# Synergistic PET/MR reconstruction with VAE constraint

Valentin Gautier, Claude Comtat, Florent Sureau, Alexandre Bousse, Voichita Maxim, Bruno Sixou

# SYNERGISTIC PET/MR RECONSTRUCTION WITH VAE CONSTRAINT

*Valentin Gautier*[1], *Claude Comtat*[2], *Florent Sureau*[2], *Alexandre Bousse*[3], *Voichita Maxim*[1], *Bruno Sixou*[1]

[1] Univ. Lyon, INSA-Lyon, UCBL, UJM-Saint-Étienne, CNRS, Inserm, CREATIS UMR 5220, U1206, Lyon, France.

[2] BioMaps, Université Paris-Saclay, CEA, CNRS, Inserm, SHFJ, 91401 Orsay, France.

[3] LaTIM, INSERM U1101, Université de Bretagne Occidentale, 29238 Brest, France

## ABSTRACT

In this work, we propose a synergistic PET/MR reconstruction method based on the ADMM algorithm and a pre-trained bimodal Variational Auto Encoder (VAE) as a constraint. Because of the multiple modalities, balancing the VAE's loss becomes a challenge. To solve this, we adapt an adaptive loss balancing method and apply it to the training of the VAE. We evaluate our approach on 2D slices from 44 different patients and show that the presented approach performs particularly well on low dose/highly undersampled data.

***Index Terms*—** Tomography, reconstruction, deep learning, VAE, synergistic, PET, MRI

## I. INTRODUCTION

Positron emission tomography (PET)/magnetic resonance (MR) imaging is a recent multimodal imaging technique that is gaining attention for its ability to visualize both functional and anatomical information on well registered images [1]. With this hybrid imaging technique, the PET and MRI data are acquired on the same device. This can be used for various applications such as PET motion correction [2] or helping with diagnosis for diseases such as Alzheimer [**?**]. Recently, several studies have explored the joint (or synergistic) reconstruction of those two modalities. The idea is that by performing a joint reconstruction, we can make use of the mutual information between the two modalities to improve the reconstruction results of each of them.

Initially, variational methods were investigated and have shown how this type of reconstruction strategy can benefit both modalities [3]. This kind of method promotes structural similarity, such as common edges between the two modalities, through joint Total Variation (TV). However, this also leads to the suppression of modality specific features and the imprinting of one modality over the other. To overcome these challenges, Mehranian et al [4] proposed a generalized joint TV that allows the disentanglement of the two modalities while still taking advantage of the mutual information to enhance each reconstruction. Although the results are promising, the use of TV tends to produce smooth images and suppress small details, which is particularly detrimental, especially for MR images.

The introduction of deep learning allowed to learned more powerful and complex priors directly from the data. This proved to be very successful at capturing the mutual information with the example of unrolled approaches [5]. Among the deep architectures, the Variational Auto Encoder (VAE) [6] proved to be successful at capturing the mutual information of multimodal images [7]. A major challenge in training these models concerns the weighting of the various components of the loss in order to balance the influence of the two modalities and insure that both images will be properly generated.

In a previous work [8], we proposed a MR-guided PET reconstruction method relying on optimization of the prior in the latent space of a pre-trained bimodal VAE. In this work we further develop the idea of deep latent reconstruction (DLR) by solving several issues related to synergistic reconstruction. In particular, we adapt and implement the method proposed by Wang et al. in [9] to automatically balance the different components of our VAE's loss during training. We show that with this method, we obtain comparable results to a high-dose setting using low-dose or very undersampled data by leveraging the bimodal information.

