VAE constrained MR guided PET reconstruction

Valentin Gautier, Claude Comtat, Florent Sureau, Alexandre Bousse, Louise Friot-Giroux, Voichita Maxim, Bruno Sixou

VAE constrained MR guided PET reconstruction

Valentin Gautier[1], Claude Comtat[2], Florent Sureau[2], Alexandre Bousse[3], Louise Friot–Giroux[1], Voichita Maxim[1], and Bruno Sixou[1]

[1]Université de Lyon, INSA-Lyon, UCBL 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69621, 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 investigate a deep learning PET-MR joint reconstruction method based on the ADMM algorithm. The a priori information to regularize the inverse problem is obtained with a VAE trained with high-quality images. Adaptive choice of the Lagrangian parameter ensures good convergence properties of the method. The proposed approach is tested on simple cases. It outperforms the classical MLEM for high noise levels.

1 Introduction

Dual imaging positron emission tomography (PET)-magnetic resonance imaging (MRI) scanners have been investigated recently as an imaging modality offering both functional and anatomical information. With this hybrid imaging technique, the PET and MRI data are simultaneously acquired. Several works have investigated the synergistic reconstruction of PET and MRI data in order to improve the reconstruction results obtained with conventional independent reconstruction approaches. The idea is to exploit the common features and similarities between PET and MR images.

Variational methods have been investigated by Ehrhardt, Thielemans, Pizarro, et al. [1]. Structural similarity, joint sparsity and alignment of the gradients of the two images is promoted through a Total Variation (TV) prior. Improvements have been proposed to overcome cross-talk artifacts [2]. Recently, Mehranian, Belzunce, Prieto, et al. [3] have proposed a non-convex joint sparsity prior generalizing the joint TV to promote common boundaries while preserving modality-unique features. Their reconstruction framework is based on the augmented Lagrangian method with a scaling to take into account the dependence of the prior on the magnitude of the PET and MR images gradients. The performance of the algorithm is highly dependent on the PET-MR initialization and on the selection rules of hyper-parameters.

Deep learning methods have opened a new area of research for medical image reconstruction and they have allowed substantial improvements over state-of-the-art conventional methods in terms both accuracy and execution time. More specifically, for synergistic PET-MR reconstruction, deep learning approaches have been studied to overcome the limitations of these variational methods [4]. These methods are based on unrolling techniques that leverage the classical iterative algorithms used for image reconstruction. The proposed synergistic PET-MR reconstruction algorithm interconnects two networks to guide one modality with the other. Generative modeling has also been used for PET image denoising with MR images [5].

In this work we propose the deep latent reconstruction method (DLR) for synergistic PET-MR images which leverages the ADMM iterative method of Mehranian, Belzunce, Prieto, et al. The Total Variation regularization is replaced by a learned constraint obtained with a Variational Auto-Encoder (VAE) [6] trained with high-quality PET-MR images. The latent variable is used to represent the common information shared by the two imaging modalities. The proposed algorithm could be used for synergistic reconstruction although we focus on MR guided PET reconstruction in this paper.

The work is structured as follows. In the first section, we summarize the ADMM algorithm and we present the VAE we used as well as our dataset. We then present and discuss preliminary results showing that the proposed method outperforms the classical MLEM algorithm, showing promising for guided reconstruction as well as mutlimodal reconstruction.

