

Deep Learning-Based Fast TOF-PET Image Reconstruction Using Direction Information

Kibo Ote

Hamamatsu Photonics K.K.

Fumio Hashimoto (✉ fumio.hashimoto@crl.hpk.co.jp)

Hamamatsu Photonics K.K. <https://orcid.org/0000-0003-2352-0538>

Research Article

Keywords: Positron emission tomography, Image reconstruction, Deep learning, Time-of-flight, Direction information

Posted Date: December 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1025979/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Radiological Physics and Technology on February 8th, 2022. See the published version at <https://doi.org/10.1007/s12194-022-00652-8>.

1 **Deep learning-based fast TOF-PET image reconstruction**

2 **using direction information**

3 Kibo Ote and Fumio Hashimoto

4 Central Research Laboratory, Hamamatsu Photonics K.K., Hamamatsu, 434-8601,
5 Japan.

6 Correspondence should be addressed to Kibo Ote and Fumio Hashimoto;

7 fumio.hashimoto@crl.hpk.co.jp, kibou@crl.hpk.co.jp

8 Phone: +81 53-586-7111

9 Fax: +81 53-586-6180

10 **Abstract**

11 Deep learning has attracted attention for positron emission tomography (PET)
12 image reconstruction task, however, it remains necessary to further improve the image
13 quality. In this study, we propose a novel CNN-based fast time-of-flight PET (TOF-
14 PET) image reconstruction method to fully utilize the direction information of
15 coincidence events. The proposed method inputs view-grouped histo-images into a 3D
16 CNN as a multi-channel image to use the direction information of coincidence events.
17 We evaluated the proposed method using Monte Carlo simulation data obtained from a
18 digital brain phantom. Compared to the case without it, when using direction
19 information, the peak signal-to-noise ratio and structural similarity were improved by
20 1.2 dB and 0.02, at a coincidence time resolution of 300 ps. The calculation times of the
21 proposed method were significantly faster than the conventional iterative
22 reconstruction. These results indicate that the proposed method improves both the speed
23 and image quality of TOF-PET image reconstruction.

24 Keywords: Positron emission tomography, Image reconstruction, Deep learning, Time-
25 of-flight, Direction information

26 **1. Introduction**

27 Positron emission tomography (PET) is a functional imaging tool for various
28 medical applications, such as oncology, cardiology, and neurology [1]. It has a unique
29 ability to quantitatively estimate radiotracer concentrations as low as picomolar
30 concentrations; however, the radiotracer concentration cannot be directly imaged from

31 a line of response measured by the coincidence detection of annihilation photons.
32 Therefore, an image reconstruction process is required to estimate the distribution of
33 the radiotracer concentration.

34 There are two main methods for image reconstruction: analytic and iterative
35 methods [2]. The analytic method is simple and fast, although it is sensitive to statistics
36 and prone to streak artifacts. The iterative method models the noise distribution and
37 reconstructs an image by iterative updating. The iterative method improves the signal-
38 to-noise ratio (SNR) of the reconstructed image better than the analytic method;
39 however, it is computationally expensive. Therefore, an image reconstruction method
40 that improves both the speed and SNR is desired.

41 Recently, deep learning has attracted attention for improving PET image quality
42 [3–9], especially for image reconstruction tasks [10–17]. As a pioneer, Zhu et al.
43 proposed an automated transform by manifold approximation (AUTOMAP) network
44 that can directly reconstruct a tomographic image from various sensing data [18]. The
45 versatility of AUTOMAP is owing to its unique network architecture, which first
46 applies a fully connected (FC) layer to sensing data multiple times, followed by a
47 convolution layer to feature maps multiple times. However, AUTOMAP was limited to
48 2D reconstruction because the FC layer has a large number of parameters. Häggström
49 et al. developed DeepPET, which is a deep convolutional encoder–decoder network for
50 directly reconstructing PET images from PET sinogram data [19]. Although DeepPET
51 provided less noisy images than the ordered subset expectation maximization algorithm
52 (OSEM) [20], it was limited to 2D reconstruction, similar to AUTOMAP. These
53 limitations were owing to the fact that the direct mapping of sensing data to image
54 format requires a large network.

55 To realize actual 3D image reconstruction in near real-time, Whiteley et al.
56 proposed FastPET, a deep learning-based fast image reconstruction specialized for
57 time-of-flight PET (TOF-PET) [21]. FastPET first converted the coincidence events
58 into a histo-image [22] format using TOF information, and then reconstructed the final
59 image volume from the histo-image using a 3D CNN. The histo-image is a more
60 suitable format for CNN than sinograms; hence, near real-time 3D image reconstruction
61 has been enabled. However, based on the theory of analytic TOF-PET image
62 reconstruction [23,24], FastPET may compromise the SNR because the histo-image
63 does not preserve the direction information of coincidence events.

64 In this study, we propose a deep learning-based fast TOF-PET image
65 reconstruction using directional information. The proposed method inputs view-
66 grouped histo-images [22] into a 3D CNN as a multi-channel image. Using the Monte
67 Carlo simulation data of a digital brain phantom, we verified that the proposed method
68 improves the SNR while maintaining near real-time image reconstruction.