## II. MATERIALS AND METHODS

### II-A. Problem formulation

We denote $M$ the number of PET lines of response and $N$ the number of image voxels for both the PET and MR images. The unknown vector $x_{\mathrm{pet}} \in \mathbb{R}^N$ is the radioactive tracer distribution and $P \in \mathscr{M}_{M,N}(\mathbb{R})$ the detection probability matrix. The forward model considers the data $y_{\mathrm{pet}} \in \mathbb{N}^M$ as independent Poisson random variables with expected counts $\tilde{y}_{\mathrm{pet}} = Px_{\mathrm{pet}} + r + s$ where $r$ and $s$ are the expected number of randoms and scatters. The PET data fidelity is given by the negative Poisson log-likelihood:

$$D_{\mathrm{pet}}(y_{\mathrm{pet}}, x_{\mathrm{pet}}) = \sum_{i=1}^{M} \left([\tilde{y}_{\mathrm{pet}}]_i - [y_{\mathrm{pet}}]_i \log([\tilde{y}_{\mathrm{pet}}]_i)\right) \quad (1)$$

The Magnetic Resonance (MR) imaging model is $\tilde{y}_{\mathrm{mr}} = Ex_{\mathrm{mr}}$ where $x_{\mathrm{mr}} \in \mathbb{R}^N$ is the MR image, $\tilde{y}_{\mathrm{mr}}$ and $y_{\mathrm{mr}} \in \mathbb{R}^{M_v}$ are the expected k-space data and the

measurements respectively, $E = UF \in \mathscr{M}_{M_v,N}(\mathbb{C})$ is the Fourier encoding matrix consisting of the product of the discrete Fourier transform $F$ and the operator $U$ which performs a radial subsampling of factor $R$ on the k-space data. $M_v$ is the number of k-space samples. We also assume the measurements are corrupted by Gaussian noise, and we define the MR data fidelity term as:

$$D_{\mathrm{mr}}(y_{\mathrm{mr}}, x_{\mathrm{mr}}) = \frac{1}{2}\|Ex_{\mathrm{mr}} - y_{\mathrm{mr}}\|_2^2 . \tag{2}$$

Our aim is to find the PET-MR solution $(\hat{x}_{\mathrm{pet}}, \hat{x}_{\mathrm{mr}}) \in \mathbb{R}^N \times \mathbb{R}^N$ of the following minimization problem:

$$\begin{aligned}(\hat{x}_{\mathrm{pet}}, \hat{x}_{\mathrm{mr}}, \hat{z}) = \underset{x_{\mathrm{pet}}, x_{\mathrm{mr}}, z}{\arg\min}\ & D_{\mathrm{pet}}(y_{\mathrm{pet}}, x_{\mathrm{pet}}) + D_{\mathrm{mr}}(y_{\mathrm{mr}}, x_{\mathrm{mr}}) \\ \text{s.t.}\quad & (x_{\mathrm{pet}}, x_{\mathrm{mr}}) = \mathrm{Decoder}(z)\end{aligned} \tag{3}$$

where $\mathrm{Decoder}$ is the decoder part of the VAE and the latent variable $z$ is used for the low-dimensional representation of both the PET and MR images. This latent variable is used to concentrate the mutual information and optimize over a smaller, simpler space.

Denoting $\mu$ the Lagrange multiplier and $\rho = (\rho_{\mathrm{mr}}, \rho_{\mathrm{pet}})$ the Lagrangian hyperparameter, we can apply ADMM to solve this problem, which gives the following iterations:

$$\begin{aligned} x_{\mathrm{pet}}^{n+1} &= \underset{x_{\mathrm{pet}}}{\arg\min}\, D_{\mathrm{pet}}(y_{\mathrm{pet}}, x_{\mathrm{pet}}) \\ &\quad + \frac{\rho_{\mathrm{pet}}}{2}\|x_{\mathrm{pet}} - \mathrm{Decoder}(z^n)_{\mathrm{pet}} + \mu_{\mathrm{pet}}^n\|^2 \end{aligned} \tag{4}$$

$$\begin{aligned} x_{\mathrm{mr}}^{n+1} &= \underset{x_{\mathrm{mr}}}{\arg\min}\, D_{\mathrm{mr}}(y_{\mathrm{mr}}, x_{\mathrm{mr}}) \\ &\quad + \frac{\rho_{\mathrm{mr}}}{2}\|x_{\mathrm{mr}} - \mathrm{Decoder}(z^n)_{\mathrm{mr}} + \mu_{\mathrm{mr}}^n\|^2 \end{aligned} \tag{5}$$

