Denoising with graphics processing units and deep learning in non-invasive medical imaging

Naoya OISHI

Medical Innovation Center, Kyoto University Graduate School of Medicine, 53 Shogoin-kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan
noishi@kuhp.kyoto-u.ac.jp

Abstract. Medical imaging is not only essential to the diagnostic process, but also plays a very important role in determining the course and effectiveness of treatment. In the last few decades, tremendous technological innovations have been made in the field of non-invasive medical imaging. Among them, imaging methods represented by computed tomography and magnetic resonance imaging are indispensable in current clinical medicine because they can acquire biological structures and functions in three to four dimensions with high spatial resolution non-invasively. However, the acquisition of data with high spatial resolution generally leads to a decrease in the signal-to-noise ratio. A longer acquisition time is required to improve the signal-to-noise ratio. However, for non-invasive medical image acquisition in clinical settings, a long acquisition time is impractical and results in a decrease in signal-to-noise ratio, especially in high spatial resolution images. It is thus essential to develop effective denoising techniques as post-processing and also to adapt the optimal denoising method in accordance with the user's objectives.

This review provides a brief overview of denoising techniques as post-processing for medical imaging, and introduces our work on fast and accurate denoising methods using graphics processing units and denoising with deep learning.

Keywords: Medical Imaging, Non-invasive, Denoising, Graphics Processing Units, Deep Learning

1 Introduction

Medical imaging is the non-invasive imaging of the internal structures and functions of living organisms for the purposes of screening and diagnosing various diseases and medical research, among others. Since the discovery of X-rays by Röntgen in 1895, medical imaging has revolutionized medicine, and medical imaging technology has made tremendous progress in recent years. Innovations and new discoveries in ultrasound, nuclear medicine, computed tomography (CT), and magnetic resonance imaging (MRI) have made medical imaging an indispensable tool of both clinical and basic medicine. Medical imaging is one of the mathematical inverse problems in terms of imaging the characteristics (cause) of biological tissue from the observed signals (result).

Recent technological advances have made it possible to acquire medical images faster, with higher image quality, and less invasively. However, there is a trade-off between faster and less invasive imaging and higher image quality. For example, high-speed imaging leads to a lower image signal-to-noise ratio (SNR), and obtaining high-SNR images requires a longer acquisition time, which also leads to increased invasiveness in methods that involve radiation exposure, such as CT and positron emission tomography (PET). In medical imaging, obtaining accurate information is extremely important for disease diagnosis and treatment decisions. Low-quality medical images containing noise can lead to misdiagnosis (Fig. 1). However, increased acquisition time and increased invasiveness should be avoided in medical practice. It is thus important to use image processing techniques to improve the SNR, that is, to remove noise from the acquired images as post-processing.

In this review, we provide an overview of denoising techniques as post-processing of medical images, and introduce denoising techniques using general-purpose computing on graphics processing units (GPGPU) with high speed and high accuracy, and denoising techniques using deep learning, including our own work.

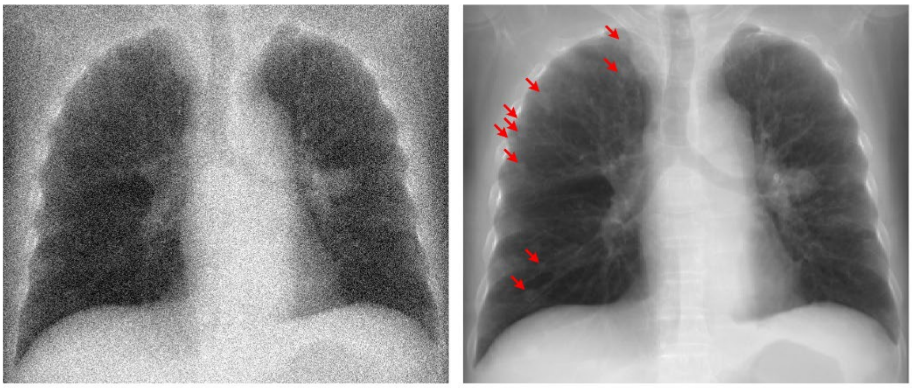

Fig. 1. Examples of low- and high-quality medical images. Left: Noisy chest X-ray. It is difficult to identify abnormal nodules from the image, which can lead to misdiagnosis. Right: The same chest X-ray without noise. Multiple nodules due to metastatic lung cancer can only be identified on a noiseless image (arrows).

