Deep Learning-Based Denoising in Brain Tumor CHO PET: Comparison with Traditional Approaches

Yucheng Zhang 1,2,†, Shuo Xu 1,2,†, Hongjia Li 3,†, Ziren Kong 4, Xincheng Xiang 1,2, Xin Cheng 5,* and Shaoyan Liu 4,*

1 Institute of Nuclear and New Energy Technology, Tsinghua University, Beijing 100084, China; yucheng-18@mails.tsinghua.edu.cn (Y.Z.); xu-s19@mails.tsinghua.edu.cn (S.X.);
inetxxc@mail.tsinghua.edu.cn (X.X.)
2 Beijing Key Laboratory of Nuclear Detection Technology, Beijing 100084, China
3 Department of Nuclear Medicine, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China; li-hj13@tsinghua.org.cn
4 Department of Head and Neck Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China; kongziren@pumc.edu.cn
5 Department of Nuclear Medicine, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China
* Correspondence: pumch_chengxin@126.com (X.C.); shaoyanliu bj@263.net (S.L.)
† These authors contributed equally to this work.

Abstract: 18F-choline (CHO) PET image remains noisy despite minimum physiological activity in the normal brain, and this study developed a deep learning-based denoising algorithm for brain tumor CHO PET. Thirty-nine presurgical CHO PET/CT data were retrospectively collected for patients with pathological confirmed primary diffuse glioma. Two conventional denoising methods, namely, block-matching and 3D filtering (BM3D) and non-local means (NLM), and two deep learning-based approaches, namely, Noise2Noise (N2N) and Noise2Void (N2V), were established for imaging denoising, and the methods were developed without paired data. All algorithms improved the image quality to a certain extent, with the N2N demonstrating the best contrast-to-noise ratio (CNR) (4.05 ± 3.45), CNR improvement ratio (13.60% ± 2.05%) and the lowest entropy (1.68 ± 0.17), compared with other approaches. Little changes were identified in traditional tumor PET features including maximum standard uptake value (SUVmax), SUVmean and total lesion activity (TLA), while the tumor-to-normal (T/N ratio) increased thanks to smaller noise. These results suggested that the N2N algorithm can acquire sufficient denoising performance while preserving the original features of tumors, and may be generalized for abundant brain tumor PET images.

Keywords: CHO; PET; denoising; deep learning; evaluation

1. Introduction

Positron emission tomography (PET) is an emerging imaging modality that detects photons produced after positron annihilation from a radionuclide tagged substrate, and has been widely applied to evaluate the tissue metabolism at a molecular level. 18F-fluorodeoxyglucose (FDG) is the most commonly used radiotracer due to the altered glucose metabolism of tumors; however, FDG displays high background activity in the normal brain because of the abundant glucose consumption, and this hinders the clinical application of FDG to evaluate brain tumors. Radiolabeled choline (CHO), therefore, has been developed as an alternative PET tracer to assess lipid metabolism, since the rapidly proliferating cells typically display elevated phospholipid production (resulting in increased choline uptake) while the physiological activity of the normal brain is low. As a result, CHO PET of the central nervous system allows a clearer visualization of
the metabolic tumor. However, the images remain noisy due to the limited number of photons captured during the PET scan, and higher counts of detected photons rely on a higher tracer dose and longer scanning time, which may give rise to the potential ionizing radiation damage to patients and medical staff or decrease the examination efficiency due to the long scanning duration. Therefore, developing a denoising method for CHO PET images of brain tumors that can produce a clearer image and retain the tumor property, would facilitate the clinical interpretation of PET data and facilitate the diagnosis of brain tumors.

Some traditional denoising methods have been applied for PET image noise reduction. Clinically, the Gaussian filter is a basic and simple approach, which convolves the reconstructed image with a Gaussian kernel. However, the Gaussian filter smooths out noise and detailed structures at the same time so this method is not edge-preserving. Multiple conventional algorithms, including bilateral filter [1], non-local means (NLM) [2], wavelet filter [3], block-matching and 3D filtering (BM3D) [4] and image-guided filter, have been developed to reduce the PET image noise while trying to preserve the details, but their clinical utility remains to be improved.