2 Materials and Methods

2.1 Forward imaging models and ADMM approach for synergistic reconstruction

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

$$D_{\text{pet}}(y_{\text{pet}}, x_{\text{pet}}) = \sum_{i=1}^{M} \left([\tilde{y}_{\text{pet}}]_i - [y_{\text{pet}}]_i \log([\tilde{y}_{\text{pet}}]_i) + \log([y_{\text{pet}}]_i!) \right) . \tag{1}$$

The Magnetic Resonance (MR) imaging model is $\tilde{y}_{\text{mr}} = Ex_{\text{mr}}$ where $x_{\text{mr}} \in \mathbb{R}^N$ is the MR image, $\tilde{y}_{\text{mr}}$ and $y_{\text{mr}} \in \mathbb{R}^{M_v}$ are the expected k-space data and the measurements respectively, $E \in \mathbb{R}^{M_v\times N}$ is the Fourier encoding matrix consisting of the product of the discrete Fourier transform F and subsampling k-space operator, M_v and N are respectively the number of k-space samples and MR image voxels. In the following, we will use fully sampled spectra. We also assume the measurements are corrupted by Gaussian noise, and we define the

MR data fidelity term :

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

Our new method of synergistic reconstruction is inspired from the method of Mehranian, Belzunce, Prieto, et al. [3] and from the approach investigated by Xie, Li, Zhang, et al. [5] for guided mono-modal reconstruction. It is well-known that variational autoencoders (VAEs) allow to reliably represent complex data in a lower dimensional space. Our hypothesis in this work is that by training a VAE to represent both modalities with a single latent variable, it should learn more about the mutual information between them. Thus, we assume that the images are the output of the decoder part of a VAE with a single input z,

$$(x_{\text{pet}}, x_{\text{mr}}) = \text{Decoder}(z) \tag{3}$$

where Decoder is the decoder part of the VAE and the latent variable z is used for the low-dimensional representation of the PET and MR images. Our aim is to find the PET-MR solution $(\hat{x}_{\text{pet}}, \hat{x}_{\text{rm}}) \in \mathbb{R}^N \times \mathbb{R}^N$ of the following minimization problem:

$$\begin{aligned}(\hat{x}_{\text{pet}}, \hat{x}_{\text{mr}}, \hat{z}) = &\underset{x_{\text{pet}}, x_{\text{mr}}, z}{\arg\min}\ D_{\text{pet}}(y_{\text{pet}}, x_{\text{pet}}) + D_{mr}(y_{\text{mr}}, x_{\text{mr}}) \\ &\text{s.t.} \quad (x_{\text{pet}}, x_{\text{mr}}) = \text{Decoder}(z)\end{aligned} \tag{4}$$

We apply the augmented Lagrangian method to the constrained optimization problem. We used the ADMM [7] algorithm to solve (4). Denoting μ the Lagrange multiplier and $\rho = (\rho_{\text{mr}}, \rho_{\text{pet}})$ the Lagrangian hyperparameter, the ADMM iterations can be rewritten:

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

$$\begin{aligned} x_{\text{mr}}^{n+1} &= \underset{x_{\text{mr}}}{\arg\min} D_{\text{mr}}(y_{\text{mr}}, x_{\text{mr}}) \\ &\quad + \frac{\rho_{\text{mr}}}{2}\|x_{\text{mr}} - \text{Decoder}(z^n)_{\text{mr}} + \mu_{\text{mr}}^n\|^2 \end{aligned} \tag{6}$$

$$z^{n+1} = \underset{z}{\arg\min}\, \|\text{Decoder}(z) - (x^{n+1} + \mu^n)\|^2 \tag{7}$$

$$\mu^{n+1} = \mu^n + x^{n+1} - \text{Decoder}(z^{n+1}) \tag{8}$$

where we denoted $x^n = (x_{\text{pet}}^n, x_{\text{mr}}^n)$, $\mu^n = (\mu_{\text{pet}}^n, \mu_{\text{mr}}^n)$ and $\text{Decoder}(z^n) = (\text{Decoder}(z^n)_{\text{pet}}, \text{Decoder}(z^n)_{\text{mr}})$. The minimization problem of Eq. (5) is solved with optimization transfer and convex surrogate function similar to the one of the classical MLEM algorithm [5]. We used the following update formula for x_{pet}^{n+1} at each pixel j:

