Unpaired Unsupervised CT Metal Artifact Reduction

Bo-Yuan Chen, Chu-Song Chen

***Abstract*—A computed tomography (CT) scan is a medical imaging technique used in radiology to get detailed body images noninvasively for diagnostic purposes. Unfortunately, artifacts caused by metallic implants appear as dark and bright streaks in CTs, decreasing diagnosis quality. This paper later facilitated the diagnosis, demonstrating unsupervised learning methods to extract metal artifacts without destroying other tissue details from the original CTs.**

I. Introduction

Recently, deep learning has made great progress in the image processing and pattern recognition field. For example, the convolutional neural network (CNN) has been applied to medical imaging for low-dose CT reconstruction and artifacts reduction. U-Net [1], a kind of fully convolutional network, has many applications in biomedical image segmentation, such as brain image segmentation ("BRATS" [2]) and liver image segmentation ("siliver07" [3]). Variations of the U-Net have also been used for medical image reconstruction [4]. Unsupervised learning [5], which has no explicit target outputs or environmental evaluations associated with each input, is suitable in this task, for it is more ideal for collecting data without ground truth. Since it is feasible to collect unpaired noisy and clean images in most real-world applications, Unpaired learning [6] is generally practical and valuable in most medical problem settings.

This paper implemented a U-Net model to attempt realizing metal artifacts reduction, which is similar to UIDNet [7]. First, this paper assumes the dark and bright streaks at CT as low-frequency noises, typically been regarded as obscure features in an image. Second, this paper displays an end-to-end unpaired image denoising network that is trained through denoising (decomposes) images with only unpaired clean (CT without artifacts) and noisy (CT with artifacts) training data. Third, the generator in this U-Net based Conditional Generative Adversarial Network (cGAN) takes random noises (Gaussian noises) paired with clean data (CT without artifacts), serving as the conditions, as input and provides pseudo-artifact data merged with clean data as output, which are necessary