Over the past several years, deep learning techniques in image processing have provided novel approaches to acquiring high-quality images in the field of nuclear medicine imaging. Supervised networks require paired datasets for input and target, and have been proposed for PET image noise reduction, employing the high-quality images to restore the noisy images [5–7]. However, the feasibility to acquire clean PET data is limited in clinical practice, and it takes additional works to obtain paired datasets which is a burden for clinical work. Therefore, unsupervised and self-supervised learning techniques, which can train the models without labels, have been applied to PET imaging. Deep image prior (DIP) [8], which employs random noise and a corrupted image as the input and target without pre-training, shows the CNN’s intrinsic ability of learning structure information from a single corrupted image, and has been developed for FDG PET denoising with the patient’s prior image as the input [9–11]. Meanwhile, networks trained by large datasets have been reported to be useful for improving PET image quality without clean data, such as Noise2Noise (N2N) and Noise2Void (N2V). N2N [12] requires two datasets that have independent zero-mean noise, and have been verified to improve the PET image quality to a similar extent compared with supervised methods (e.g., Noise2Clean) [13,14]. N2V [15] employs a noisy dataset masked by a blind-plot network as the input and the same original dataset as the corresponding target, and has been applied for simulated as well as public brain PET data [16]. Notably, the majority of deep learning denoising methods are developed for FDG PET, whose images reflect basic anatomical structures and can be corrected with MRI, while CHO PET has a lower background activity in almost all normal brain structures and highlight the metabolic tumors. Therefore, networks trained by FDG PET images prior are not applicable for CHO PET and an alternative denoising algorithm is required.

This study developed a deep learning-based denoising algorithm for brain tumor CHO-PET that exhibits sufficient performance while preserving the detail of lesions. The method was explored and validated without a paired dataset, aiming to expand its generalizability for abundant retrospective images.

2. Materials and Methods
2.1. Patients

This was a retrospective investigation on a prospective CHO PET cohort which was approved by the Institutional Review Board of Peking Union Medical College Hospital (PUMCH) (ethics code ZS-2660), and informed consent was collected from all patients. Eligible criteria of the current study include those: (1) of age $\geq 18$ years and Karnofsky Performance Score (KPS) $\geq 70$; (2) suspected of having a primary brain tumor and being planned for surgery; (3) who underwent head CHO PET/CT prior to surgery; (4) with histopathological proof that the brain lesions are primary diffuse glioma; and (5) who had no anti-tumor
treatment prior to PET/CT or surgery. Finally, 39 patients with pathological-confirmed primary diffuse glioma were enrolled for CHO PET/CT and included in the current study. The role of CHO PET quantitative parameters to distinguish the World Health Organization (WHO) grade and molecular markers were reported previously [17,18], and the current study focused on the denoising of CHO PET image.

2.2. CHO PET/CT Data Acquisition and Tumor Segmentation

$^{18}$F-fluoroethylcholine (CHO) was synthesized as described previously, and a dose of 5.55 MBq (0.15 mCi) CHO per kilogram of body weight was given intravenously. PET/CT was acquired 40–60 min after CHO injection on a Biograph 64 TruePoint TrueV PET/CT system (Siemens Medical Solutions, Erlangen, Germany). The original image was acquired with a slice thickness of 3 mm and interpolated into the DICOM data with a matrix of $336 \times 336 \times 148$ as a standard protocol, with the physical size of each pixel being approximately $1 \times 1 \times 1.5$ mm. The finally output DICOM data were processed by Gaussian post-filtering. Standard uptake value (SUV) of each pixel normalized by body weight and decay factor was subsequently calculated.

Tumors were semiautomated segmented using 3D slicer 4.10.2 (https://www.slicer.org/, accessed on 13 November 2019) by a nuclear medicine physician as previously reported. Briefly, three spherical reference regions of interest (ROI) were first manually placed on the contralateral normal cortex to calculate the maximum and mean SUV ($N_{\text{max}}$ and $N_{\text{mean}}$) of the normal brain. The tumoral ROI was semiautomatically defined for regions with SUV/$N_{\text{mean}}$ > 2.0 and SUV/$N_{\text{max}}$ > 1.0 if the lesion displayed significant CHO activity, or delineated on the CT image and co-registered to CHO PET image if the lesions did not reveal significant CHO uptake. Manual editing was performed to ensure the continuity of the ROI and to remove the structures with physiological CHO uptake.