$$\begin{aligned}\left[x_{\text{pet}}^{n+1}\right]_j = \frac{1}{2}\Bigg(&[\text{Decoder}(z^n)_{\text{pet}}]_j - [\mu_{\text{pet}}^n]_j - \frac{p_j}{\rho_{\text{pet}}} \\ &+ \sqrt{\left(\text{Decoder}(z^n)_{\text{pet}\,j} - [\mu_{\text{pet}}^n]_j - \frac{p_j}{\rho_{\text{pet}}}\right)^2 + \frac{4p_j[x_{\text{pet,em}}^{n+1}]_j}{\rho_{\text{pet}}}}\Bigg)\end{aligned} \tag{9}$$

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

$$[x_{\text{pet,em}}^{n+1}]_j = \frac{[x_{\text{pet}}^n]_j}{p_j}\sum_i P_{ij}\frac{[y_{\text{pet}}]_i}{[Px_{\text{pet}}^n]_i + [r]_i + [s]_i} \tag{10}$$

The update of the latent variable in Eq. (7) is obtained with a simple gradient descent with a gradient step S which is easily performed using Tensorflow's Gradient API [8]. This step size S has to be tuned depending on the input data.

The scaling of the 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. We thus chose to rescale the output of the decoder by using a scaling factor equal to the mean of the current reconstructed image.

The additional ADMM penalty parameter ρ 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 [9]. 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 $\varepsilon = 0.02$.

2.2 Dataset

The data used for training the VAE and testing our method consists in 840 co-registered 2D brain [^{18}F]FDG PE 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 256×256 and are considered our references. The data used for the reconstructions are generated from the images x_{pet} according to the following:

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

where α is a factor used to control the noise level . The signal to noise ratio is lower for lower values of α. 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) and are then used as the input data for our reconstruction method.

2.3 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.

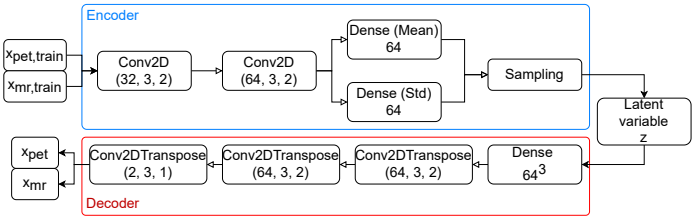

Figure 1: Architecture of the VAE.

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 VAE we considered in our work is a β-VAE [10] which allows us to balance the training of the model between latent space regularization and data fidelity. On top of that, we use a L_2 loss for the data fidelity term with a weighting parameter to balance the two modalities' contributions to the loss. 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).

2.4 Experiments

For preliminary tests, we have fixed the MR image to the reference image and only reconstructed the PET image. We then compare the reconstruction results to the ones obtained with the classical MLEM algorithm [11].
We initialize the algorithm with the 10th iteration of MLEM and initialize the latent variable by using the encoder part of the VAE on the initial PET image and the reference MR image. The forward and backward projections are handled by the ASTRA toolbox [12] with a parallel geometry. The test reconstructions were performed on 10 slices from the test set.

3 Results

Figure 2 shows a slice from the test set reconstructed with our approach together with the ground truth image and the reconstruction given by the MLEM algorithm . Qualitatively, the DLR approach outperforms MLEM on these very deteriorated data.
Figure 3 shows the evolution of the mean squared error (MSE) and of the constraint $\|\text{Decoder}(z^n)_{\text{pet}} - x^n_{\text{pet}}\|$ as a function of the ADMM iterations for the slice shown in figure 1 for the proposed method during one reconstruction. We also show the best NRMSE obtained by MLEM run for 30 iterations for comparison: with the adaptive update of the

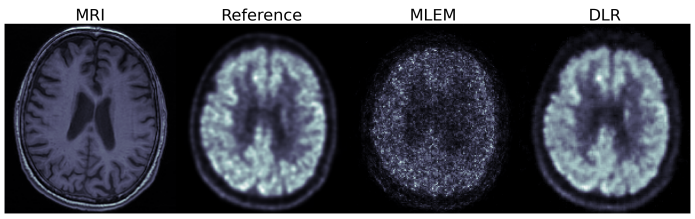

Figure 2: Ground truth image and reconstructed images obtained with MLEM and the deep latent reconstruction approach for $\alpha = 10^5$.

