image.

Table 1
Comparison among the noisy image, bilateral filtering image, and CAN approximation image

Factor	Description	Noisy	Bilateral	CAN
SSIM (Structural similarity index)	SSIM evaluates the visual impact of 3 image characteristics: contrast, structure, and luminance in opposition to reference images. The nearer SSIM value to 1 is the superior of tested image to agree with that reference images	0.76253	0.91578	0.92673
NIQE (Naturalness Image Quality Evaluator)	Perceptual image quality is evaluated with Lesser NIQE rates pointing to better perceptual quality	12.1865	7.22606	6.18105
PSNR (peak signal-to-noise ratio)	Larger PSNR values generally indicate better image quality	20.3254	26.421	26.786



Fig. 12. The applications of the deep learning CAN approximation method: *a* – foggy input image; *b* – defogged output image

6. Discussion of the results of developed CAN Network

In this paper, the use of a multi-scale context aggregation network (CAN) for performing the filter imaging process is investigated. In particular, a developed image processing approximate is used for fog-removal operation on CCTV road images on a training dataset. The insights of the architecture development by applying the CAN learning approach are outlined below.

The noisy image has been generated by adding Gaussian white-noise to the reference with at least 256 by 256 pixels in size image. Bilateral filtering was used to mitigate the noise and preserve edge sharpness, which requires specifying the smoothing degree to equal the pixel values variance as shown in Fig. 9. The produced image of the developed multi-scale deep learning CAN approximation operation is the target de-noised image that applied on the bilateral filtered noisy image as shown in Fig. 10, which shows significant improvements in image clearness by eliminating more noise than the common bilateral filtering method. However, both methods maintain edge sharpness for images. The effectiveness of the CAN network was evaluated with four key factors as listed in Table 1.

The success of the training increased as the evaluation of the objective function progressed. The objective function contains the operating cost over the training period equaling one episode of training. The multi-scale CAN architecture turns out to be an applicable approach for selecting promising. The evaluation parameters were obtained when the training was close to 100 iterations which are sufficient to train the developed network successfully. The developed CAN approximation learning approach shows potential as a promising approach for complex noisy images.

The limitation of this study is that it requires input noisy images with resolution no less than 256×256×3 that may not meet some CCTV datasets. However, this disadvantage can be eliminated in the future by combining this network with another flexible deep learning architecture.

7. Conclusions

1. A bilateral filtering approximation is created from a noisy input image by adding Gaussian white-noise to a reference CCTV image.

2. The performance of the developed CAN approximation operator on the bilateral filtering noisy image is proven when improving both the noisy reference image and a CCTV foggy image.

3. The three image evaluation metrics (SSIM, NIQE, and PSNR) evaluate the developed CAN approximation visually and quantitatively. The ratio of the SSIM, NIQE, and PSNR values of the CAN operator to the Bilateral

filtering were (0.92673/0.76253, 6.18105/12.1865, and 26.786/20.3254) respectively, which indicate the improvement when comparing the created de-noised image over the reference image.

Acknowledgments

I am grateful to all of those with whom I have had the pleasure to work during this and other related projects.

References

1. Kwon, H. (2021). MedicalGuard: U-Net Model Robust against Adversarially Perturbed Images. *Security and Communication Networks*. doi: <https://doi.org/10.1155/2021/5595026>
2. Zhu, G., Fu, J., Dong, J. (2020). Low Dose Mammography via Deep Learning. *Journal of Physics: Conference Series*. doi: <https://doi.org/10.1088/1742-6596/1626/1/012110>
3. Liu, H., Wu, J., Lu, W., Onofrey, J. A., Liu, Y.-H., Liu, C. (2020). Noise reduction with cross-tracer and cross-protocol deep transfer learning for low-dose PET. *Physics in Medicine & Biology*, 65 (18). doi: <https://doi.org/10.1088/1361-6560/abae08>
4. Chen, Q., Xu, J., Koltun, V. (2017). Fast Image Processing with Fully-Convolutional Networks. 2017 IEEE International Conference on Computer Vision (ICCV). doi: <https://doi.org/10.1109/ICCV.2017.273>
5. Sharma, S., Tang, B., Ball, J. E., Carruth, D. W., Dabir, L. (2020). Recursive multi-scale image deraining with sub-pixel convolution based feature fusion and context aggregation. *IEEE Access*. doi: <https://doi.org/10.1109/ACCESS.2020.3024542>
6. Kim, J., Kim, J., Jang G.-J., Lee, M. (2017). Fast learning method for convolutional neural networks using extreme learning machine and its application to lane detection. *Neural Networks*, 87. doi: <https://doi.org/10.1016/j.neunet.2016.12.002>
7. Missert, A. D., Yu, L., Leng, S., Fletcher, J. G., McCollough, C. H. (2020). Synthesizing images from multiple kernels using a deep convolutional neural network. *Med Phys*, 47 (2). doi: <https://doi.org/10.1002/mp.13918>
8. Klyuzhin, I. S., Cheng, J.-C., Bevington, C., Sossi, V. (2020). Use of a Tracer-Specific Deep Artificial Neural Net to Denoise Dynamic PET Images. *IEEE Transactions on Medical Imaging*, 39 (2). doi: <https://doi.org/10.1109/TMI.2019.2927199>
9. Zhang, J., Zhao, Y., Wang, J., Chen, B. (2020). FedMEC: Improving Efficiency of Differentially Private Federated Learning via Mobile Edge Computing. *Mobile Networks and Applications*, 25, 2421–2433. doi: <https://doi.org/10.1007/s11036-020-01586-4>
10. Mehranian, A., Wollenweber, S. D., Walker, M. D., Bradley, K. M., Fielding, P. A., Su, K.-H. et. al. (2022). Image enhancement of whole-body oncology [18F]-FDG PET scans using deep neural networks to reduce noise. *European Journal of Nuclear Medicine and Molecular Imaging*, 49, 539–549. doi: <https://doi.org/10.1007/s00259-021-05478-x>
11. Lim, H., Chun, I. Y., Dewaraja, Y. K., Fessler, J. A. (2020). Improved Low-Count Quantitative PET Reconstruction With an Iterative Neural Network. *IEEE Transactions on Medical Imaging*, 39 (11), 3512–3522. doi: <https://doi.org/10.1109/TMI.2020.2998480>
12. Deeba, F., Zhou, Y., Dharejo, F. A., Du, Y., Wang, X., Kun, S. (2021). Multi-scale Single Image Super-Resolution with Remote-Sensing Application Using Transferred Wide Residual Network. *Wireless Personal Communications*, 120, 323–342. doi: <https://doi.org/10.1007/s11277-021-08460-w>
13. Kromrey, M.-L., Tamada, D., Johno, H., Funayama, S., Nagata, N., Ichikawa, S. et. al. (2020). Reduction of respiratory motion artifacts in gadoxetate-enhanced MR with a deep learning-based filter using convolutional neural network. *European Radiology*, 30, 5923–5932. doi: <https://doi.org/10.1007/s00330-020-07006-1>
14. Grabowski, D., Czyżewski, A. (2020). System for monitoring road slippery based on CCTV cameras and convolutional neural networks. *Journal of Intelligent Information Systems*, 55, 521–534. doi: <https://doi.org/10.1007/S10844-020-00618-5>