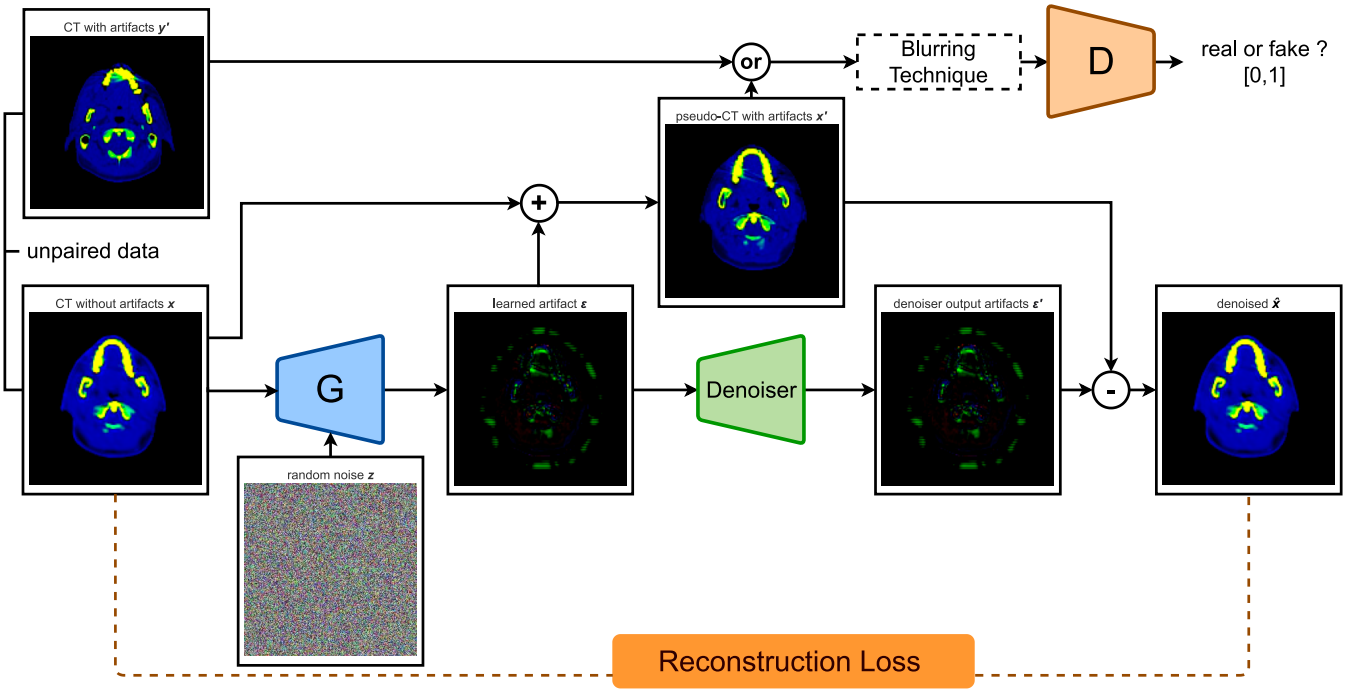

Fig. 1. An overview of the proposed framework. Given unpaired CT images, U-Net-based Conditional Generative Adversarial Network (cGAN) takes random noises with clean images (without artifacts) as conditional input to the generator (G in the figure), and the generator(G in the figure) gives pseudo-artifacts merged with clean images as output, noise images. Then, the discriminator (D in the figure) distinguishes the generated noise images from the actual noise images (with artifacts) . Finally, the U-Net-based denoising network (Denoiser in the figure) removes such pseudo-artifact from the noise images, similar to the existing denoising methods based on paired images.

procedures in this structure. Fourth, the U-Net-based denoising network removes the artifacts from the previous integrated data from the generator, similar to the existing denoising methods based on paired images.

The significant contributions of this methodology include: (1)We can treat feature-like components in images as low-frequency noises. (2) We presented a framework trained end-to-end for feature decomposition in images without paired supervision data. (3) We presented an image blurring method to make the GAN model focus on the feature-like information in images.

II. Related Work

A. Metal artifact reduction (MAR) algorithms

Some metal artifact reduction (MAR) algorithms have been developed to reduce metal artifacts [8]. Haewon Nam [9]claims these are categorized into two groups: projection completion methods, acquiring a high-quality prior image is essential, and iterative methods, which high computational cost is a practical obstacle.

B. Deep neural networks for unpaired image denoising

Recently, methods based on deep neural networks have performed better than traditional prior based methods [9]. However, the good performance achieved by the above methods is based on the premise that a large paired training data set is available. But this premise is not true in many practical situations. Hence, it's essential to seek the help of methods based on unpaired data.

C. CT artifact reduction in GAN

After Qingsong Yang introduced new CT image denoising method based on the generative adversarial network (GAN) [9]. GAN have been wide used in low-dose CT image denoising. Zhiwei Hong introduced UIDNet [9] made a big progress on unpaired image denoising. This paper then takes advantage of UIDNet by using adversarial unpaired image reducing feature like components(metal artifacts) without destroying other tissue images in the original CT with two U-Nets. An overview of proposed network is shown in (Fig. 1).

III. Proposed Method

The artifacts learning network is a Generative Adversarial Network (GAN). It learns the dark and bright streaks distribution in the CT Y' and transforms the clean data(without artifacts) X to its corresponding pseudo-artifact X'. It consists of a generator G and a discriminator D. The generator G is trained to generate samples that are closed to real data from a random noise sample and D is trained to distinguish whether a sample is generated by G or from real data. In the original GAN, D and G are trained to solve the following min max optimization problem.

Where E($\cdot$) denotes the expectation operator, and P_{data} and P_z are the distributions of real data and random noise. The plain GAN generates images just from a random noise sample z, which is not a proper option for us here because

$$\min_G \max_D \mathcal{L}_{GAN}(D,G) = \mathbb{E}_{x\sim P_{data}(x)}\left[log(D(x))\right] + \mathbb{E}_{z\sim P_z(z)}\left[log(1-D(G(z)))\right]$$

noise is often related to image content in most real world situations. Therefore, we have to revise the generator such that the generated artifacts is related to the clean data X. cGAN, the generator can generate samples based on some condition c, which could be any kind of auxiliary information. In this case, we treat a clean image X as the condition c and perform the conditioning by feeding it into the generator as an additional input apart from the random sample z.

A. Generative Network Architecture

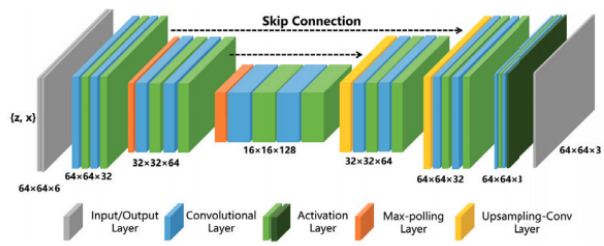

Fig. 2. The generator/ denoiser architecture of the proposed network.