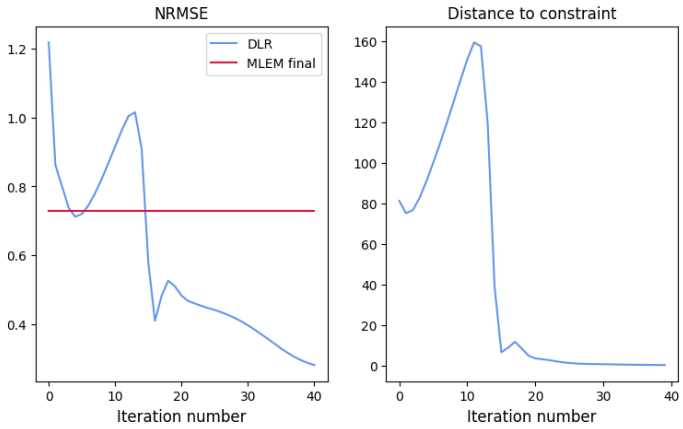

Figure 3: Evolution of the NRMSE, SSIM, data fidelity terms and constraint as a function of iterations for $\alpha = 10^5$. The MLEM NRMSE is displayed for comparison.

Lagrangian parameter ρ, the behavior of the error metrics is highly nonlinear. The regularizing effect of the constraint is obtained after a few iterations when ρ increases significantly. The large decrease of the constraint corresponding to the VAE is concomitant with the MSE decrease. It should be noted that simultaneously the data fidelity term increases which means that once a good latent vairable has been found, the improvement is due to the constraint. The projection on the image manifold learned with the autoencoder is thus efficient to reduce the noise and the artifacts on the reconstructed image.
The quality of the reconstructions was evaluated quantitatively using the the normalised root mean squared error (NRMSE) and SSIM for several noise levels. The results obtained for several Poisson noise levels are displayed in table 1 together with the value of the gradient step S chosen empirically. The deep latent reconstruction method clearly outperforms the MLEM approach for lower signal to noise ratio.

4 Discussion

In this work, we have presented preliminary results obtained with a new deep latent reconstruction approach. The VAE constrained reconstruction framework achieves better performance compared with MLEM for the various noise levels investigated. The main advantage of the method

α	S	MLEM		DLR	
		NRMSE	SSIM	NRMSE	SSIM
10^5	10^{-2}	1.47	0.79	0.88	0.83
2×10^5	5×10^{-3}	1.1	0.82	0.7	0.86
5×10^5	10^{-3}	0.68	0.86	0.65	0.87

Table 1: Comparison of the mean for NRMSE and SSIM for the MLEM method and the proposed DLR method.

is that the regularizing effect is learned and not based on penality terms like TV or joint TV regularization. By using a 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 image. Moreover, the approach is based on a limited number of hyper-parameters thanks to the automatic update method adopted for the ADMM algorithm. It should be noted that the known convergence properties of the classical ADMM algorithm are not guaranteed since the optimization method uses both a convex surrogate and a non linear constraint.

For the presented MR-aided reconstruction, several aspects will be further investigated. First, we would like to get more data and use more realistic simulations to evaluate the method in a clinically accurate context.