2 Denoising with graphics processing units

2.1 Denoising methods for medical imaging

Image processing techniques for denoising can be broadly classified into two categories: one based on unsupervised learning and the other on supervised learning.

The former mainly uses methods that take advantage of redundancy in image patterns, while the latter uses deep learning and other techniques that have made remarkable technological progress in recent years. Both methods require an enormous amount of computation to achieve high-precision denoising, and since it is difficult for ordinary central processing units (CPUs) to perform real-time processing as required in the medical field, they are used only for simple denoising methods. Therefore, the use of GPGPU, a technology that converts the functions of graphics processing units (GPUs), which are good at repeatedly applying relatively simple numerical calculations to a large amount of data in parallel, is a solution to this problem. It enables high-speed and high-precision denoising.

Another advantage of utilizing GPGPU for denoising is its use as an image viewer. In the medical field, medical images are displayed on an image viewer, and physicians view the images and make a diagnosis. If a denoising filter can be applied and displayed in real-time to the level of denoising required by the physician, it is possible to obtain the optimal signal for the detection of abnormalities. However, highly accurate denoising requires an enormous amount of computation, making it difficult to adapt the filter in real-time. Therefore, the use of a GPGPU not only enables high-speed computation, but also further speeds up the process by allowing images computed on the GPU to be displayed directly on the screen without having to return them to the CPU (Fig. 2).

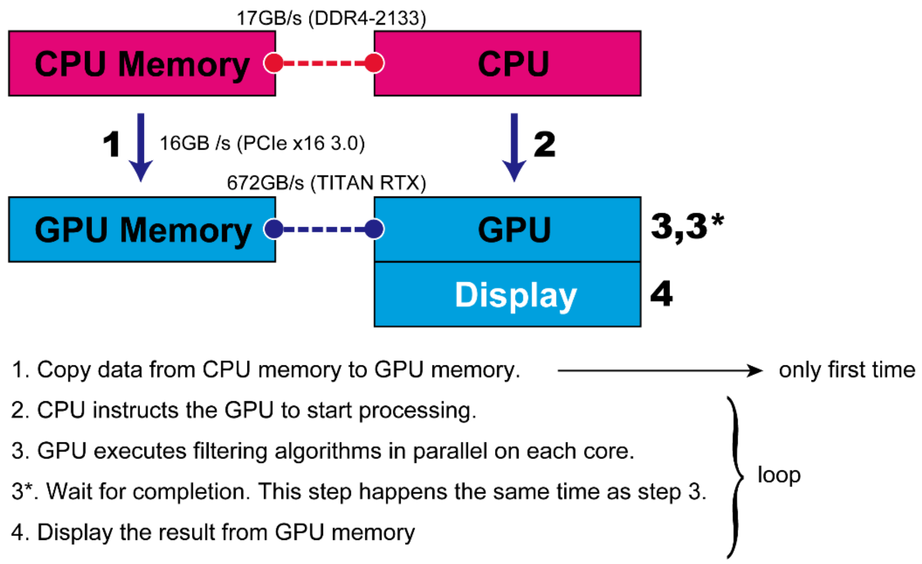

Fig. 2. General-purpose computing on graphics processing units (GPGPU) workflow in real-time filtering for an image viewer. The workflow requires data transfer from central processing unit (CPU) memory to graphics processing unit (GPU) memory only for the first time, which can be a bottleneck.

2.2 Perpendicular Gaussian filter for denoising

Although nowadays many medical images can be acquired as three-dimensional (3D) images, it is necessary to display certain cross sections, such as axial, coronal, and sagittal sections, as 2D images in order to display them in an image viewer. In this case, efficient denoising can be achieved, especially in medical images obtained with high spatial resolution, by taking into account the adjacent cross-sectional information that is not used for display. Blurred boundaries can also be avoided by averaging with stronger weights in the neighborhood of the currently displayed cross section. This filter is called a perpendicular Gaussian filter. This needs to be computed for each section, and since each computation is large, the use of GPGPU is indispensable for real-time display.