The generator is a UNet (Ronneberger, Fischer, and Brox 2015) like network that takes a clean image X and random noise sample z as six channels of the input. It is widely known that the training of a GAN model tends to be unstable. Hence, it is difficult for the network to directly generate high quality pseudo-artifact CT X' with artifact similar to the given CT with artifacts Y' without introducing artifacts. To ease the training process, we set the generator to generate only artifacts and then add it to the clean image X to obtain the X'. Here we implicitly assumed that all the artifacts have the same beam-like distribution. The generator architecture (Fig. 2) consists of an encoder network and a decoder network with skip connections between them. The encoder part just follows the general CNN architecture with repeated conv-pool units, i.e., each unit consists of two 3×3 convolutional layers followed by a rectified linear unit (ReLU) (He et al. 2015) activation layer and a 2 × 2 max-pooling layer for downsampling. The number of feature channels in the first convolutional layer is 32 and it doubles at each max-pooling layer. The decoder part performs opposite processing with upsampling or deconvolutional layers. Here, we use an "upconv" operation consisting of an upsampling layer followed by a 2 × 2 convolutional layer to double the feature map size and halve the number of feature channels. Then, with a skip connection, it is concatenated with the corresponding layers in the encoder part, and is followed by two 3 × 3 convolutional layers and a ReLU activation layer. At the end, a 1 × 1 convolutional layer

is used to output the target image with the desired number of channels.

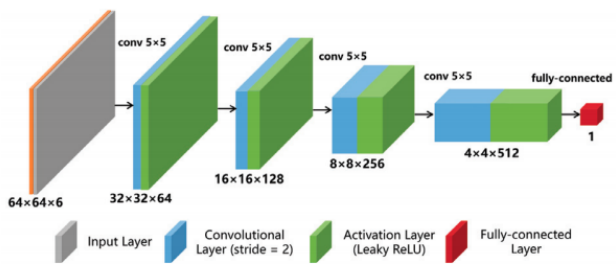

Fig. 3. The discriminator architecture of the proposed network.

B. Discriminative Network Architecture

The discriminator takes a real CT with artifacts from Y' or generated pseudo-artifact from X' as input and returns the probability that the image is sampled from real noisy data. The architecture of the discriminator is illustrated in (Fig. 3) which is similar to DCGAN (Radford, Metz, and Chintala 2015). In the training process, we use 64 $\times$ 64 image patches to train the GAN network. The images fed to the discriminator are X' and Y' that may have different image contents. To make the discriminator focus more on noise, we apply a sharpening technique by subtracting a image from its local-mean filtered version. The sharpened images and original images serve as two group of channels of the input for the discriminator. This sharpening technique is critical to the success of model in performance. The kernel size for calculating local mean is a hyper-parameter and is set to 3 $\times$ 3 in our experiments.

C. Denoiser Network Architecture

The cGAN in above subsection helps us to construct the paired clean and pseudo-artifact data X, X'. Now we perform image denoising similar to the existing paired methods. Here, this paper adopt a network architecture U-Net which is same as generator previous to help improve the performance in image denoising. Our network takes artifact ϵ as the input and outputs the residual artifact ϵ_i that it has learned. The denoised clean image $\bar{x}$ is then obtained by subtracting the artifact ϵ_i from the input artifact ϵ. The involved objective function is written below. where N is the size of the training

$$\mathcal{L}_{denoiser} = \frac{1}{2N}\sum_{i=1}^{N} ||(\,\overline{x} - x_i)||_F^2$$
$$\overline{x} = x_i' - \epsilon_i$$
$$\epsilon_i = f_\Theta(\,\boldsymbol{\varepsilon}\,)$$

data, θ denotes the network parameter we need to learn and the subscript i means the i-th sample of the data and F denotes Frobenius norm. The detailed network architecture is the same in (Fig. 2). Finally, the artifact learning and denoising modules are integrated together and trained end-to-end jointly.

IV. Experiment

In this section, we evaluate network's performance on real world data which is real medical CT images. Our experiments are divided into several parts and is compared with several representative methods in each part. (1) We evaluate our model' s performance on real photographs to show that it is capable of dealing with testing world artifact, which is supervised. (2) Our model is applied try to remove artifact to demonstrate our model' s capability in handling even more complicated noise. In the following, we denote our method with the blurring technique as Net, with the sharpening technique as Net-S and the version without the technique as Net-N. Semi-supervised(paired partially) with same structure as SS-Net. Changing Denoiser from U-Net to DNN network as Net-U2D, and same DNN network with semi-supervised(paired partially) learning as SS-Net-U2D.

A. Training and Test Data

Data includes 92 patients with each patient's brain nearly 50-70 CT images(total 2620 images)with resize size [3, 320, 320], and each patient contains about 3-5 metal artifacts. Then, 12 patients among it has 57 images in total are supervised denoised pair. First, divide those 12 patients to test data(rest to train). Second, Divides each of train and test into noise (with metal artifacts) and clear (without metal artifacts), and sends to the data loader. In order to balance the data, concatenate the noise 8 times. Although CT image is i dimension, still treat it as RGB to indicate the novelty of this network.
Train noise: 255, Test noise: 57, Total noise: 312
Train clear: 1899, Test clear: 409, Total clear: 2308

TABLE I
METRICS OUTPUT OF EACH METHOD.