69 **2. Histo-image generation**

70 The basis of analytic TOF-PET image reconstruction is deconvolution after
71 accumulating coincidence events into a histo-image format using TOF information. The
72 problem dealt with the process of accumulating events into a histo-image. Accordingly,
73 two methods were considered: the most likely annihilation position (MLAP) [21,23]
74 and confidence weighting (CW) [23] methods.

75 In both methods, the MLAP of the coincident event connecting two detector
76 elements whose coordinates are \vec{P}_1 and \vec{P}_2 is estimated using the TOF information.

77

$$\vec{P}_{MLAP} = \frac{\vec{P}_1 + \vec{P}_2}{2} + c \frac{\Delta t}{2} \frac{\vec{P}_1 - \vec{P}_2}{\|\vec{P}_1 - \vec{P}_2\|}, \quad (1)$$

78 where c and Δt denote the speed of light and TOF information, respectively. The
 79 MLAP method simply accumulates the event in the voxel nearest to \vec{P}_{MLAP} . The CW
 80 method accumulates the event as a line weighted by the TOF response function centered
 81 on \vec{P}_{MLAP} , as illustrated in Figure 1.

82 By mathematically modeling the variance of the voxel value of an analytically
 83 reconstructed image of a uniform disk phantom, it was concluded that the CW method
 84 is optimal in terms of SNR [23,24]. The MLAP method is not optimal because the high-
 85 resolution information in the vertical direction (Figure 1 (b)) of the coincidence event is
 86 lost after accumulating the events into the histo-image, using the MLAP method. In
 87 other words, the direction of the coincidence event contains information about
 88 resolution heterogeneity. Another reason to choose the CW method is continuity with
 89 conventional non-TOF-PET image reconstruction. In other words, the CW method
 90 tends to the non-TOF-PET image reconstruction method, as the coincidence time
 91 resolution (CTR) increases to infinity.

92 **3. Methodology**

93 In this study, we adopted an angular-view grouping [22,25] to introduce directional
 94 information in the TOF-PET image reconstruction using CNN.

95 **3.1. Angular-view grouping**

96 Although the CW method is optimal from the perspective of SNR, it is time
 97 consuming because it requires ray tracing. In this study, we introduce angular-view
 98 grouping [22,25] to implement direction information. In this scheme, the events are
 99 divided into N groups, depending on the angle of coincidence. The group of events is
 100 separately accumulated in N histo-images using MLAP methods. Figure 2 shows an
 101 example of angular-view grouping at $N = 8$. Using angular-view grouping, we can
 102 preserve the direction information of coincidence events as view-grouped histo-images
 103 without ray tracing. Note that the angular-view grouping in this study is performed in
 104 the azimuthal angle, and not in the oblique angle.

105 The azimuthal angle φ and view group n can be calculated with following
 106 equations:

$$107 \quad \varphi = \begin{cases} \cos^{-1} \frac{y_2 - y_1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}} & x_2 - x_1 \geq 0 \\ \pi - \cos^{-1} \frac{y_2 - y_1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}} & \text{otherwise} \end{cases}, \quad (2)$$

$$108 \quad n = \left\lceil \frac{\varphi}{\pi} N + 0.5 \right\rceil, \quad (3)$$

109 where $(x_1, y_1), (x_2, y_2)$ are 2D coordinates of \vec{P}_1 and \vec{P}_2 , respectively, and $\lceil \cdot \rceil$
 110 denotes an operator of the round down.

111 3.2. TOF-PET image reconstruction using CNN

112 Figure 3 shows an outline of the proposed method. Raw coincidence events are
 113 accumulated in N view-grouped histo-images by the above-mentioned angular-view

114 grouping. Subsequently, the 3D CNN receives N view-grouped histo-images and one
115 attenuation map as an $(N + 1)$ channel image, and outputs the final image. A 3D CNN
116 can use the direction information of coincidence events because the view-grouped histo-
117 images preserve the direction information of coincidence events. The view-grouped
118 histo-images are normalized for the detection efficiency variation. Then, the attenuation
119 and scatter are corrected inside the 3D CNN through end-to-end learning of mapping of
120 the view-grouped histo-images, which are not corrected for scatter and attenuation, to
121 the phantom images of ground truth.

122 Figure 4 illustrates the network architecture of the 3D CNN used in this study. We
123 employed a 3D U-net [26] architecture because it is suitable for medical image
124 processing. To increase the speed of inference, we reduce the trainable parameters of
125 the proposed architecture to 3% of the FastPET architecture. The 3D U-net comprises
126 an encoder, a decoder, and skip connections.

127 The encoder part extracts useful features for image reconstruction through
128 convolution, non-linear activation, and down-sampling. The combination of $3 \times 3 \times 3$,
129 3D convolution, and a leaky rectified linear unit (LReLU) is repeated twice before
130 down-sampling. Down-sampling is performed by a $3 \times 4 \times 4$, 3D convolution with
131 stride $(1, 2, 2)$, followed by LReLU. At each down-sampling stage, the x and y sizes of
132 the feature maps are halved, and the number of channels is doubled.

133 The decoder part reconstructs the final image from the feature maps through
134 convolution, nonlinear activation, and up-sampling. The combination of $3 \times 3 \times 3$, 3D
135 convolution and LReLU is repeated twice before up-sampling. Up-sampling is

136 performed by a $3 \times 4 \times 4$, 3D transpose convolution with stride (1, 2, 2), followed by
137 LReLU. At each up-sampling stage, the x and y sizes of the feature maps are doubled,
138 and the number of channels is halved. The final image is reconstructed by a $3 \times 3 \times 3$,
139 3D convolution with one channel output.