Optical coherence tomography (OCT) is a non-invasive technique that uses near-infrared light. It is a medical imaging technique that uses the scattering of reflected light to obtain three-dimensional images of biological tissue. While this technique can obtain 3D images with extremely high spatial resolution, it is also essential to use denoising techniques in conjunction because of the large influence of noise. Fig. 3 shows an OCT image of the skin of a healthy subject's palm. While it has extremely high spatial resolution, noise is noticeable. By applying a perpendicular Gaussian filter to each cross section, highly accurate denoising can be achieved, and structures such as sweat glands can be clearly seen. By applying this filter to OCT images of medaka brains, we were able to clearly delineate cerebral vasculature of 7 to 23 μm in size [1].

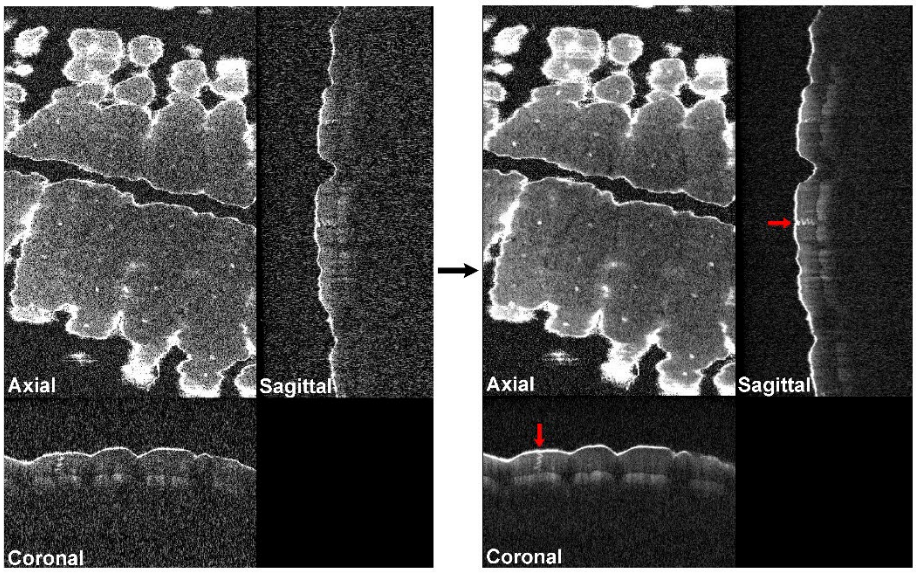

Fig. 3. Examples of optical coherence tomography (OCT) images of healthy skin of the palm of the hand. Left: Original OCT images in axial, coronal and sagittal sections. Right: Denoised images with a perpendicular Gaussian filter under general-purpose computing on graphics processing units. Sweat glands, which are three-dimensionally coiled tubular structures, can be easily identified (red arrows).

2.3 Non-local means filter

Not only medical images but also natural images are known to have spatial redundancy in image patterns, and image noise can be efficiently reduced by an averaging process using this property. The Gaussian filter, a major noise reduction filter, is based on the principle of signal averaging using spatial redundancy in images, and is widely used in medical imaging such as PET image reconstruction and functional MRI (fMRI) analysis because of its relatively high processing speed. However, this filter has the major drawback of blurring edges because it also averages dissimilar data (Fig. 4, middle panel). The non-local means (NLM) filter, proposed by Baudes in 2005, averages similar data in the image by weighting the data with similar spatial patterns, thereby reducing the blurring of edges [2]. This filter achieves efficient denoising while avoiding edge blurring (Fig. 4, right panel).