	Method	**PSNR**	**MAE**
semi-supervised	SS-Net-$_{cv:1}$	46.22	1.073
	SS-Net-$_{cv:2}$	46.25	1.056
	SS-Net-$_{cv:3}$	46.08	1.029
	SS-Net-U2D-$_{cv:1}$	45.97	1.090
	SS-Net-U2D-$_{cv:2}$	45.87	1.104
	SS-Net-U2D-$_{cv:3}$	46.02	1.074
unsupervised	Net-U2D	45.54	1.12
	Net-N	45.94	1.076
	Net-S	45.97	1.096
	Net	46.04	1.078

B. Artifact reduction

Average metrics of the artifact reduced CT images of each denoiser method's output calculate with the supervised paired real clear(without artifact) CT images(Table. I). The supervised paired real clear(without artifact) and noise(with artifact) CT images' average PSNR is 47.86, MAE is 0.81. All results are show in(Figure. 4 - 8), and the format are all the same as (Figure. 4)

V. Conclusions

A. Denoiser

It can be observed that there are indeed the characteristics of learned artifact in Learned noise which usually be generated at the part of teeth, but there is also a lot of unnecessary information, such as the outer line in Denoiser output artifact(Figure. 4). Although CT images remove part of metal artifact in Denoiser which is DNN network, but it may also remove key information in the brain from CT images. Hence, reconstruct noise with U-Net can avoid this problem.

B. Generator(G)

The learned artifact image generated by the Generator can roughly generate the brighter areas in the CT image, and learn the characteristics of the dark and bright streaks at CT.

C. Blurring technique

Originally, by applying AvgPool2D blur image to make G learn higher frequency information, modifying +AvgPool2D to -AvgPool2D to achieve sharpening image. The result turns out that the Generator(G) tend to learn higher frequency noise which is far from achieving the goal of this task.

References

[1] ”U-Net code”. https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/

[2] ”MICCAI BraTS 2017: Scope — Section for Biomedical Image Analysis (SBIA) — Perelman School of Medicine at the University of Pennsylvania”. www.med.upenn.edu. Retrieved 2018-12-24.

[3] ”SLIVER07 : Home”. www.sliver07.org. Retrieved 2018-12-24.

[4] Andersson J, Ahlstrm H, Kullberg J (September 2019). ”Separation of water and fat signal in whole-body gradient echo scans using convolutional neural networks”. Magnetic Resonance in Medicine

[5] Unsupervised Learning, Peter Dayan. MIThttp://www.gatsby.ucl.ac.uk/ dayan/papers/dun99b.pdf

[6] Unpaired Learning of Deep Image Denoising, Xiaohe Wu, Ming Liu, Yue Cao, Dongwei Ren, Wangmeng Zuo. https://arxiv.org/abs/2008.13711

[7] End-to-End Unpaired Image Denoising with Conditional Adversarial Networks, Zhiwei Hong, Xiaocheng Fan, Tao Jiang, Jianxing Feng. https://ojs.aaai.org//index.php/AAAI/article/view/5834

[8] G . Joselin Retna Kumar et. al, “Design and Control of Ball on Plate System” International Journal of Control Theory and Applications, Vol. 9(34) 2016, pp. 765-778, December 2016, ISSN: 0974-5572.

[9] Deep Learning vs. Traditional Computer Vision, Niall O’ Mahony, Sean Campbell, Anderson Carvalho, Suman Harapanahalli, Gustavo Velasco Hernandez, Lenka Krpalkova, Daniel Riordan, Joseph Walsh. https://arxiv.org/pdf/1910.13796.pdf.

[10] A. Knuplez, A. Chowdhury and R. Svecko, ”Modeling and control design for the ball and plate system,” IEEE International Conference on Industrial Technology, 2003, Maribor, Slovenia, 2003, pp. 1064-1067 Vol.2, doi: 10.1109/ICIT.2003.1290810.