The brain was segmented based on CT image with the threshold to delineate the skull and floodfill function of OpenCV (version 4.5.3, https://opencv.org/, accessed on 6 October 2021) to fill the interior region (which represented the brain tissue). The CT-based brain segmentation was co-registered to the PET data using the resize function of OpenCV, and the normal brain was defined by subtracting the tumor segmentation from the brain segmentation.

2.3. Tumor Feature Definition

Five traditional features, namely, SUVmax, SUVmean, metabolic tumor volume (MTV), total lesion activity (TLA) and tumor-to-normal contralateral cortex activity ratio ($T/N$ ratio), were defined to quantify the metabolic characteristics of the tumor [18]. SUVmax, SUVmean and TLA represent the maximum, mean and total radioactivity of the tumoral ROI, while the $T/N$ ratio indicates the ratio of the tumor SUVmax to mean SUV of the contralateral brain ($N_{\text{mean}}$). MTV represents the volume of ROI and remained unchanged before and after postprocessing. Changes of SUVmax, SUVmean, TLA and $T/N$ ratio during postprocessing were calculated to reflect the influence of denoising to the original feature of the tumor, and Pearson correlation coefficients of SUVmax, SUVmean, MTV and TLA with denoising performance were computed to evaluate whether the denoising was tumor feature-dependent.

2.4. Denoising (Postprocessing)

Two conventional denoising methods (BM3D and NLM) and two deep learning-based approaches (N2N and N2V) were established for imaging denoising. BM3D and NLM are conventional denoising approaches that are supposed to be the basic reference in the field of image restoration. NLM filtering uses the weighted average value of similar blocks with square neighborhood of 10, the search window size of 21 and the Gaussian filtering parameter of 5. BM3D realized denoising by nonlocal-based grouping, collaborative filter and aggregation, with the standard deviation of additive white Gaussian noise set to 15.
N2N and N2V employed U-Net architecture as shown in Figures 1 and 2. The retrospective nature of the study only allows the original PET image to be the target, and N2N added Gaussian noise to create the input image, while N2V masked each input patch with blind-spot networks to implement the network possible. L2 loss function was minimized by Adam optimizer and utilized for both algorithms.

**Figure 1.** The network architecture of N2N network. Abbreviations: Conv, convolution; Deconv, deconvolution; ReLU, rectified linear unit; LeakyReLU, leaky rectified linear unit.

**Figure 2.** The network architecture of N2V network. Abbreviations: Conv, convolution; Deconv, deconvolution; BN, batch normalization; ReLU, rectified linear unit.
2.5. Denoising Evaluation

The contrast-to-noise ratio (CNR) shows the ratio of the contrast between tumor region and the normal brain of the same patient, defined as

\[
\text{CNR} = \frac{M_{\text{tumor}} - M_{\text{norm}}}{\text{SD}_{\text{norm}}} \tag{1}
\]

where \(M_{\text{tumor}}\) stands for the mean pixel value of the tumor region while \(M_{\text{norm}}\) reflects the mean value of complement pixels (the pixels outside the tumor). \(\text{SD}_{\text{norm}}\) is the standard deviation of the complement pixels.

To evaluate the performance of different methods, CNR improvement ratio is subsequently defined

\[
\text{CNR improvement ratio} = \frac{\text{CNR}_{\text{denoised}} - \text{CNR}_{\text{original}}}{\text{CNR}_{\text{original}}} \times 100\% \tag{2}
\]

where \(\text{CNR}_{\text{denoised}}\) and \(\text{CNR}_{\text{original}}\) present the CNR of denoised data and original PET data, respectively.

Inspired by the concept of thermodynamics, Shannon [19] proposed entropy as a measure of information. Entropy \((H)\) presents the statistical randomness of the gray-scale image and characterizes the image texture, defined as

\[
H = -\sum_{x \in \chi} p(x) \log_2[p(x)] \tag{3}
\]

where \(p(x)\), calculated by the gray level histogram, represents the normalized probability density of the pixels which have the value of \(x\), \(\chi\) is the scale of gray levels.