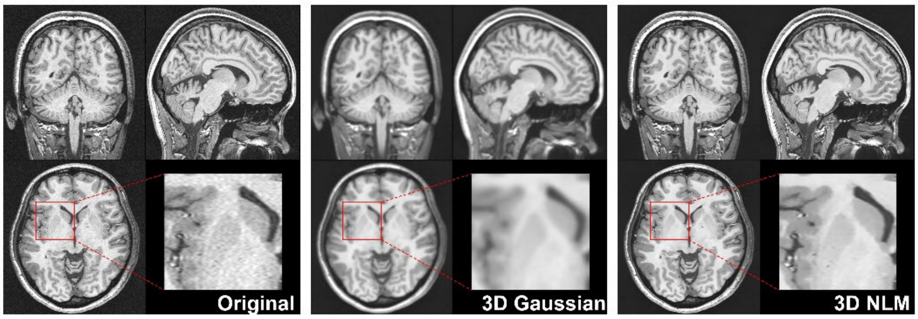

Fig. 4. Effects of denoising filters on brain magnetic resonance imaging (MRI) of a healthy subject. Left: Original MRI with extremely high spatial resolution (0.5×0.5×0.5 mm), but with noticeable noise. Middle: MRI denoised using a common three-dimensional (3D) Gaussian filter. Although the noise is reduced, the edges are blurred and detailed structures cannot be distinguished. Right: MRI denoised using the 3D non-local means (NLM) filter, which is extremely effective in denoising while avoiding blurred edges.

The NLM filter adjusts the value of each pixel with a weighted average of other pixels with similar geometric patterns in the neighborhood. Since the pixels in this image are highly correlated while the noise is generally independently and identically distributed (i.i.d.), averaging these pixels reduces the noise component and yields pixels that are close to the ideal value. However, a major drawback of the NLM filter is its huge computational complexity, which is especially apparent in 3D images. The NLM filter requires the calculation of the distance (similarity) between each voxel in the 3D image and all voxels in the spatial region defined as the neighborhood. In other words, if the size of the 3D image is N^3, the size of the search region is $(2S+1)^3$, and the size of the neighborhood region is $(2N_i+1)^3$, the complexity of the filter algorithm is on the order of $O\big(N(2S+1)(2N_i+1)\big)^3$. In fact, in the first application of 3D NLM to 3D MRI, Coupe et al. reported a computation time of 6 h on a 3 GHz CPU with an image size of $181\times217\times181$ and minimum values for S and N_i ($S=5, N_i=1$) [3]. Our implementation of the same 3D MRI with $S=7, N_i=1$ and a CPU (Xeon W-2295 3.0 GHz, 1 thread) takes approximately 301 s, which is faster than the previous implementation, but is not fast enough to be used for real-time viewing. Therefore, we implemented a 3D NLM using GPGPU and achieved increased speed of approximately 1,000 times faster than the CPU (Table 1). The same 3D MRI with $S=7, N_i=1$ takes about 0.28 s, which is 0.001 s for one slice, a level that poses no problem for real-time display.

Table 1. Comparison of three-dimensional non-local means (3D NLM) filter processing time using general-purpose computing on graphics processing units (GPGPU) and central processing unit (CPU). The magnetic resonance imaging (MRI) image size is 181×217×181 and the neighborhood size is fixed at 3 ($N_i=1$).

Search voxels/blocks	**Computational Time (in s)** **GPGPU** 1 x GeForce TITAN RTX	**Computational Time (in s)** **CPU** Xeon W-2295 3.0 GHz, 1 thread	**Ratio** **CPU/GPGPU**
5^3 = 225 voxels	0.17 (0.0009 s/slice)		
7^3 = 343 voxels	0.28 (0.001 s/slice)	301.2 (1.66 s/slice)	1043.4
11^3 = 1,331 voxels	0.73 (0.004 s/slice)		
15^3 = 3,375 voxels	1.5 (0.008 s/slice)		
21^3 = 9,261 voxels	3.7 (0.02 s/slice)		