$$z^{n+1} = \underset{z}{\arg\min}\, \|\mathrm{Decoder}(z) - (x^{n+1} + \mu^n)\|^2 \tag{6}$$

$$\mu^{n+1} = \mu^n + x^{n+1} - \mathrm{Decoder}(z^{n+1}) \tag{7}$$

where we denoted $x^n = (x_{\mathrm{pet}}^n, x_{\mathrm{mr}}^n)$, $\mu^n = (\mu_{\mathrm{pet}}^n, \mu_{\mathrm{mr}}^n)$ and $\mathrm{Decoder}(z^n) = (\mathrm{Decoder}(z^n)_{\mathrm{pet}}, \mathrm{Decoder}(z^n)_{\mathrm{mr}})$ . The minimization problem of Eq. (4) is solved with optimization transfer and a convex surrogate function similar to the one used for the classical MLEM algorithm [10]. We used the following update formula for $x_{\mathrm{pet}}^{n+1}$ at each pixel $j$:

$$\begin{aligned}\left[x_{\mathrm{pet}}^{n+1}\right]_j = \frac{1}{2}\Bigg( & [\mathrm{Decoder}(z^n)_{\mathrm{pet}}]_j - [\mu_{\mathrm{pet}}^n]_j - \frac{p_j}{\rho_{\mathrm{pet}}} \\ & + \sqrt{\left([\mathrm{Decoder}(z^n)_{\mathrm{pet}}]_j - [\mu_{\mathrm{pet}}^n]_j - \frac{p_j}{\rho_{\mathrm{pet}}}\right)^2 + \frac{4p_j[x_{\mathrm{pet,em}}^{n+1}]_j}{\rho_{\mathrm{pet}}}}\Bigg)\end{aligned} \tag{8}$$

where $p_j = \sum_i P_{i,j}$ and $x_{\mathrm{pet,em}}^{n+1}$ is obtained by doing one MLEM step:

$$[x_{\mathrm{pet,em}}^{n+1}]_j = \frac{[x_{\mathrm{pet}}^n]_j}{p_j}\sum_i P_{ij}\frac{[y_{\mathrm{pet}}]_i}{[Px_{\mathrm{pet}}^n]_i + r_i + s_i} \tag{9}$$

Equation 5 is a penalized least square problem. This problem can be solved using first order optimality condition. We get the following equation :

$$(\rho_{\mathrm{mr}}I + E^HE)x_{\mathrm{mr}}^{n+1} = E^Hy_{\mathrm{mr}} + \rho_{\mathrm{mr}}(\mathrm{Decoder}(z^n)_{\mathrm{mr}} - \mu_{\mathrm{mr}}^n) \tag{10}$$

The solution of this equation can be obtained with a few iterations of the conjugate gradient algorithm, which is used extensively for MR image reconstruction [11].

The update of the latent variable in Eq. (6) is performed with a gradient descent. We use the Adam algorithm [12] that includes tuning of the step size. The scaling of the $\mathrm{Decoder}$ output is also an important issue. It is common practice to train the decoder on normalized data but the image that we are trying to reconstruct is not necessarily normalized. Thus, we have to rescale independently the decoded images during the iterations to match the current reconstruction of the corresponding modalities. The additional ADMM penalty parameter $\rho$ has a strong influence on the convergence rate and is often chosen empirically based on some validation data. In this work, we implemented the adaptive update scheme proposed recently in [13] independently for $\rho_{\mathrm{pet}}$ and $\rho_{\mathrm{mr}}$. Its principle is to balance the relative primal and dual residuals while taking into account the scaling properties of the ADMM problem. We have also implemented the stopping criterion proposed in the same paper with $\epsilon = 0.02$.