3. Results

3.1. Baseline Characteristics

There were 38.5% \((n = 15)\), 25.6% \((n = 10)\), 35.9% \((n = 14)\) patients diagnosed with WHO grade II, III and IV primary diffuse gliomas, respectively, and the CHO activity of the lesions increased progressively with tumor grade. The baseline characteristics of the 39 enrolled patients were displayed in Table 1.

| Characteristics | Grade II \((n = 15)\) | Grade III \((n = 10)\) | Grade IV \((n = 14)\) | \(p\) Value |
|-----------------|---------------------|---------------------|---------------------|-----------|
| Age             | 42.87 ± 9.96        | 47.70 ± 16.89       | 55.29 ± 12.47       | 0.045     |
| Sex             | /                   | /                   | /                   | 0.153     |
| Male            | 6 (40.0%)           | 8 (80.0%)           | 8 (57.1%)           | /         |
| Female          | 9 (60.0%)           | 2 (20.0%)           | 6 (42.9%)           | /         |
| SUVmax          | 0.62 ± 0.65         | 1.59 ± 1.16         | 2.69 ± 0.92         | <0.001    |
| SUVmean         | 0.141 ± 0.202       | 0.512 ± 0.438       | 0.877 ± 0.365       | <0.001    |
| MTV             | 34.0 ± 27.7         | 36.1 ± 32.4         | 56.6 ± 40.6         | 0.171     |
| TLA             | 6.20 ± 15.55        | 27.30 ± 38.16       | 50.91 ± 46.25       | 0.006     |

Notes: Fisher’s exact test or analysis of variance (ANOVA) test, as appropriate, were applied to calculate the statistical difference. Abbreviations: SUV, standardized uptake value; MTV, metabolic tumor volume; TLA, total lesion activity.

3.2. Computation Time

The training time and parameters for N2N and N2V, and the computation time for N2N, N2V, BM3D and NLM are shown in Table 2. We utilized an integral image algorithm with memory optimization [20] to accelerate NLM, while other algorithms (BM3D, N2N and N2V) were accelerated by GPU. This experiment was carried out on a computer with NVIDIA GeForce RTX3070 GPU and Intel(R) Core(TM) i7-10700F CPU @ 2.90 GHz. According to the computation time, BM3D and NLM are difficult to meet the needs of
real-time computing, and deep learning-based methods (N2N and N2V) are faster and may have wider application scenarios.

Table 2. Computation time for N2N, N2V, BM3D and NLM.

| Algorithm | Training Time (Dataset: 39 Patients) | Parameters | Computation Time (Each Patient) |
|-----------|--------------------------------------|------------|---------------------------------|
| N2N       | 26.0 h                               | 2.63 MB    | 25.44 s (GPU)                   |
| N2V       | 30.2 h                               | 14.5 MB    | 27.24 s (GPU)                   |
| BM3D      | /                                    | /          | 112.03 s (GPU)                  |
| NLM       | /                                    | /          | 255.72 s (CPU)                  |

N2N: optimizer: ADAM (default = [0.9, 0.99, 1 × 10⁻⁸]); batch size: 50; epoch: 1000; N2V: optimizer: ADAM (default = [0.9, 0.99, 1 × 10⁻⁸]); batch size: 50; epoch: 500.

3.3. Denoising Performance

All algorithms improved the image quality to a certain extent, and Figure 3 shows the comparison of the input image and denoised results. Both traditional algorithms (NLM and BM3D) enhanced the image quality to a similar extent, with mean CNRs of 3.85 ± 3.06 and 3.86 ± 3.06, respectively, while BM3D displayed a higher CNR improvement ratio (1.62% ± 3.41%) than NLM (1.37% ± 3.03%). On the other hand, deep learning methods (N2N and N2V) presented higher CNRs and CNR improvement ratios than NLM and BM3D. N2V resulted in the CNR at 3.89 ± 3.12 and the CNR improvement ratio at 7.75% ± 2.07%. N2N reflected clearer output images with higher CNR (4.05 ± 3.45) and CNR improvement ratio (13.60% ± 2.05%), both of which outperformed other methods. Values and distributions of CNR and CNR improvement ratio are demonstrated in Table 3 and Figure 4, respectively.