As an example of the NLM filter, a segmentation using a brain MRI from the BrainWeb database (http://www.bic.mni.mcgill.ca/brainweb/) is shown in Fig. 5. This MRI has an image size of 181×217×181. Segmentation is a typical form of medical imaging processing of brain MRI, which separates the brain into gray matter, white matter, and cerebrospinal fluid. This is important for detecting brain atrophy that can occur in neurodegenerative diseases such as Alzheimer's disease. Segmentation is typically performed using a Gaussian mixture model based on signal intensity values, but noise affects its accuracy. In the middle panel of Fig. 5, 9% noise is added, which reduces the segmentation accuracy. The 3D NLM process improves the segmentation accuracy (Fig. 5, right panel). We applied the 3D NLM filter to reduce the amount of contrast agent, which could be toxic, used in repeated CT-angiography in a rat model of carotid artery occlusion [4]. We also adapted the 3D NLM filter to capture minute changes in brain structures using MRI in a stress model rat [5].

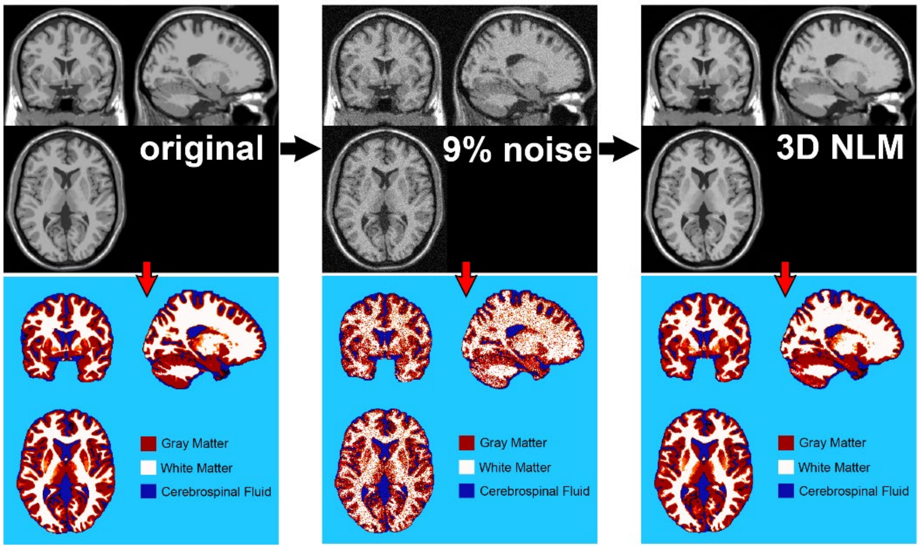

Fig. 5. Effect of the non-local means (NLM) filter on brain segmentation using three-dimensional magnetic resonance imaging (3D MRI). Left: Segmentation into gray matter, white matter, and cerebrospinal fluid using a Gaussian mixture model based on signal intensities. Since the data are generated by a simulation without noise, accurate segmentation is performed. Middle: Image with 9% Gaussian noise added. The noise has reduced the segmentation accuracy. Right: The 3D NLM filter applied to the noise image. The segmentation accuracy is close to that of the original image due to efficient denoising that avoids blurring of the edges.

3 Denoising with deep learning

3.1 Denoising methods with deep learning

The emergence of artificial intelligence (AI) in medicine is expected to have an impact comparable to the breakthrough discoveries of vaccines, anesthesia, sterilization, X-rays, antibiotics, and deoxyribonucleic acid (DNA) in the history of medicine. The recent tremendous progress in AI technology has been driven by deep learning. Deep learning is a method in which a multi-layered neural network architecture transforms input information into multiple levels of abstraction, automatically learning from the data representations that have been designed by humans in conventional machine learning, such as feature extraction. Among others, the development of convolutional neural network (CNN) technology for image pattern recognition tasks drove early deep learning techniques. CNNs have also been applied to various problems in medical imaging, such as lesion detection and classification, segmentation, and image reconstruction, of which noise reduction is a representative example.

Denoising with deep learning can be categorized as an image-to-image (I2I) translation task. I2I translation is the task of learning a model between images from a source domain and images from a target domain. The goal of I2I for denoising is to convert an input noisy image in a source domain to a target domain with the corresponding denoised image. I2I translation can be categorized into two main learning methods: supervised and unsupervised I2I translation. The supervised I2I translation can also be divided into two major networks: encoder-decoder network and generative adversarial network (GAN).