### II-B. Dataset

The data used for training the VAE and testing our method consists in 840 co-registered 2D brain [$^{18}$F]FDG PET images (20 minutes acquisition) and T1-weighted MR images extracted from 44 acquisitions on a clinical hybrid PET/MR scanner (Signa PET/MR, GE Healthcare) of patients with dementia or epilepsy. These images are of shape $256 \times 256$ and are considered our references. Since our goal is to achieve the quality of a high dose reconstruction in a low dose setting, we consider those high dose images as the references and we simulate low dose data from them. The low dose PET data $y_{\mathrm{pet}}$ are generated according to the following:

$$y_{pet} = \mathrm{Poisson}\left(\frac{\alpha}{\|x_{\mathrm{pet}}\|_1}(Px_{\mathrm{pet}} + r + s)\right) \tag{11}$$

where $\alpha$ is a factor used to control the noise level. The signal to noise ratio is lower for lower values of $\alpha$. We add $r$ (random) and $s$ (scatter) events so that they correspond to 1% of the total number of observed events . The resulting sinograms are of shape 256 (number of bins) $\times$ 60 (number of angles).

For the MRI, the data are generated according to the following :

$$y_{\mathrm{mr}} = Ex_{\mathrm{mr}} + b \tag{12}$$

where $b$ is some random Gaussian noise. We can control the level of noise we add through $b$ and we simulate a shorter

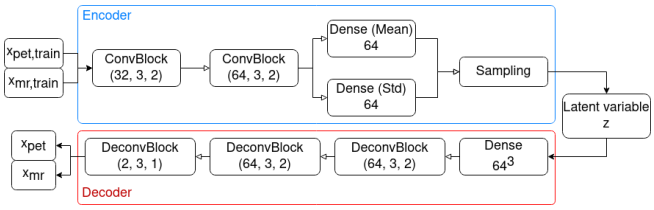


**Fig. 1**. Architecture of the VAE.

acquisition time by increasing the subsampling factor in the operator $E$.

### II-C. VAE structure and training

In this work, we use VAEs to generate PET and MR images. A VAE is a latent space generative model based on the variational Bayesian inference originally proposed by Kingma and Welling [6]. The structure of our VAE is displayed in figure 1. Each convolution and transposed convolution is described with the format (number of channels, filter size, stride) and the dense layers are described by their number of neurons. The sampling layer is used to perform the reparametrization trick [6] and takes a mean vector and a standard deviation vector to sample a latent variable $z$. We use a multichannel input CNN, which treats PET/MR images as two separate input channels. The training loss of our VAEs can be written as:

$$\mathscr{L}(\theta) = \mathscr{L}_{KL}(\theta) + \lambda_{\text{pet}}\mathscr{L}_{rec}^{\text{pet}}(\theta) + \lambda_{\text{mr}}\mathscr{L}_{rec}^{\text{mr}}(\theta) \qquad (13)$$

where $\mathscr{L}_{KL}(\theta)$ is the KL loss, $\mathscr{L}_{rec}^{\text{pet}}(\theta)$ and $\mathscr{L}_{rec}^{\text{mr}}(\theta)$ are the reconstruction losses for the PET and MR modality respectively and $\lambda_{\text{pet}}$ and $\lambda_{\text{mr}}$ are the weighting coefficients associated with each modality.

The challenge with training such a VAE is that the different components of the loss can have completely different scales. Typically, in our experiments, there are four orders of magnitude between the reconstruction loss and the KL loss. To tackle this issue, we use the method proposed in [9]. By using the gradients of the different losses we can update $\lambda_{\text{pet}}$ and $\lambda_{\text{mr}}$ before each gradient descent step in order to get each loss at the same scale. This training strategy was crucial to obtain satisfactory results.

The training was implemented using the open-source library Keras 2.2.5 with Tensorflow backbone and performed on an NVIDIA RTX A2000 mobile. The network is trained for 500 epochs using the Adam optimizer with a learning rate of $10^{-3}$ and a batch size of 32. The images from our dataset were used as the high-quality references and the VAE was trained on them. The dataset was split into 3 parts: one for training, one for validation (20% of the data) and one for testing (10% of the data).

### II-D. Experiments

We test our method by reconstructing 2D slices with various simulated acquisition times. We then compare the reconstruction results of the PET image with MLEM [14] and the ones for the MR image with the simple inverse Fourier transform.