Figure 3. The input and denoised PET images of a WHO grade IV tumor (A), a WHO grade III tumor (B), and the normal brain region of a brain tumor patient (C), with the window width set as the maximum and minimum value of each image. In accordance with their mathematical definition, NLM and BM3D blurred the speckle and the tumor simultaneously, which decreased the noise of the normal brain but also removed the details of the tumor. Conversely, N2V and N2N removed the speckles to provide clearer images, and the tumor structure was better preserved. Among the four algorithms, N2N exhibited the clearest normal brain and better protected tumor details.
Table 3. Values of CNRs, CNR improvement ratios and entropy presented by N2N, N2V, BM3D and NLM.

| Denoising Parameters | N2N       | N2V       | BM3D      | NLM       |
|----------------------|-----------|-----------|-----------|-----------|
| CNR                  | 4.05 ± 3.45 | 3.89 ± 3.12 | 3.86 ± 3.06 | 3.85 ± 3.06 |
| CNR improvement ratio| 13.60% ± 2.05% | 7.75% ± 2.07% | 1.62% ± 3.41% | 1.37% ± 3.03% |
| Entropy              | 1.68 ± 0.17 | 2.42 ± 0.16 | 1.99 ± 0.19 | 2.05 ± 0.20 |

Figure 4. Distributions of CNRs (A) and CNR improvement ratios (B) for the results of BM3D, N2N, N2V and NLM.

Figure 5 showed the distributions of entropy for different methods. Higher entropy is suggestive of more information presented in the image, and N2V preserved more information with higher entropy (2.42 ± 0.16) than other methods. From another perspective, higher entropy is also correlated with a disordered image, and therefore, N2N exhibited the best denoising performance with the lowest entropy (1.68 ± 0.17).

Figure 5. Distributions of entropy for BM3D, NLM, N2N and N2V results.
3.4. Influence of Denoising on Tumor Features

SUVmax, SUVmean, TLA and T/N ratio are the most crucial parameters for malignancy evaluation in clinical settings, and minimum changes were identified in SUVmax, SUVmean, and TLA. SUVmax measures the maximum radioactivity of the tumor region, which represents the highest malignancy, and this remained unchanged after denoising (change ratio of 6.29% for N2N, 4.77% for N2V, 3.02% for BM3D, and 4.86% for NLM). Little changes were noticed in the SUVmean, which reflects the average radioactivity of the tumor region, with the change ratio of 0.85% for N2N, 0.58% for N2V, 2.25% for BM3D, and 1.52% for NLM. N2N and N2V reflected smaller changes than conditional methods in TLA, while all these methods exhibited little changes within 3% compared to the original images. In accordance with the denoising purpose, the T/N ratios increased for all algorithms, with the improvement ratio of 20.93% and 5.46% for N2N and N2V, while fewer changes were identified in conventional methods (2.54% for BM3D and 1.51% for NLM). Table 4 showed the mean improvement ratio of tumor features with outliers detected and removed. Figure 6 showed the distributions of tumor features for denoised data.

**Table 4.** The change ratios of SUVmax, SUVmean, TLA and T/N ratio for N2N, N2V, BM3D and NLM.

| Change Ratios | N2N   | N2V   | BM3D  | NLM   |
|---------------|-------|-------|-------|-------|
| SUVmax        | 6.29% | 4.77% | 3.02% | 4.86% |
| SUVmean       | 0.85% | 0.58% | 2.25% | 1.52% |
| TLA           | 0.85% | 0.58% | 2.25% | 1.52% |
| T/N ratio     | 20.93%| 5.46% | 2.54% | 1.51% |

Figure 6. Distributions of the improvement ratios of SUVmax (A), SUVmean (B), TLA (C) and T/N ratio (D).
There is also no correlation of CNR improvement ratio with tumor size, SUVmax, SUVmean and TLG, with the Pearson correlation coefficient ranging from \(-0.110\) to \(-0.031\) for CNR improvement ratio. Correlation coefficients of CNR and CNR improvement ratio with tumor features are displayed in Table 5.