One current limitation of the method is that the latent variable z update from Eq.7 is based on a simple gradient descent with a rough estimate of the gradient step. This sometimes leads to updates outside of the known latent space and may cause the algorithm's divergence. Additional constraints could be studied to improve the search of the optimal latent variable. The simple VAE used could also be improved to generate less blurred images. The VAE that was used here is known to produce blurry images, which is good enough to handle PET images but not enough for MR images. With some improvement, we could also handle MR reconstruction and improve on the PET one. Variants like VAE GAN, InfoVAE [13] or even diffusion models [14] exist in the literature and could lead to improving the quality of the generated images as well as a better use of the mutual information. In the end, the proposed framework is highly flexible and each of its components can be improved individually.

5 Conclusion

Our aim in this paper is to improve the fusion of the complementary information in PET/MR images. We have investigated a network-constrainted 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 an MR image for guidance, we can find a suitable latent variable to represent our denoised PET data. In future work, we will consider the joint reconstruction of PET and MR images and compare our results with other deep learning based reconstructions approaches such as unrolling.

6 Acknowledgment

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

References

[1] M. J. Ehrhardt, K. Thielemans, L. Pizarro, et al. "Joint reconstruction of PET-MRI by exploiting structural similarity". en. *Inverse Problems* 31.1 (Jan. 2015), p. 015001. DOI: 10.1088/0266-5611/31/1/015001.

[2] F. Knoll, M. Holler, T. Koesters, et al. "Joint MR-PET Reconstruction Using a Multi-Channel Image Regularizer". *IEEE Transactions on Medical Imaging* 36.1 (2017), pp. 1–16. DOI: 10.1109/TMI.2016.2564989.

[3] A. Mehranian, M. A. Belzunce, C. Prieto, et al. "Synergistic PET and SENSE MR Image Reconstruction Using Joint Sparsity Regularization". en. *IEEE Transactions on Medical Imaging* 37.1 (Jan. 2018), pp. 20–34. DOI: 10.1109/TMI.2017.2691044.

[4] G. Corda-D'Incan, J. A. Schnabel, and A. J. Reader. "Syn-Net for Synergistic Deep-Learned PET-MR Reconstruction". en (Oct. 2020), pp. 1–5. DOI: 10.1109/NSS/MIC42677.2020.9508086.

[5] Z. Xie, T. Li, X. Zhang, et al. "Anatomically aided PET image reconstruction using deep neural networks". *Medical Physics* 48.9 (2021), pp. 5244–5258. DOI: https://doi.org/10.1002/mp.15051.

[6] D. P. Kingma and M. Welling. "Auto-Encoding Variational Bayes" (2013). DOI: 10.48550/ARXIV.1312.6114.

[7] S. Boyd, N. Parikh, E. Chu, et al. "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers". *Foundations & Trends in Machine Learning* 3.1 (2010), pp. 1–122.

[8] S. Ruder. "An overview of gradient descent optimization algorithms" (2016). DOI: 10.48550/ARXIV.1609.04747.

[9] B. Wohlberg. "ADMM Penalty Parameter Selection by Residual Balancing" (2017). DOI: 10.48550/ARXIV.1704.06209.

[10] I. Higgins, L. Matthey, A. Pal, et al. "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework" (2017).

[11] L. A. Shepp and Y. Vardi. "Maximum Likelihood Reconstruction for Emission Tomography". *IEEE Transactions on Medical Imaging* 1.2 (1982), pp. 113–122. DOI: 10.1109/TMI.1982.4307558.

[12] W. van Aarle, W. J. Palenstijn, J. Cant, et al. "Fast and flexible X-ray tomography using the ASTRA toolbox". *Opt. Express* 24.22 (2016), pp. 25129–25147. DOI: 10.1364/OE.24.025129.

[13] S. Zhao, J. Song, and S. Ermon. "InfoVAE: Information Maximizing Variational Autoencoders" (2017). DOI: 10.48550/ARXIV.1706.02262.

[14] K. Pandey, A. Mukherjee, P. Rai, et al. *DiffuseVAE: Efficient, Controllable and High-Fidelity Generation from Low-Dimensional Latents*. 2022. DOI: 10.48550/ARXIV.2201.00308.