3.2 Denoising with the encoder-decoder network

A well-known encoder-decoder network is U-Net, which is derived from convolutional neural networks reported by Ronneberger et al. [6]. U-Net comprises an encoder-decoder network with skip connections. The input image is scaled down toward the lower layer (encoder) and enlarged as it returns to the upper layer (decoder). The encoder increases the ability to abstract the content of the image, whereas the decoder increases the ability to generate an image from the encoder's information. Skip connections can transfer non-abstract raw information, especially high-frequency signal information, directly to the final output layer. The name U-Net is derived from the shape of the overall architecture, which resembles the letter "U" in English. Although U-Net was originally developed for segmentation of biological tissue images, its usefulness has led to various applications in medical imaging, including denoising. We have also developed a multi-tasking deep learning approach that, in addition to identifying ischemic core regions in cerebrovascular disorders using U-Net, combines another deep learning model from the abstract representation reduced in dimensionality by U-Net to predict prognosis. We reported that the prognosis of mechanical thrombectomy for large-vessel occlusion was significantly improved by this method over the conventional method [7].

Fig. 6 shows an example of brain MRI denoising using U-Net. Recently, parallel imaging techniques using multi-channel phased array coils have been used to speed up MRI scans by acquiring only a portion of the k-space data collected by the MRI system. Although a variety of reconstruction algorithms have been developed, position-dependent variations in noise values in the reconstructed image occur, especially with sparse sample data. In other words, the noise is not spatially constant, making effective denoising difficult with conventional denoising techniques. In the example, a standard structural image (Fig. 6, first column) was acquired for 10 healthy subjects, which were imaged with a 3 Tesla MRI system for 6 min, and a noisy structural image (Fig. 6, second column), which was acquired in 3 min using a high-speed imaging method with a parallel imaging technique. The images were trained and validated by 5-fold cross-validation. Whereas U-Net achieves highly accurate denoising including at the center of the image (Fig. 6, 3rd column), noise in the center still remains with conventional denoising techniques such as the 3D NLM method (Fig. 6, 4th column). Although U-Net is a supervised learning method and thus requires a large amount of data for general-purpose denoising, it can be a useful method for situations where image acquisition protocols are somewhat fixed, such as in the medical field, because a training model can be constructed with a small number of data sets.

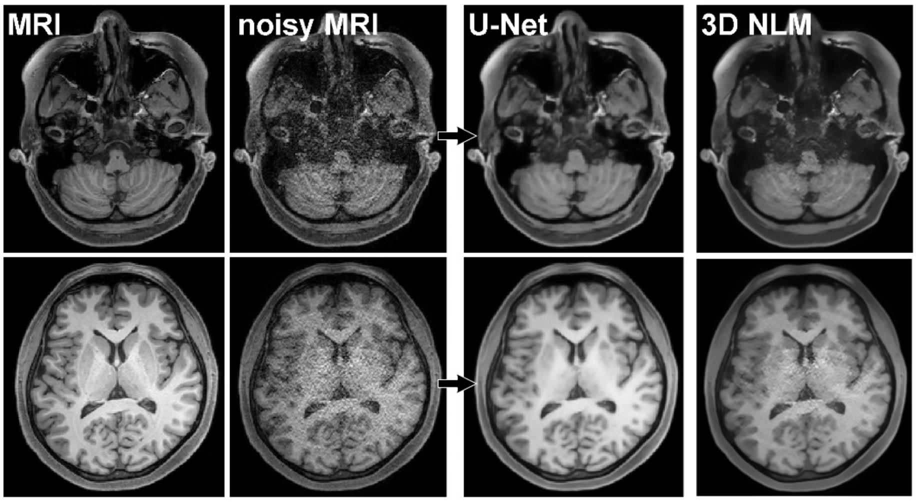

Fig. 6. Example of brain magnetic resonance imaging (MRI) denoising using U-Net. First column: Standard three-dimensional (3D) structural image acquired in 6 min with a 3T-MRI system. Second column: Noisy structural image acquired in 3 min by high-speed imaging method using parallel imaging technique. 3rd column: Denoised image using 3D U-Net. Denoising is achieved throughout the image, including the central region. 4th column: Denoising image using the 3D non-local means (NLM) filter. Although noise in the periphery has been denoised, noise in the center still remains.