Table 5. Correlation of CNR and CNR improvement ratio with MTV, SUVmax, SUVmean and TLA.

| Correlation   | MTV          | SUV Max        | SUV Mean        | TLA          |
|---------------|--------------|----------------|-----------------|--------------|
| N2N Improvement ratio CNR | -0.032 (p = 0.660) | -0.071 (p = 0.323) | -0.073 (p = 0.311) | -0.058 (p = 0.415) |
|               | 0.431 (p < 1 × 10^{-5}) | 0.425 (p < 1 × 10^{-5}) | 0.461 (p < 1 × 10^{-5}) | 0.449 (p < 1 × 10^{-5}) |
| N2V Improvement ratio CNR | -0.031 (p = 0.666) | -0.066 (p = 0.355) | -0.066 (p = 0.355) | -0.058 (p = 0.421) |
|               | 0.443 (p < 1 × 10^{-5}) | 0.453 (p < 1 × 10^{-5}) | 0.498 (p < 1 × 10^{-5}) | 0.473 (p < 1 × 10^{-5}) |
| NLM Improvement ratio CNR | -0.110 (p = 0.126) | -0.107 (p = 0.134) | -0.111 (p = 0.120) | -0.103 (p = 0.151) |
|               | 0.441 (p < 1 × 10^{-5}) | 0.456 (p < 1 × 10^{-5}) | 0.490 (p < 1 × 10^{-5}) | 0.468 (p < 1 × 10^{-5}) |
| BM3D Improvement ratio CNR | -0.036 (p = 0.618) | -0.074 (p = 0.302) | -0.071 (p = 0.325) | -0.059 (p = 0.407) |
|               | 0.441 (p < 1 × 10^{-5}) | 0.451 (p < 1 × 10^{-5}) | 0.488 (p < 1 × 10^{-5}) | 0.470 (p < 1 × 10^{-5}) |

Notes: The values of tumor features change ratios were calculated with outliers that differed from the median by more than three scaled median absolute deviations detected and removed.

4. Discussion

CHO PET is an effective molecular-level lipid metabolism evaluation approach for tumor diagnosis. PET images are recognized by limited spatial resolution, and this paper established conventional and deep learning methods for CHO PET denoising. All proposed algorithms reduced noise to a certain extent, and deep learning methods (N2N and N2V) preserved more details of the tumor region compared to conventional filters (BM3D and NLM). N2N had superior performance among all algorithms, with a higher CNR improvement ratio of 13.60% ± 2.05%, lower entropy of 1.68 ± 0.17 and fewer changes in tumor features (SUVmax, SUVmean and TLA). This proposed method provided clearer images for physician’s diagnosis with good preservation of the representative features in tumor region, which can be expected to be applied to other low background activity radiotracer data.

The conventional filters reduced noise in the spatial domain, which calculates the pixel value by weighted average of correlated pixels. Though there are algorithms, such as BM3D and NLM, to search the correlated pixels and estimate the weight of the pixels, they cause blurring of results when the noise level is high. As presented in Figure 3, the normal brain regions of conventional method results have unwanted indistinction. Deep learning-based methods, on the other hand, are data-driven and learn to remove noise from images directly instead of filtering the neighborhood pixels to smooth both noise and detail information. Although it has narrower application scenarios, a well-trained deep learning algorithm may present a better performance in executing a specific assignment. In our study, the deep learning-based networks outperformed the conventional filters with higher CNRs and CNR improvement ratios, indicating the capability of deep learning for the denoising of brain tumor CHO PET data.

Previous studies are always based on FDG PET data [21], for FDG is the most widely applied radiotracer with an enormous amount of application. The injection dose of CHO is similar to the dose of FDG, but the uptake approaches of the two kinds of tracer have essential differences. FDG, as a glucose analog that can be absorbed but not further metabolized, can detail the major brain structures due to the significant glucose assumption in the cortex. CHO, on the other hand, reflects phospholipid membrane production, which is generally low in both the cortex and medulla. Therefore, CHO PET images present low background metabolism in the whole brain, that are supposed to obtain less information from brain tissue. Comparing CHO PET with FDG PET, the total uptake of radiation tracer is lower and the characteristics of the images and noise are different. Therefore, supervised denoising networks trained by FDG PET images may have difficulty
with the generalization to CHO PET images, which have different characteristics compared to FDG.

A variety of neural networks have been designed to produce denoised images, and the majority of them are trained by paired data such as short-scan-time and long-scan-time data. Although these approaches are expected to have better performance on the denoising result, the limitation of these methods still narrows their practical application. Supervised denoising networks rely on the experiments bringing out long-scan-time data or down-sampled data to generate paired datasets. However, the short-scan-time data or down-sampled data are not always accessible for the retrospective dataset which are post-processed by Gaussian filters. Hence, the techniques that require ground-truth data are difficult to be adopted in clinical practice. In contrast, N2N and N2V denoising networks are self-supervised/unsupervised learning approaches, and they have the potential capability of clinical generalization. Our denoising algorithm was developed from the finally outputted DICOM data, which is different from some of the previous studies that denoise as one of the post-process steps. Processing the final DICOM data provides clearer images with important tumor features preserved, and it gives us opportunities to deal with the abundant retrospective (original-existing) data. In addition to CHO PET, amino acid PET was also recognized for its low background activity and high tumor-to-normal contrast in brain tumors, and our denoising method may also be utilized for amino acid PET to produce a clearer image.