140 The feature maps of the encoder part before down-sampling are added to the feature
141 maps of the decoder part after up-sampling through a skip connection.

142 **4. Experimental setup**

143 **4.1. Simulation dataset**

144 Monte Carlo simulations were performed with our own Monte Carlo simulator
145 to create the dataset. Segmented brain MRI images of 20 normal subjects downloaded
146 from BrainWeb [27] were used to make digital brain phantoms. The contrast of
147 radioactivity between gray matter, white matter, and cerebrospinal fluid was set to 1 :
148 0.25 : 0.05 based on the [¹⁸F]FDG contrast. The attenuation coefficients of the soft
149 tissue and bone were set to 0.00958 mm^{-1} and 0.0151 mm^{-1} , respectively.

150 In this study, we used the brain-dedicated PET scanner described in [28] as the
151 detector arrangement for the simulation. A detector ring with a diameter of 486.83 mm
152 was constructed with detector units of 28 and 4 in the ring and axial directions,
153 respectively. Each detector unit had a 16×16 array of cerium-doped lutetium–yttrium
154 oxyorthosilicate (LYSO) crystals. The size of each LYSO crystal was $3.14 \text{ mm} \times 3.14$
155 $\text{mm} \times 20 \text{ mm}$. The image size was $70 \times 128 \times 128$ voxels with a voxel size of 3.221×3
156 $\times 3 \text{ mm}^3$. An energy resolution of 15% and an energy window of 400–650 keV were
157 assumed. A total of $181.12 \pm 6.08 \text{ M}$ counts, including scatter events, were collected for

158 each subject using 3D acquisition. CTR values of 100, 300, and 600 ps were simulated.

159 The number of ring was 72, including the gap between the detector units in the axial

160 direction. The maximum ring difference was set to ± 66 .

161 We split the 20 subjects into 15 and 5 for training and testing, respectively. In

162 addition, the training data were split into 12 for real training and 3 for validation to

163 monitor the validation loss during training.

164 **4.2. Network training**

165 We trained the 3D CNN for 500 epochs using the Adam optimizer with $\beta_1 = 0.5$.

166 In this study, the phantom images were used as the training label. The mean squared

167 error was used as the loss function. We considered 64 updates using the mini-batch with

168 a batch size of 32 as one epoch. We randomly cropped the $64 \times 64 \times 64$ sub-image from

169 the original $70 \times 128 \times 128$ image during training as data augmentation. We monitored

170 the loss of validation data during training and chose the model with the minimum

171 validation loss for testing. In the testing phase, the original $70 \times 128 \times 128$ image was

172 used as the network input.

173 **4.3. Evaluation**

174 We reconstructed the five test data using the list-mode dynamic row-action

175 maximum likelihood algorithm (List-DRAMA) [29,30] and the proposed method of $N =$

176 1, 2, 4, 8, and 16. List-DRAMA was performed with two main iterations and 40

177 iterations. List-DRAMA was calculated in parallel using eight cores of Intel Xeon

178 X5680 3.33 GHz. The proposed method was calculated with an NVIDIA Quadro P6000
 179 graphics board using Chainer 7.7.0 (<https://chainer.org/>).

180 For quantitative evaluation, we measured the peak signal-to-noise ratio (PSNR)
 181 and structural similarity (SSIM) of reconstructed images, \hat{x} , using a phantom image, x ,
 182 as a reference.

183 PSNR is an indicator of the voxel-level image similarity, and it was calculated
 184 as:

$$185 \quad \text{PSNR} = 10 \log_{10} \frac{\left(\max_{j \in R} x_j \right)^2}{\frac{1}{N_R} \sum_{j \in R} (\hat{x}_j - x_j)^2} \text{ (dB)}, \quad (4)$$

186 where j , R and N_R denote an index of the voxel, an entire brain region, and the
 187 number of voxels inside the entire brain region, respectively.

188 SSIM is an indicator of the image brightness, contrast, and structural
 189 similarities, and it was calculated as:

$$190 \quad \text{SSIM} = \frac{1}{N_R} \sum_{j \in R} \frac{(2\mu_{j\hat{x}}\mu_{jx} + c_1)(2\sigma_{j\hat{x}x} + c_2)}{(\mu_{j\hat{x}}^2 + \mu_{jx}^2 + c_1)(\sigma_{j\hat{x}}^2 + \sigma_{jx}^2 + c_2)}, \quad (5)$$

191 where μ_j and σ_j are the mean and standard deviation of a small local region
 192 (patch) around the j -th voxel, $\sigma_{j\hat{x}x}$ is the covariance between the patches around the j -th
 193 voxel of the reconstructed image and the phantom image. In this study, the patch size
 194 was $7 \times 7 \times 7$ voxels, $c_1 = (0.01L)^2$, and $c_2 = (0.03L)^2$, where L is the dynamic range
 195 of the phantom image.

196 5. Results and Discussion

Figure 5 shows the PSNR and SSIM relative to the number of views at CTRs of 100, 300, and 600 ps. The PSNR and SSIM increased as the number of views increased. This indicated that the direction information is beneficial in improving the SNR even with the deep learning-based image reconstruction method.