APPENDIX

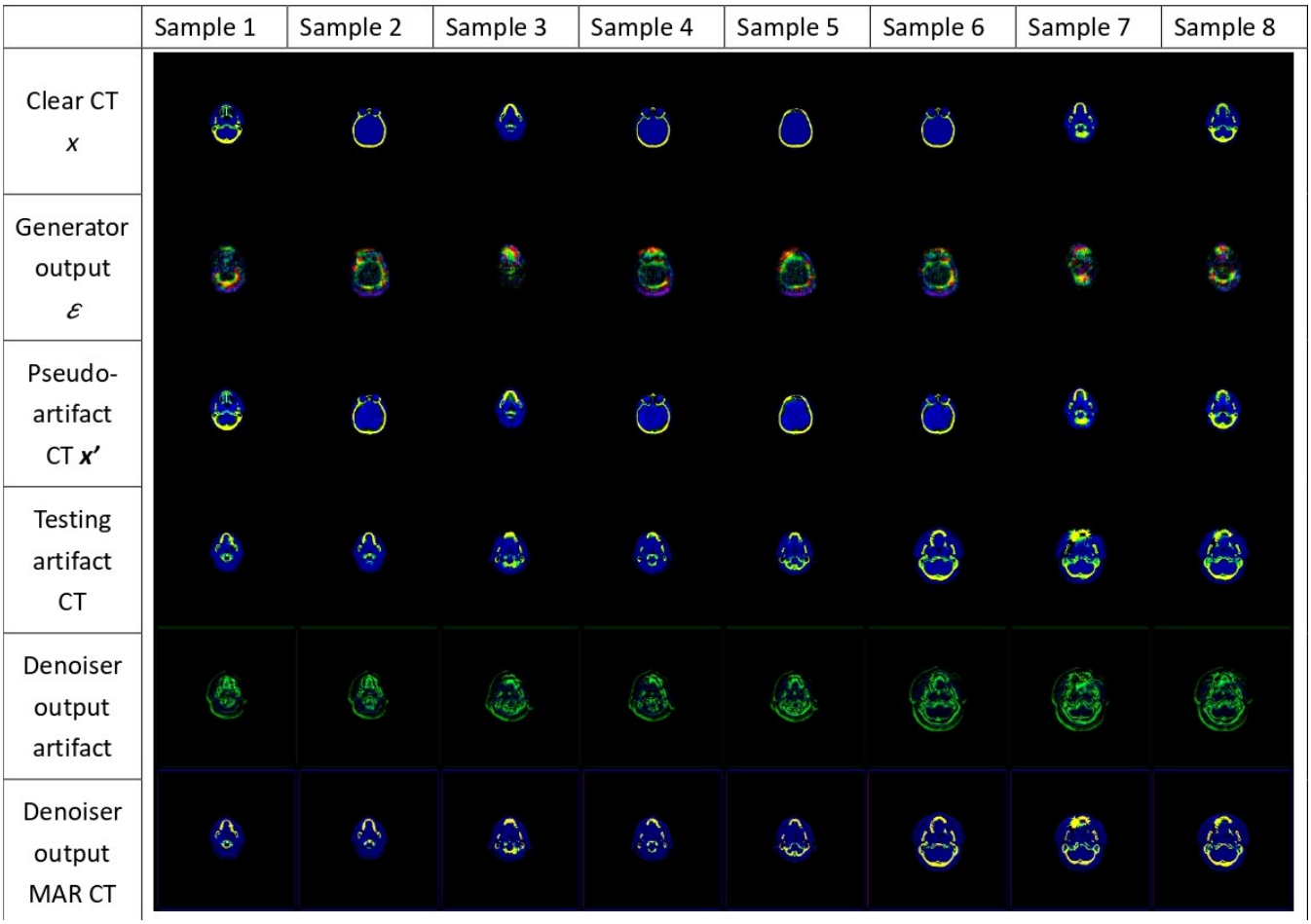

Fig. 4. semi-supervised learning with Denoiser modifying U-Net to DNN