3.3 Denoising with the generative adversarial network

A deep learning algorithm for image generation, GAN, was presented by Goodfellow et al. in 2014 [8]. GAN consists of two models, a generator and a discriminator, and the idea is that these two models learn adversarially. These two models are typically implemented in neural networks. The generator attempts to learn the distribution of true examples for generating new images. The discriminator, usually a binary classifier, learns to discriminate as accurately as possible between the generated image and the true image. Learning converges when the discriminator can no longer distinguish between the true image and the image generated by the generator. Since the publication of GAN, a vast number of GAN-derived methods have been proposed, such as conditional GAN (cGAN), which can be extended to a conditional model by conditioning both the discriminator and the generator on additional information. Isola et al. reported a GAN that performs image transformation by inputting corresponding images as pairs and identifying whether they are pairs, which they called Pix2Pix [9]. This is a type of cGAN because it conditions the corresponding images as additional information, and it is one of the I2I translations in terms of performing image transformations.

Fig. 7 shows an example of brain MRI denoising using Pix2Pix. In this example, as in the example in Fig. 6, a standard 3D structural image (Fig. 7, first column) and a noisy structural image (Fig. 7, second column) were acquired for 10 healthy subjects. Each was acquired in 6 min with a 3T-MRI system and 3 min with a high-speed imaging method using parallel imaging technology. The images were trained and validated by 5-fold cross-validation. Pix2Pix achieves more accurate denoising than the denoising using U-Net described in 3.2 (Fig. 7, 3rd and 4th columns). In particular, it should be noted that Pix2Pix reconstructs structures such as the cerebellar fissures and putamen (Fig. 7, 3rd column, red arrows), which are almost completely lost in the noisy structural image. This means that the concept of normal brain structure was effectively learned by GAN.

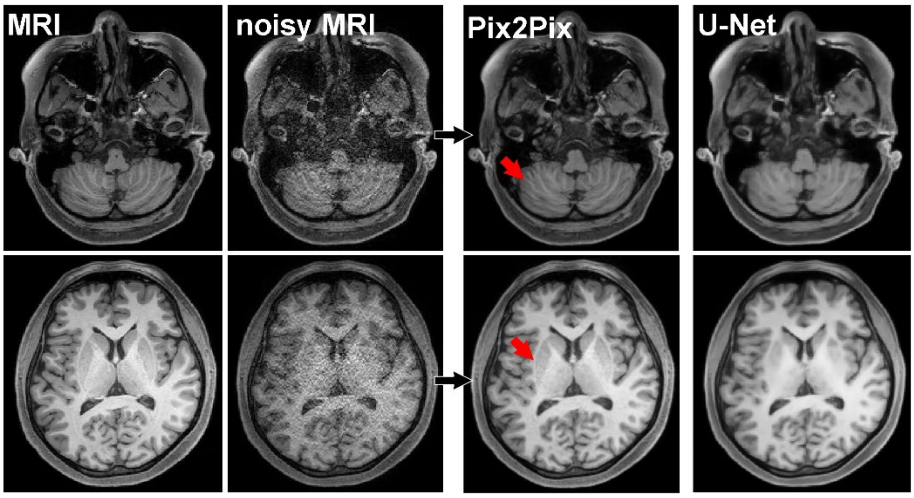

Fig. 7. Example of brain magnetic resonance imaging (MRI) denoising using Pix2Pix. First column: Normal structural image taken in 6 min with a 3 Tesla MRI system. Second column: Noisy structural image acquired in 3 min by high-speed imaging method using parallel imaging technique. 3rd column: Noise-reduced image using Pix2Pix. High-precision noise reduction was achieved, and structures such as the cerebellar fissures and putamen, which were almost completely lost in the noisy structural image, were restored (red arrows). 4th column: Denoised image using U-Net. Although the noise is removed with high-precision, it is worse than Pix2Pix, including the restoration of microstructures.