The PSNR at a CTR of 100 ps slightly decreased when the number of views reached 12 or more. It is known that the number of views required for TOF-PET image reconstruction decreases as CTR improves [25,31]. The number of views required for the analytic TOF-PET image reconstruction was obtained [25],

$$205 \quad N > 2\pi\sigma_T/d_0, \quad (6)$$

where σ_T is the standard deviation of the TOF response function, and d_0 denotes the required spatial resolution. If $\sigma_T = 6.37$ mm, which corresponds to a CTR of 100 ps, and $d_0 = 4.5$ mm, which corresponds to a 1.5 voxel width, then $N > 6$ is sufficient. The above results are consistent with this theory. Therefore, the optimal number of views for each TOF-PET scanner can be easily estimated using the above formula.

Figure 6 shows a comparison of reconstructed images between the proposed method and the other methods at a CTR of 300 ps. The results of FastPET correspond to the results of the proposed method using 1 view. The results of the proposed method are the case with eight views. The proposed method improved both PSNR and SSIM and provided sharper images than the other methods. The proposed method recovered a finer structure than that of the other methods, as indicated by a red square. These results indicate that the spatial resolution of the reconstructed image by the proposed method was improved by the high-resolution information in the vertical direction of coincidence events.

220 Table 1 shows a comparison of calculation times, PSNRs, and SSIMs between the
221 proposed method and the other methods for a CTR of 300 ps. The calculation time of
222 the proposed method is almost the same as that of FastPET. In addition, the proposed
223 method was three orders of magnitude faster than List-DRAMA. These results
224 indicate that the proposed method has the capability of near real-time TOF-PET
225 image reconstruction with high image quality.

226 One of the limitations of this study is that it is applied only for the simulation
227 dataset. We will collect experimental data for training neural networks using the brain
228 PET scanner. From this study, it seems that the findings of theory of image
229 reconstruction are also useful for deep learning-based methods. For example, view-
230 grouped histo-images could be beneficial in improving PET image quality using
231 unsupervised CNN framework [6–8,12].

232 In this study, we used the angular-view grouping with MLAP instead of CW. The
233 CW method is expected to improve the SNR from the principle of TOF-PET image
234 reconstruction, however, it is impractical from the perspective of calculation cost
235 because the strict calculation of CW requires ray-tracing event-by-event. The angular-
236 view grouping can be considered as the fast approximation of the CW method, and its
237 performance is similar to that of the CW method as the number of views increases. In
238 this study, we input the tuple of view-grouped histo-images and attenuation maps to
239 the 3D CNN similar to FastPET [21]. When no attenuation map is used for the CNN
240 input, the accuracies of scatter and attenuation corrections could be degraded because
241 the Compton scattering and photon attenuation are governed by the attenuation map.

242 Table 1: Comparison of calculation times, PSNRs, and SSIMs between the proposed method and other
243 methods at 300 ps CTR

Reconstruction method	Number of views	Calculation time (s)	PSNR (dB)	SSIM
List-DRAMA	-	731.2±9.02	18.71±0.41	0.818±0.011
FastPET	1	0.225±0.002	19.48±0.17	0.921±0.001
Proposed method	4	0.229±0.003	20.44±0.17	0.936±0.001
	8	0.235±0.003	20.95±0.33	0.941±0.002
	12	0.241±0.003	20.98±0.14	0.941±0.001
	16	0.246±0.004	21.00±0.15	0.943±0.002

244

245 **6. Conclusions**

246 We proposed a deep learning-based fast TOF-PET image reconstruction using
 247 direction information. We input the view-grouped histo-images to the 3D CNN to use
 248 the direction information. We evaluated the proposed method using Monte Carlo
 249 simulation data from a digital brain phantom. The proposed method achieved better
 250 PSNR and SSIM results, recovered finer structures than the other methods, and required
 251 a sub-second calculation time. These results indicate that the proposed method is
 252 beneficial in both the speed and image quality of TOF-PET image reconstruction.

253 **Conflicts of Interest**

254 The authors are employees of Hamamatsu Photonics K.K. The company had no control
 255 over the interpretation, writing, or publication of this work.