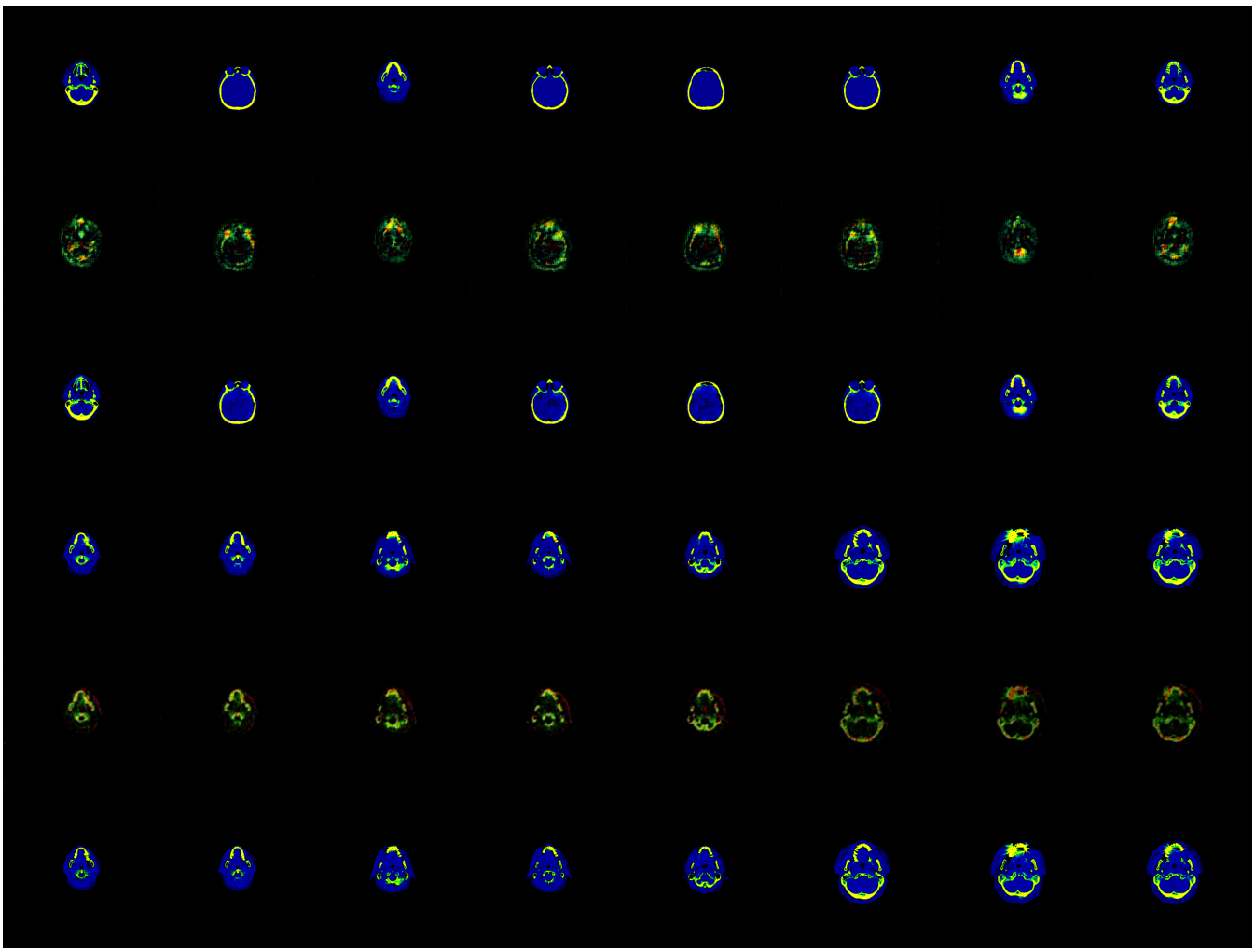

Fig. 5. semi-supervised learning with Denoiser using U-Net

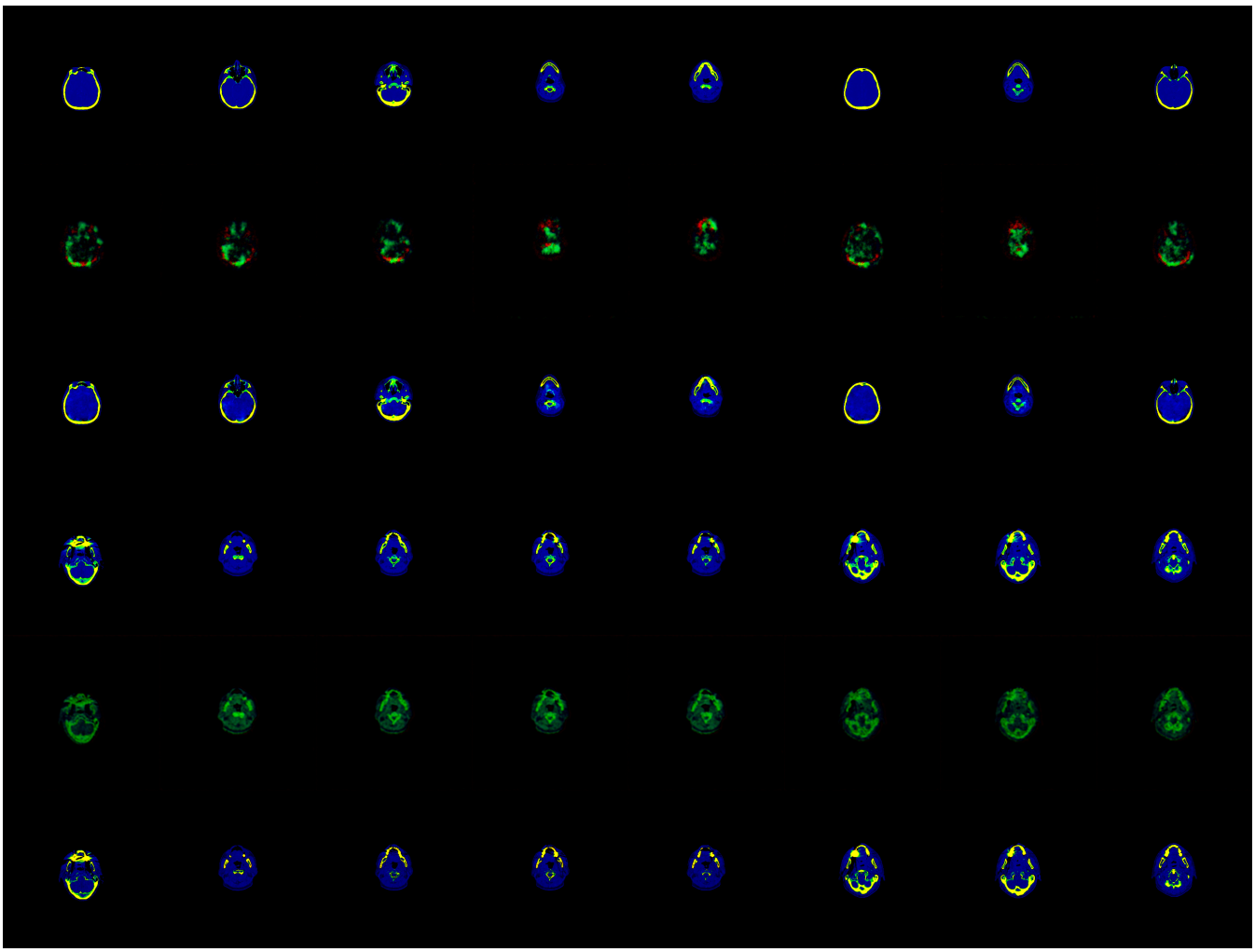