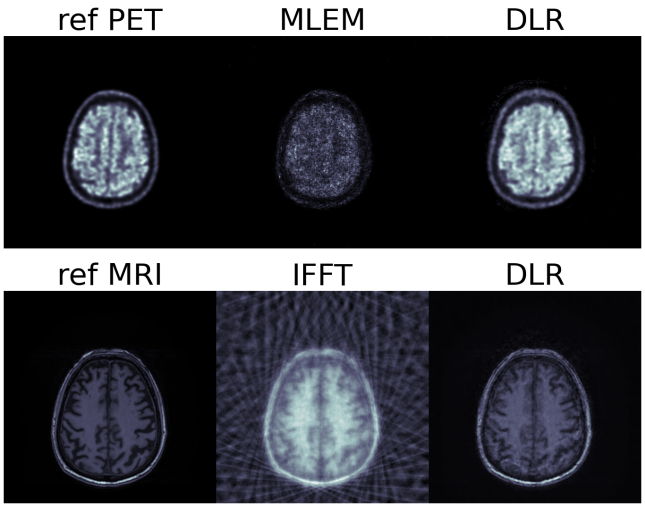


**Fig. 2**. Ground truth images and reconstructed images obtained for $\alpha = 10^5$ and a subsampling factor of 40.

We initialize the algorithm with the 10th iteration of MLEM for the PET and the inverse Fourier transform for the MRI. The initial latent variable is given by using the encoder part of the VAE on the initial PET and MR images. The forward and backward projections are handled by the ASTRA toolbox [15] with a parallel geometry. The test reconstructions were performed on a test set of 46 slices never seen by the VAE.

## III. RESULTS

Figure 2 shows a slice from the test set reconstructed with our approach together with the ground truth images and the reconstructions given by the MLEM algorithm for PET and IFFT for the MRI. Visually, we can see that we manage to retrieve a lot of lost information but we still lose some high frequency details. This can be partly attributed to the very nature of the VAE that compresses the data and loses the high frequencies in the process.

Figure 3 shows the evolution of the distance to the constraint $\|\text{Decoder}(z^n)_{\text{pet}} - x_{\text{pet}}^n\|$ as a function of the ADMM iterations for the slice shown in figure 2 for the proposed method during one reconstruction. With the adaptive update of the Lagrangian parameter $\rho$, the behavior of the error metrics is highly nonlinear. The regularizing effect of the constraint is obtained after a few iterations when $\rho$ increases significantly. The large decrease of the distance to the VAE's output is concomitant with the MSE decrease. It should be noted that simultaneously the data fidelity term increases which means that once a good latent variable has been found, the improvement is due to the constraint. The projection on the image manifold learned with the autoencoder is

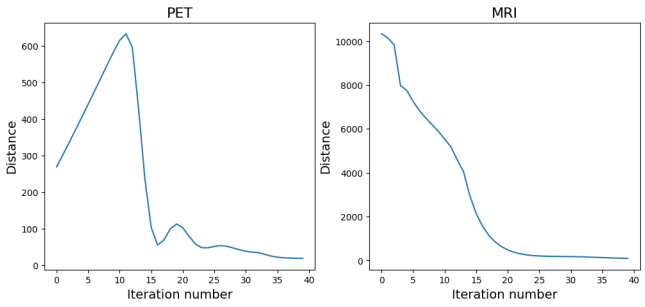


**Fig. 3**. Evolution of the distance to the constraint for $\alpha = 10^5$ and $R = 20$.

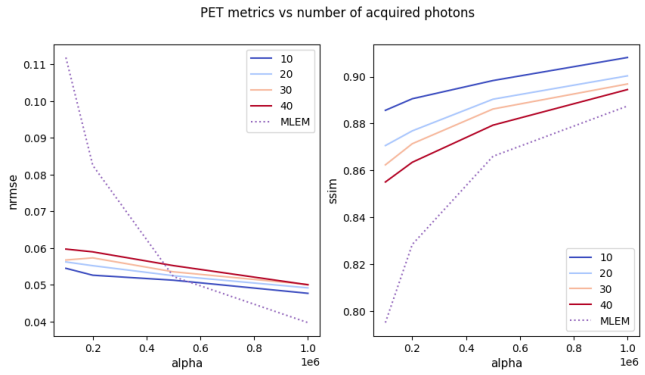


**Fig. 4**. Evolution the NRMSE and SSIM with the photon counts of the PET reconstruction, for different undersampling rates $R$.