256 **Statement of human and animal rights**

257 This study did not include research on human subjects or animals.

258 **References**

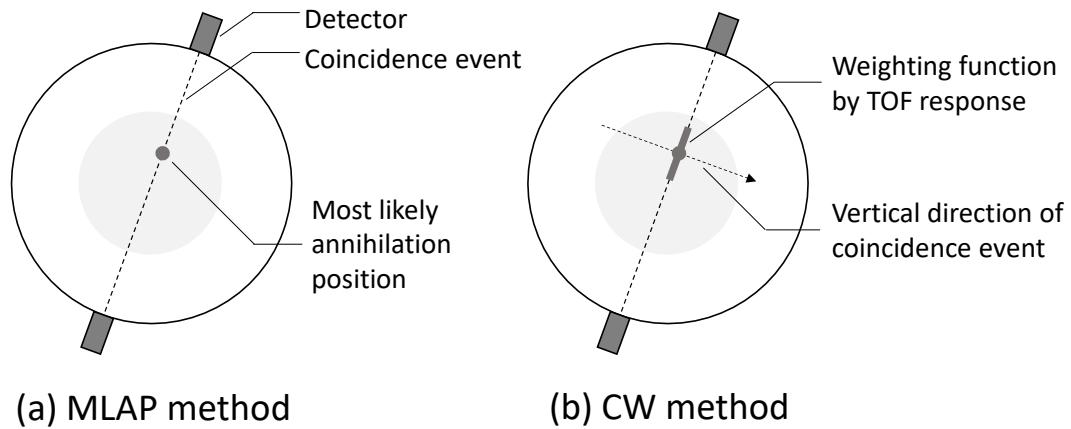
- 259 1. Phelps ME. PET: molecular imaging and its biological applications. New York:
260 Springer; 2012
- 261 2. Defrise M, Kinahan PE. Data acquisition and image reconstruction for 3D PET in
262 The Theory and Practice of 3D PET. Dordrecht: Springer; 1998
- 263 3. Wang Y, Yu B, Wang L, Zu C, Lalush D S, Lin W, et al. 3D conditional generative
264 adversarial networks for high-quality PET image estimation at low dose. Neuroimage
265 2018;174:550–62
- 266 4. Chen K T, Gong E, de Carvalho Macruz F B, Xu J, Boumis A, Khalighi M, et al. Ultra-
267 low-dose ¹⁸F-Florbetaben amyloid PET imaging using deep learning with multi-
268 contrast MRI inputs. Radiology 2019;290(3):649–56
- 269 5. Gong K, Guan J, Liu C C, Qi J. PET image denoising using a deep neural network
270 through fine tuning. IEEE Trans Radiat Plasma Med Sci 2018;3(2):153–61
- 271 6. Hashimoto F, Ohba H, Ote K, Teramoto A, Tsukada H. Dynamic PET image denoising
272 using deep convolutional neural networks without prior training datasets. IEEE
273 Access 2019;7:96594–603
- 274 7. Hashimoto F, Ohba H, Ote K, Kakimoto A, Tsukada H, Ouchi Y. 4D deep image prior:
275 Dynamic PET image denoising using an unsupervised four-dimensional branch
276 convolutional neural network. Phys Med Biol 2021;66(1):015006
- 277 8. Hashimoto F, Ito M, Ote K, Isobe T, Okada H, Ouchi Y. Deep learning-based
278 attenuation correction for brain PET with various radiotracers. Ann Nucl Med
279 2021;35(6):691–701
- 280 9. Sanaat A, Shiri I, Arabi H, Mainta I, Nkoulou R, Zaidi H. Deep learning-assisted ultra-
281 fast/low-dose whole-body PET/CT imaging. Eur J Nucl Med Mol Imaging
282 2021;48(8):2405–15

- 283 10. Yang B, Ying L, Tang J. Artificial neural network enhanced Bayesian PET image
284 reconstruction. *IEEE Trans Med Imaging* 2018;37(6):1297–309
- 285 11. Gong K, Guan J, Kim K, Zhang X, Yang J, Seo Y, et al. Iterative PET Image
286 Reconstruction Using Convolutional Neural Network Representation. *IEEE Trans*
287 *Med Imaging* 2019;38(3):675–85
- 288 12. Gong K, Catana C, Qi J, Li Q. PET image reconstruction using deep image prior.
289 *IEEE Trans Med Imaging* 2019;38(7):1655–1665
- 290 13. Whiteley W, Luk W K, Gregor J. DirectPET: full-size neural network PET
291 reconstruction from sinogram data. *J Med Imaging* 2020;7(3):032503
- 292 14. Reader AJ, Corda G, Mehranian A, da Costa-Luis C, Ellis S, Schnabel JA. Deep
293 learning for PET image reconstruction. *IEEE Trans Radiat Plasma Med Sci*
294 2021;5(1):1–25
- 295 15. Hu Z, Xue H, Zhang Q, Gao J, Zhang N, Zou S, et al. DPIR-Net: Direct PET image
296 reconstruction based on the Wasserstein generative adversarial network. *IEEE Trans*
297 *Radiat Plasma Med Sci* 2021;5(1):35–43
- 298 16. Kandarpa VSS, Bousse A, Benoit D, Visvikis D. DUG-RECON: A Framework for
299 Direct Image Reconstruction Using Convolutional Generative Networks. *IEEE Trans*
300 *Radiat Plasma Med Sci* 2021;5(1):44–53
- 301 17. Mehranian A, Reader AJ. Model-based deep learning PET image reconstruction using
302 forward-backward splitting expectation maximization. *IEEE Trans Radiat Plasma
303 Med Sci* 2021;5(1):54–64
- 304 18. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-
305 transform manifold learning. *Nature* 2018;555(7697):487–92