Entropy is considered to represent the variance of an image, with a higher value indicating a higher image disorder of the same subject. For N2N network, which tend to learn the low-frequency signal to fulfill the accuracy of norm regression, the output of this network presented low entropy. The N2V approach is based on the assumption that the noise of images is pixel-wise independent, so that it can filter out the unpredictable noise with the structured noise left. While the uptake of normal brain tissue is assumed to be low and evenly distributed, the high uptake spots in the normal brain region are considered to be noise, which seems not to be pixel-wise independent. As a result, the entropy of the N2V output was higher. BM3D and NLM had similar entropy scores which were lower than N2V, for they sacrificed the details compared to N2V.

The denoising methods reflected little changes in tumor features such as SUVmax, SUVmean and TLA, suggesting a minimum influence on tumor grading and threshold of diagnosis. Therefore, the methods are feasible for the clinical experience and research findings. Conversely, there is also no correlation of the CNR improvement ratio with the tumor features, indicating the denoising algorithms are robust and the denoising performances of the normal brain are not influenced by the intrinsic tumor feature.

This study has a few limitations. First, ground-truth images or reference images are not available for the evaluation of denoising results, so the proposed method was not compared to supervised networks. Second, the assumption of N2N network that Gaussian noise was added to the noisy images is not precise, so the method can be improved in the future research. Third, the network has not been verified for other tracers, different types of tumors and other tissues. Forth, the denoising methods for 3D data will be developed, and developing new networks and comparing with more effective algorithms will be the topic of our research in the future.