Fig. 6. Unsupervised learning with Denoiser using U-Net

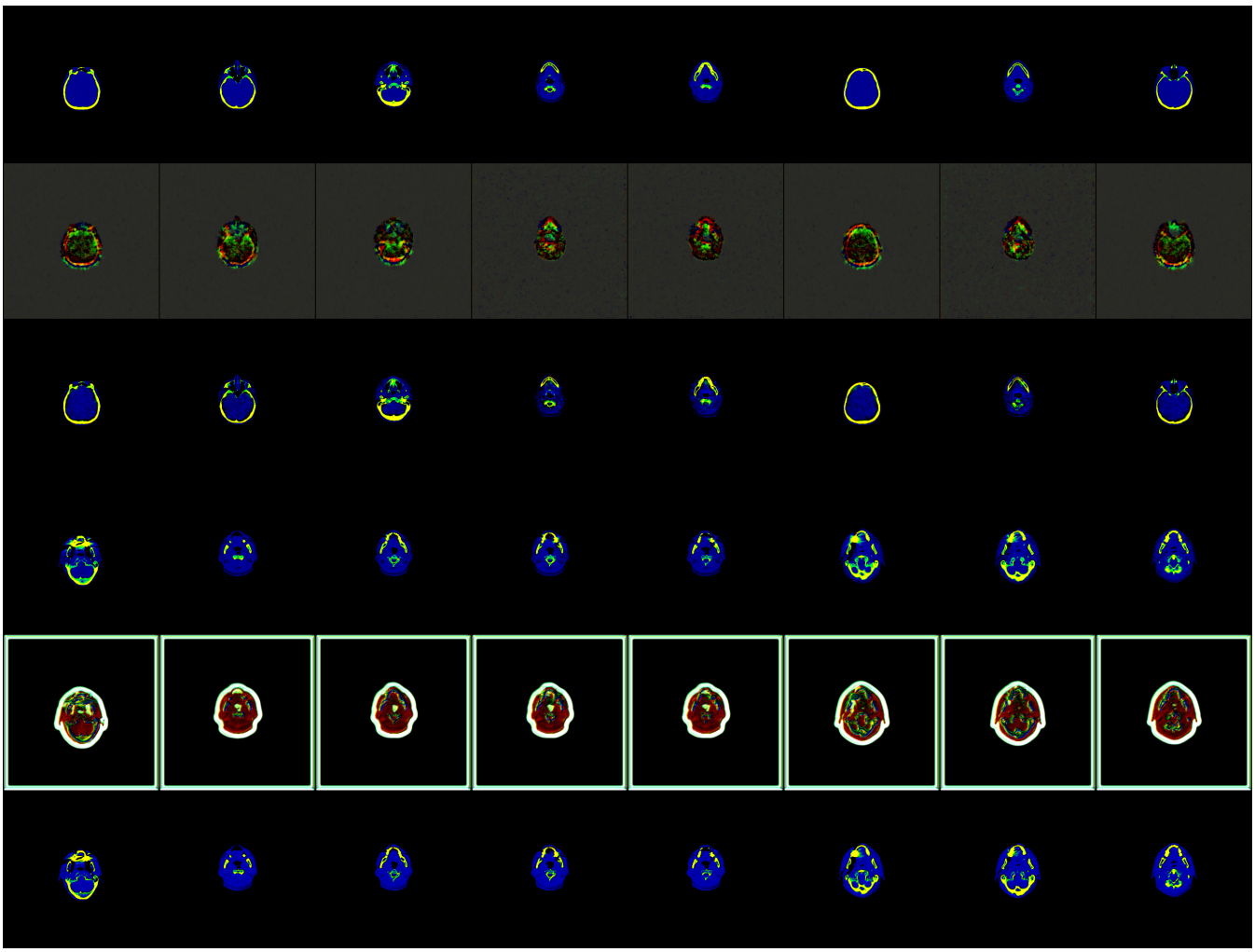

Fig. 7. Unsupervised learning with Denoiser using DNN