- 306 19. Häggström I, Schmidlein CR, Campanella G, Fuchs TJ. DeepPET: A deep encoder–
307 decoder network for directly solving the PET image reconstruction inverse problem.
308 *Med Image Anal* 2019;54:253–62
- 309 20. Hudson HM, Larkin RS. Accelerated image reconstruction using ordered subsets of
310 projection data. *IEEE Trans Med. Imaging* 1994;13(4):601–9
- 311 21. Whiteley W, Panin V, Zhou C, Cabello J, Bharkhada D, Gregor J. FastPET: near real-
312 time reconstruction of PET histo-image data using a neural network. *IEEE Trans*
313 *Radiat Plasma Med Sci* 2021;5(1):65–77
- 314 22. Matej S, Surti S, Jayanthi S, Daube-Witherspoon ME, Lewitt RM, Karp JS. Efficient
315 3-D TOF PET reconstruction using view-grouped histo-images: DIRECT—Direct
316 image reconstruction for TOF. *IEEE Trans Med. Imaging* 2009;28(5):739–51
- 317 23. Snyder DL, Thomas LJ, Ter-Pogossian MM. A matheematical model for positron-
318 emission tomography systems having time-of-flight measurements. *IEEE Trans Nucl*
319 *Sci* 1981;28(3):3575–83.
- 320 24. Tomitani T. Image reconstruction and noise evaluation in photon time-of-flight
321 assisted positron emission tomography. *IEEE Trans Nucl Sci* 1981;28(6):4581–9
- 322 25. Tanaka E Line-writing data acquisition and singal-to-noise ratio in time-of-flight
323 positron emission tomography. *IEEE Comput Soc* 1982;448:101–8.
- 324 26. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning
325 dense volumetric segmentation from sparse annotation. *Medical Image Computing*
326 and Computer Assisted Intervention (MICCAI) LNCS 2016;9901:424–32

- 327 27. Collins DL, Zijdenbos AP, Kollokian V, Sled JG, Kabani NJ, Holmes CJ, Evans AC.
328 Design and construction of a realistic digital brain phantom. IEEE Trans Med.
329 Imaging 1998;17(3):463–8
- 330 28. Saito A, Yoshikawa E, Omura T, Yamanaka T, Ote K, Isobe T, et al. Development
331 of a brain PET scanner with motion correction using motion capture technology. IEEE
332 Nuclear Science Symposium and Medical Imaging Conference 2018;M-07-146
- 333 29. Tanaka E, Kudo H. Subset-dependent relaxation in block-iterative algorithms for
334 image reconstruction in emission tomography Phys Med Biol 2003;48(10):1405–22.
- 335 30. Nakayama T, Kudo H. Derivation and implementation of ordered-subsets algorithms
336 for list-mode PET data. IEEE Nuclear Science Symposium and Medical Imaging.
337 Conference Record 2005;3540-3
- 338 31. Vandenberghe S, Daube-Witherspoon ME, Lewitt RM, Karp JS. Fast reconstruction
339 of 3D time-of-flight PET data by axial rebinning and transverse mashing. Phys Med
340 Biol 2006;51(6):1603–21
- 341

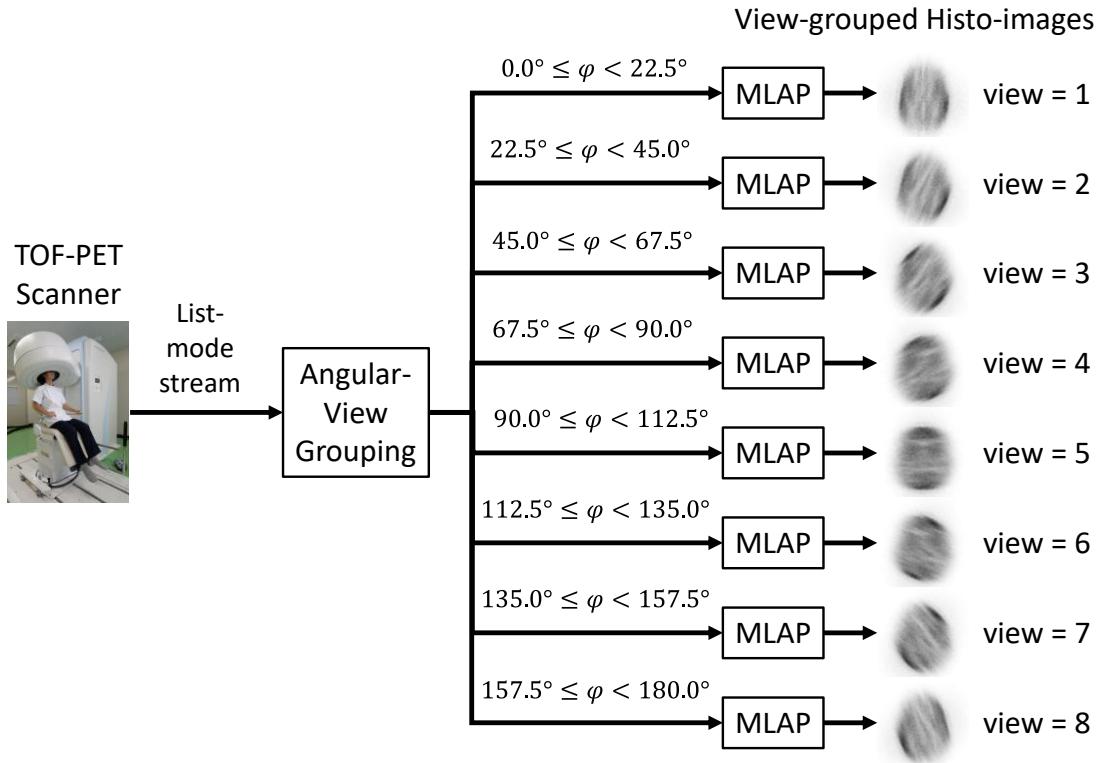
342 Figures



343

344 Figure 1: Event accumulation methods of TOF-PET. The MLAP method accumulates events into the
 345 voxel nearest to the MLAP estimated using TOF information (a). The CW method accumulates events as
 346 a line weighted by the TOF response function centered on the most likely annihilation position (b). The
 347 CW method is optimal in terms of SNR, but is time-consuming because it requires ray tracing.