4 Conclusions

In this paper, we review a rapid and efficient method for denoising medical images using GPU and deep learning, including our own work. Table 2 lists the advantages and disadvantages of the denoising methods introduced in the paper. With the tremendous improvement in computer processing speed in recent years, real-time processing of computationally demanding high-precision noise reduction algorithms, which was difficult in the past, can now be realized using GPGPU. Furthermore, by using deep learning, even information lost due to noise can be recovered. This is a noise reduction technique at a level previously unthinkable. We believe that these techniques will generally be adapted to medical imaging in the future and will be beneficial to both clinical and basic medicine.

Table 2. Advantages and disadvantages of the denoising methods introduced in this paper

Denoising methods	Advantages	Disadvantages
Perpendicular Gaussian filter	The concept is simple. No training data are required. On a two-dimensional display, not only denoising but also 3D structural information of the previous and next slices can be visualized.	It is not suitable for 3D image analysis because it cannot be represented as a 3D structure.
Non-local means filter	By changing the parameters, it can adapt to images with a wide range of noise levels. Suitable for 3D image analysis because denoising with edge preserving can be expressed as a 3D structure. No training data required.	Noise remains when there is spatial non-uniformity in the noise properties.
Encoder-decoder network	Among deep learning algorithms, it is relatively stable over training. It can be adapted to cases where there is spatial non-uniformity in the noise properties.	Training data are required. It cannot be adapted to images with noise with different properties in comparison to the training data, and a new training model needs to be built additionally.
Generative adversarial network	It is possible to achieve denoising with restoration of even minute structures that have lost most of their information in the image.	Training data is required. It cannot be adapted to images with noise with different properties in comparison to the training data, and a new training model needs to be built additionally. Parameters can be difficult to adjust and it is sometimes unstable to train.

Acknowledgments

We would like to thank all of the participants in this study and all research collaborators for their invaluable work in data collection and analysis. This work was supported in part by JSPS KAKENHI grant numbers JP21K07593 and JP18K07712.

References

1. Suzuki, T., Ueno, T., Oishi, N., Fukuyama, H.: Intact in vivo visualization of telencephalic microvasculature in medaka using optical coherence tomography. Scientific Reports. 10(1), 19831 (2020).
2. Buades, A., Coll, B., Morel, JM.: A review of image denoising algorithms, with a new one. Multiscale Modeling & Simulation. 4(2), 490–530 (2005).
3. Coupe, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C.: An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. IEEE Transactions on Medical Imaging. 27(4), 425–441 (2008).
4. Kitamura, A., Fujita, Y., Oishi, N., Kalaria, RN., Washida, K., Maki, T., Okamoto, Y., Hase, Y., Yamada, M., Takahashi, J., Ito, H., Tomimoto, H., Fukuyama, H., Takahashi, R., Ihara, M.: Selective white matter abnormalities in a novel rat model of vascular dementia. Neurobiol Aging. 33(5), e25–e35 (2012).
5. Yoshii, T., Oishi, N., Ikoma, K., Nishimura, I., Sakai, Y., Matsuda, K., Yamada, S., Tanaka, M., Kawata, M., Narumoto, J., Fukui, K.: Brain atrophy in the visual cortex and thalamus induced by severe stress in animal model. Scientific Reports. 7(1), 12731 (2017).
6. Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv. 1505.04597 (2015).
7. Nishi, H., Oishi, N., Ishii, A., Ono, I., Ogura, T., Sunohara, T., Chihara, H., Fukumitsu, R., Okawa, M., Yamana, N., Imamura, H., Sadamasa, N., Hatano, T., Nakahara, I., Sakai, N., Miyamoto, S.: Deep learning-derived high-level neuroimaging features predict clinical outcomes for large vessel occlusion. Stroke. 51(5), 1484–1492 (2020).
8. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Neural Information Processing Systems, 2672–2680 (2014).
9. Isola, P., Zhu, JY., Zhou, T., Efros, AA.: Image-to-image translation with conditional adversarial networks. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 5967–5976 (2017).