5. Conclusions

In this work, deep learning-based networks trained by full dose original images were applied to CHO PET denoising, demonstrating a higher overall image quality compared with conventional approaches. The proposed N2N network was more effective in removing the noise with tumor features, keeping the detailed structures preserved, compared to NLM, BM3D and N2V.
Author Contributions: Conceptualization, Y.Z. and Z.K.; Data curation, H.L.; Formal analysis, Y.Z. and S.X.; Methodology, Y.Z. and S.X.; Project administration, Z.K.; Resources, H.L., X.C. and S.L.; Visualization, Y.Z., S.X. and X.X.; Writing—original draft, Y.Z. and S.X.; Writing—review & editing, H.L., Z.K., X.C. and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (grant number 81201121) and the Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (grant numbers 2018-I2M-3-001).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of Peking Union Medical College Hospital.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The code and model are available online (https://github.com/xushuo0629, accessed on 10 May 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hofheinz, F.; Langner, J.; Beuthien-Baumann, B.; Oehme, L.; Steinbach, J.; Kotzerke, J.; Van den Hoff, J. Suitability of bilateral filtering for edge-preserving noise reduction in PET. EJNMMI Res. 2011, 1, 23. [CrossRef] [PubMed]
2. Dutta, J.; Leahy, R.M.; Li, Q. Non-Local Means Denoising of Dynamic PET Images. PLoS ONE 2013, 8, e81390. [CrossRef] [PubMed]
3. Boussion, N.; Le Rest, C.C.; Hatt, M.; Visvikis, D. Incorporation of wavelet-based denoising in iterative deconvolution for partial volume correction in whole-body PET imaging. Eur. J. Nucl. Med. Mol. Imaging 2009, 36, 1064–1075. [CrossRef] [PubMed]
4. Chen, L.L.; Gou, S.P.; Yao, Y.; Bai, J.; Jiao, L.; Sheng, K. Denoising of low dose CT image with context-based BM3D. In Proceedings of the 2016 IEEE Region 10 Conference (TENCON), Singapore, 22–25 November 2016; pp. 682–685. [CrossRef]
5. Gong, K.; Guan, J.; Liu, C.-C.; Qi, J. PET Image Denoising Using a Deep Neural Network Through Fine Tuning. IEEE Trans. Radiat. Plasma Med. Sci. 2019, 3, 153–161. [CrossRef] [PubMed]
6. Chen, K.T.; Gong, E.; de Carvalho Macruz, F.B.; Xu, J.; Boumis, A.; Poston, K.L.; Sha, S.J.; Greicius, M.D.; Mormino, E.; et al. Ultra-Low-Dose 18F-Florbetaben Amyloid PET Imaging Using Deep Learning with Multi-Contrast MRI Inputs. Radiology 2019, 290, 649–656. [CrossRef] [PubMed]
7. Schaefferkoetter, J.; Yan, J.; Ortega, C.; Sertic, A.; Lechtman, E.; Eshet, Y.; Metser, U.; Veit-Haibach, P. Convolutional neural networks for improving image quality with noisy PET data. EJNMMI Res. 2020, 10, 105. [CrossRef] [PubMed]
8. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Deep Image Prior. Int. J. Comput. Vis. 2020, 128, 1867–1888. [CrossRef]
9. Hashimoto, F.; Ohba, H.; Ote, K.; Teramoto, A.; Tsukada, H. Dynamic PET Image Denoising Using Deep Convolutional Neural Networks Without Prior Training Datasets. IEEE Access 2019, 7, 96594–96603. [CrossRef]
10. Gong, K.; Catana, C.; Qi, J.; Li, Q. PET Image Reconstruction Using Deep Image Prior. IEEE Trans. Med. Imaging 2019, 38, 1655–1665. [CrossRef] [PubMed]
11. Schaefferkoetter, J.; Yan, J.; Ortega, C.; Sertic, A.; Lechtman, E.; Eshet, Y.; Metser, U.; Veit-Haibach, P. Convolutional neural networks for improving image quality with noisy PET data. EJNMMI Res. 2020, 10, 105. [CrossRef] [PubMed]
12. Lehtinen, J.; Munkberg, J.; Hasselgren, J.; Laine, S.; Karras, T.; Aittala, M.; Aila, T. Noise2Noise: Learning image restoration without clean data. arXiv 2018, arXiv:1803.04189.
13. Yie, S.Y.; Kang, S.K.; Hwang, D.; Lee, J.S. Self-supervised PET Denoising. Nucl. Med. Mol. Imaging 2020, 54, 299–304. [CrossRef] [PubMed]
14. Chan, C.; Zhou, J.; Yang, L.; Qi, W.; Asma, E. Noise to Noise Ensemble Learning for PET Image Denoising. In Proceedings of the 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), Manchester, UK, 26 October–2 November 2019; pp. 1–3. [CrossRef]
15. Krull, A.; Buchholz, T.-O.; Jug, F. Noise2Void—Learning Denoising From Single Noisy Images. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 2124–2132. [CrossRef]
16. Song, T.-A.; Yang, F.; Dutta, J. Noise2Void: Unsupervised denoising of PET images. Phys. Med. Biol. 2021, 66, 214002. [CrossRef] [PubMed]
17. Kong, Z.; Jiang, C.; Liu, D.; Chen, W.; Ma, W.; Cheng, X.; Wang, Y. Quantitative Features From CHO PET Distinguish the WHO Grades of Primary Diffuse Glioma. Clin. Nucl. Med. 2021, 46, 103–110. [CrossRef] [PubMed]
18. Kong, Z.; Zhang, Y.; Liu, D.; Liu, P.; Shi, Y.; Wang, Y.; Zhao, D.; Cheng, X.; Wang, Y.; Ma, W. Role of traditional CHO PET parameters in distinguishing IDH, TERT and MGMT alterations in primary diffuse gliomas. Ann. Nucl. Med. 2021, 35, 493–503. [CrossRef] [PubMed]
19. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* 1948, 27, 379–423. [CrossRef]

20. Froment, J. Parameter-Free Fast Pixelwise Non-Local Means Denoising. *Image Process. Line* 2014, 4, 300–326. [CrossRef]

21. Liu, J.; Malekzadeh, M.; Mirian, N.; Song, T.-A.; Liu, C.; Dutta, J. Artificial Intelligence-Based Image Enhancement in PET Imaging: Noise Reduction and Resolution Enhancement. *PET Clin.* 2021, 16, 553–576. [CrossRef] [PubMed]