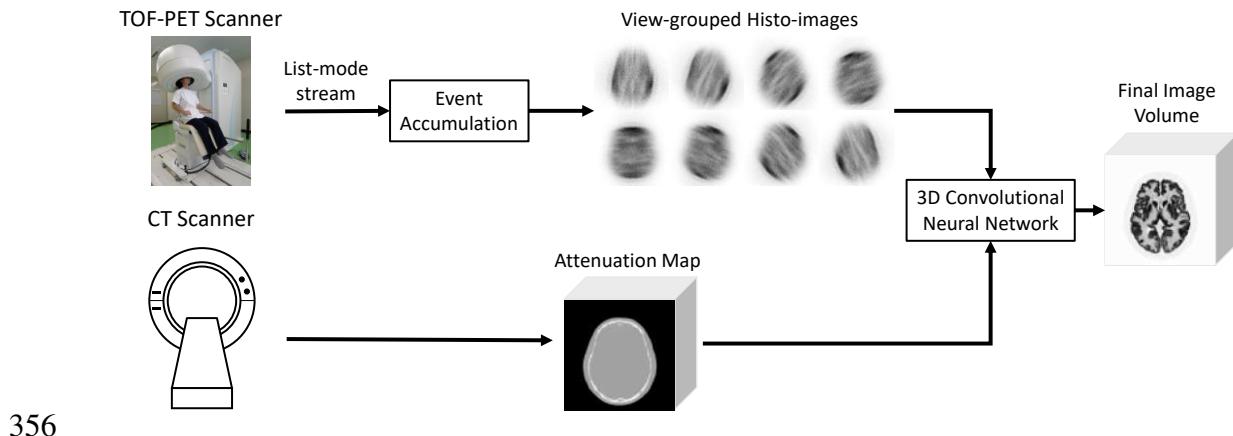
348



349

350 Figure 2: Event accumulation of TOF-PET using angular-view grouping. Events are divided into N
 351 groups depending on the angle of coincidence. The groups of events are separately accumulated in N
 352 histo-images using the MLAP method. In this example, $N = 8$. View-grouped histo-images preserve the
 353 direction information of the coincidence event without ray tracing. φ is an azimuthal angle of a
 354 coincidence event.

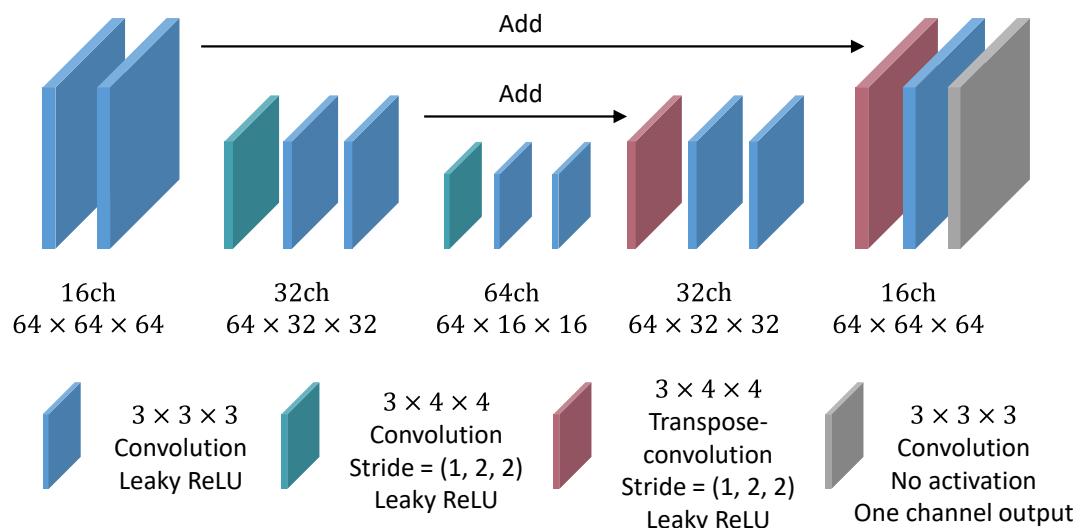
355



356

357 Figure 3: Deep learning-based fast TOF-PET image reconstruction using direction information. Three-
 358 dimensional CNN receives N view-grouped histo-images and one attenuation-map as an $(N + 1)$ channel
 359 image, and outputs the final image volume. Hence, 3D CNN can use the direction information to improve
 360 the SNR of the final image volume.