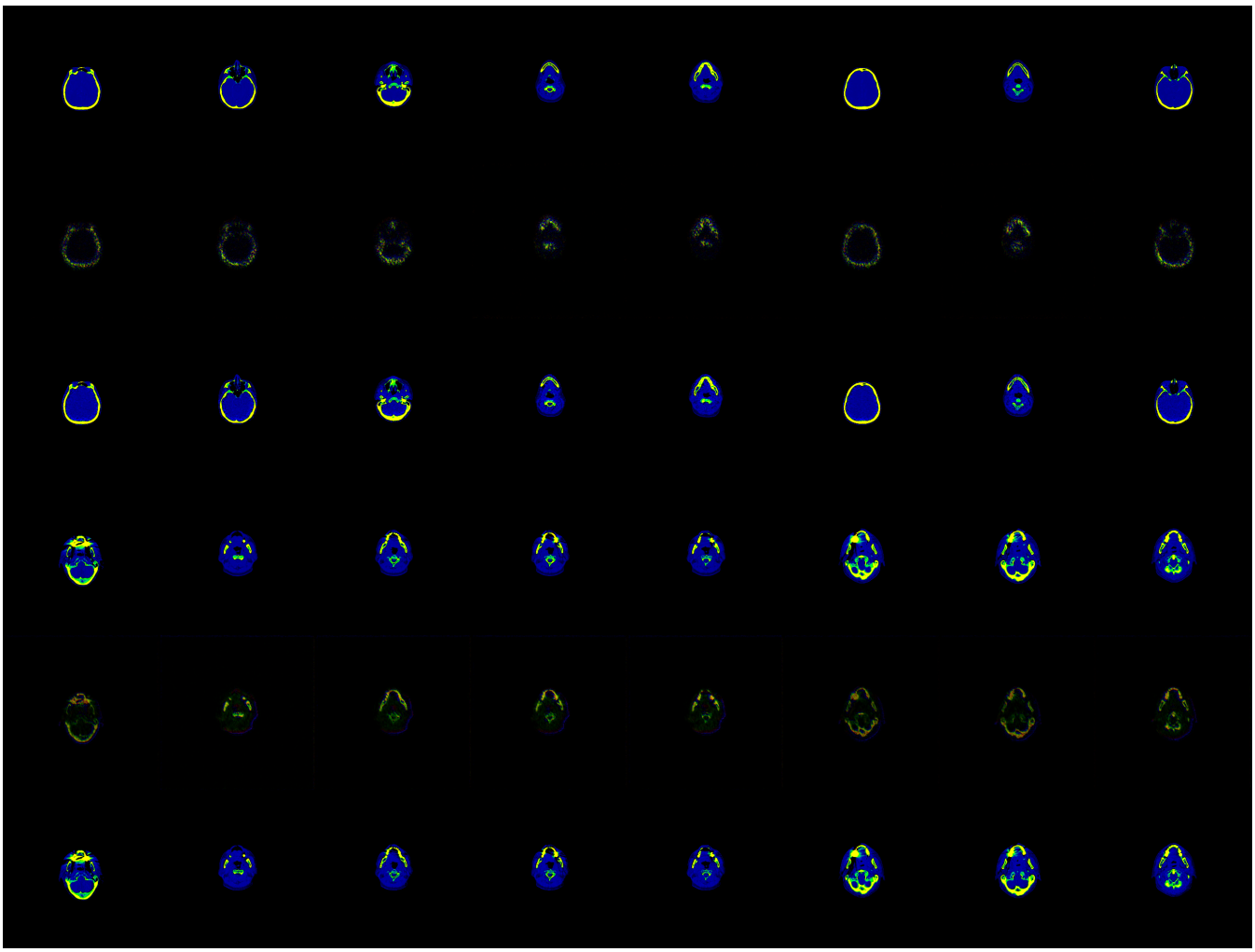

Fig. 8. Unsupervised learning with Denoiser using U-Net modifying Blurring technique to sharpening technique.