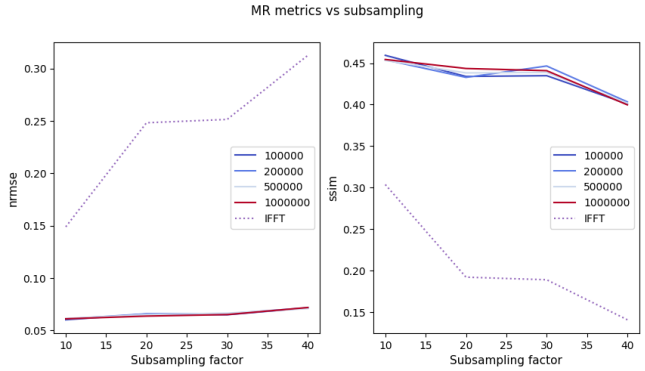


**Fig. 5**. Evolution of the NRMSE and SSIM with the undersampling rate for the MR reconstruction, for different numbers of photons.

thus efficient to reduce the noise and the artifacts on the reconstructed image.

The quality of the reconstructions was evaluated quantitatively using the normalised root mean squared error (NRMSE) and (Structural SIMilarity) SSIM for several noise levels and undersampling rates R. These metrics are plotted taking the undersampling rate or noise level of the other modality as parameter. In figure 4, for the PET modality, we plot the evolution of the NRMSE and of the SSIM with the numbers of photons for the proposed method, for different undersampling rates R and we compare the results with the ones obtained with MLEM. Similarly, Figure 5 displays the evolution of the error metrics with the subsampling factor R, for various photons counts, and we plot also the errors with the naive IFFT for comparison. We can see that the proposed approach performs well especially in the case of a lot of missing data for both modalities. We can also see on figure 4 how reducing the subsampling factor (which corresponds to a more accurate acquisition) leads to a slightly better PET reconstruction. A similar effect is less obvious but can also be seen for the MR image on figure 5. This shows the effect of the multimodality on the reconstruction while hinting at a greater impact of the MR modality on the PET one than the opposite.

## IV. DISCUSSION

The main advantage of the proposed method is that the regularizing effect is learned and not based on penality terms like TV or joint TV regularization. By using a multimodal VAE, we get a latent variable that sums up the mutual information between the two modalities. With enough data it should be possible to get a latent space from which we can generate any PET/MR images. Moreover, the approach is based on a limited number of hyper-parameters thanks to the automatic update method adopted for the ADMM algorithm. One of the current drawbacks of the method is that the quality of the reconstructed images fully depends on the images generated by the VAE. Firstly, the reconstructed images are thus a bit blurry because of the very nature of the VAE. Second, there are mismatches with the data that are created by the VAE. This is particularly true with tumours for example since the VAE has not been trained on these anomalies and thus, cannot generate them. To try and address these challenges, we are most interested in the recent diffusion models which could enable a sharper reconstruction and a better generation conditioned on the measured data [16], [17].

## V. CONCLUSION

Our aim in this paper is to improve the fusion of the complementary informations in PET/MR images. We have investigated a network-constrained image reconstruction method where a pre-trained multi-channel input VAE trained with high-quality images is used to represent feasible PET and MR images. We show that using the two modalities, we are able to find a latent variable that represents denoised images despite a lot of missing data. In future works, we will try to get better results by using new architectures and we will better quantify the impact of the bimodality on the quality of the reconstruction.

## VI. ACKNOWLEDGMENT

We acknowledge financial support form the French National Research Agency (ANR) under grant ANR-20-CE45-0020 (ANR MULTIRECON).

## VII. COMPLIANCE WITH ETHICAL STANDARDS

This study was performed in line with the principles of Helsinki. Approval was granted by the executive board of the CEA/SHFJ department (2022/11/10).