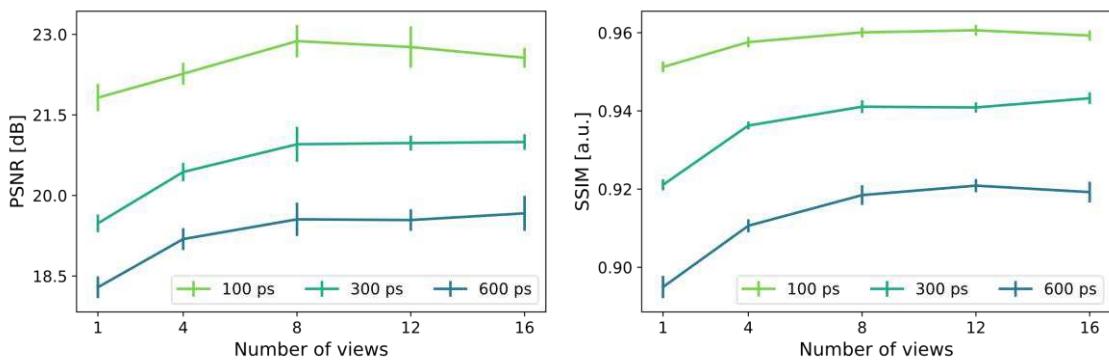
361



362

363 Figure 4: Network architecture of this study.

364

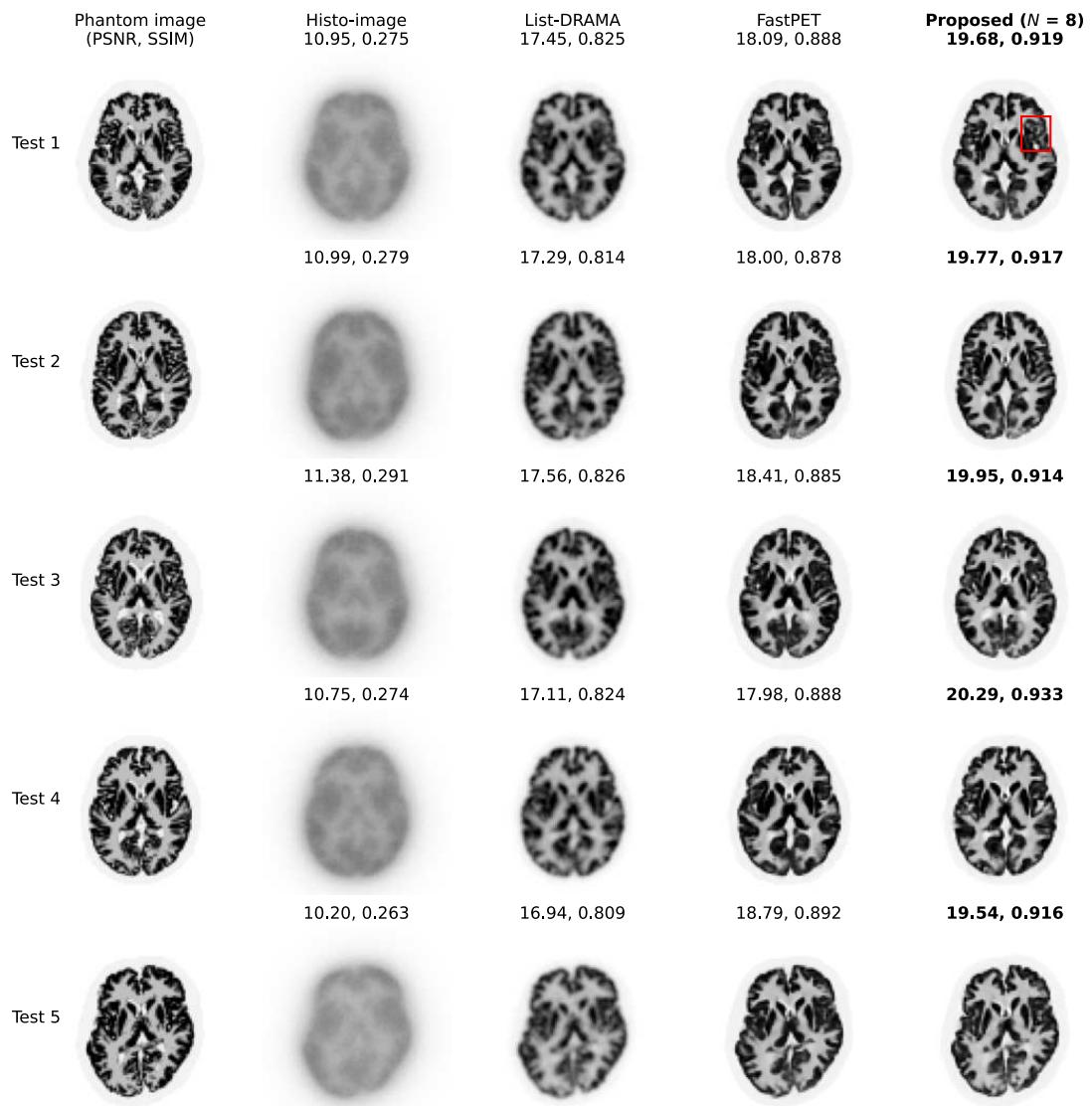


365

366 Figure 5: Effects of number of views to PSNR (left) and SSIM (right), at CTRs of 100, 300, and 600 ps.

367 PSNR and SSIM increases as the number of views increases.

368



369

370 Figure 6: Comparison of reconstructed images between the proposed method and the other methods at a
 371 CTR of 300 ps. From left to right, the phantom image, histo-image, List-DRAMA, and FastPET, which
 372 corresponds to the proposed method using 1 view, and the proposed method using 8 views. The histo-
 373 image was made by accumulating all events into single array using TOF information. The images were
 374 tagged with PSNR and SSIM. The red square indicates the region that proposed method recovered a finer
 375 structure than that of the other methods.