## VIII. REFERENCES

[1] Ciprian Catana, Alexander Drzezga, Wolf-Dieter Heiss, and Bruce R. Rosen, "PET/MRI for Neurologic Applications," *Journal of Nuclear Medicine*, vol. 53, no. 12, pp. 1916–1925, 2012, Publisher: Society of Nuclear Medicine _eprint: https://jnm.snmjournals.org/content/53/12/1916.full.pdf.

[2] Jinsong Ouyang, Quanzheng Li, and Georges El Fakhri, "Magnetic Resonance-Based Motion Correction for Positron Emission Tomography Imaging," *Seminars in Nuclear Medicine*, vol. 43, no. 1, pp. 60–67, 2013.

[3] Matthias J Ehrhardt, Kris Thielemans, Luis Pizarro, David Atkinson, Sébastien Ourselin, Brian F Hutton, and Simon R Arridge, "Joint reconstruction of PET-MRI by exploiting structural similarity," *Inverse Problems*, vol. 31, no. 1, pp. 015001, Jan. 2015.

[4] Abolfazl Mehranian, Martin A. Belzunce, Claudia Prieto, Alexander Hammers, and Andrew J. Reader, "Synergistic PET and SENSE MR Image Reconstruction Using Joint Sparsity Regularization," *IEEE Transactions on Medical Imaging*, vol. 37, no. 1, pp. 20–34, Jan. 2018.

[5] Guillaume Corda-D'Incan, Julia A. Schnabel, and Andrew J. Reader, "Syn-Net for Synergistic Deep-Learned PET-MR Reconstruction," pp. 1–5, Oct. 2020.

[6] Diederik P Kingma and Max Welling, "Auto-encoding variational bayes," 2013.

[7] Masahiro Suzuki and Yutaka Matsuo, "A survey of multimodal deep generative models," *Advanced Robotics*, vol. 36, no. 5-6, pp. 261–278, Mar. 2022, arXiv:2207.02127 [cs, stat].

[8] Valentin Gautier, Claude Comtat, Florent Sureau, Louise Bousse, Alexandre Friot-Giroux, Voichita Maxim, and Bruno Sixou, "Vae constrained mr guided pet reconstruction," in *Fully3D 2023*, 2023.

[9] Sifan Wang, Yujun Teng, and Paris Perdikaris, "Understanding and mitigating gradient pathologies in physics-informed neural networks," Jan. 2020, arXiv:2001.04536 [cs, math, stat].

[10] Zhaoheng Xie, Tiantian Li, Xuezhu Zhang, Wenyuan Qi, Evren Asma, and Jinyi Qi, "Anatomically aided pet image reconstruction using deep neural networks," *Medical Physics*, vol. 48, no. 9, pp. 5244–5258, 2021.

[11] Klaas P. Pruessmann, Markus Weiger, Peter Börnert, and Peter Boesiger, "Advances in sensitivity encoding with arbitrary $k$ -space trajectories: SENSE With Arbitrary $k$ -Space Trajectories," *Magnetic Resonance in Medicine*, vol. 46, no. 4, pp. 638–651, Oct. 2001.

[12] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," 2017.

[13] Brendt Wohlberg, "Admm penalty parameter selection by residual balancing," 2017.

[14] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Transactions on Medical Imaging*, vol. 1, no. 2, pp. 113–122, 1982.

[15] Wim van Aarle, Willem Jan Palenstijn, Jeroen Cant, Eline Janssens, Folkert Bleichrodt, Andrei Dabravolski, Jan De Beenhouwer, K. Joost Batenburg, and Jan Sijbers, "Fast and flexible x-ray tomography using the astra toolbox," *Opt. Express*, vol. 24, no. 22, pp. 25129–25147, Oct 2016.

[16] Kushagra Pandey, Avideep Mukherjee, Piyush Rai, and Abhishek Kumar, "Diffusevae: Efficient, controllable and high-fidelity generation from low-dimensional latents," 2022.

[17] A. Lopez-Montes, T. McSkimming, W. Zbijewski, A. Skeats, C. Delnooz, B. Gonzales, M. Marie, J.H. Siewerdsen, and A. Sisniega, "Diffusion posterior sampling-based reconstruction for stationary ct imaging of intracranial hemorrhage," in *Fully3D 2023*, 2023.