Convergent Plug-and-Play algorithms for positron emission tomography reconstruction

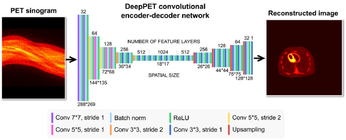
M. Savanier F. Sureau C. Comtat

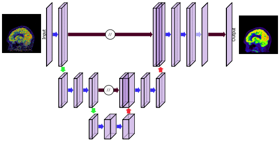
BioMaps, Université Paris-Saclay, CEA, CNRS, Inserm, SHFJ

Deep CosmoStat Days, January 2025

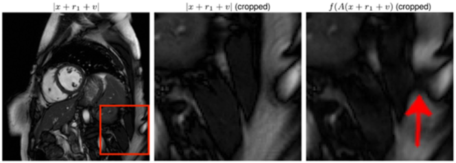
Deep Learning and medical image reconstruction

- Wide diversity of Deep Learning techniques (DL) to solve inverse problems with promising experimental results



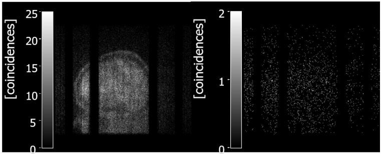


- Stable and plausible instabilities ("hallucinations") observed in medical image reconstruction with DL (Antun'20, Gottschling'20)



Stable instabilities (Vegard'20)

- **Ill-posed tomographic inverse problem with Poisson data** $\Rightarrow$ *Instabilities at low-dose?*



Effect of dose reduction (/60) on measured data

- Learning/validating in a typically low data regime in a medical context ($\mathcal{O}(10-100)$ exams) $\Rightarrow$ *Robustness?*
- Large scale tomographic 3D/4D inverse problem ($\mathcal{O}(10^7-10^8)$ variables) $\Rightarrow$ *Numerical efficiency?*

Aim

Develop & validate **robust and numerically efficient** low-count PET reconstruction schemes using DL

Developing robust DL methods

Strategy

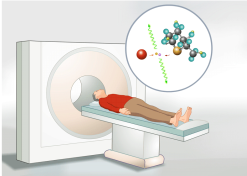
- Focus on learning what need to be learned (not the forward model!)
- Focus on supervised learning even though small research databases (constrained learning, fewer parameters)
- Use tools from statistics, optimization to understand robustness issues and propose a robust reconstruction method
- Develop validation tools for PET DL reconstructed images?

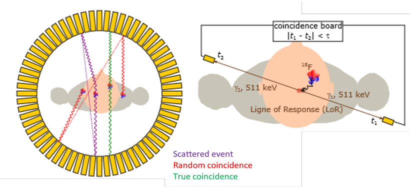
Hybrid techniques:

- Reconstruction bricks/layers from convex optimization
- Learned adaptive (implicit) regularization
- More control on the reconstruction (mathematical characterization)

	Unfolding	Synthesis	Plug-and-Play
Learning	End-to-end	offline	offline
Optimisation with network	End-to-end	Yes	No
Memory load	$\propto N_{unroll} N_{params}$	$\propto N_{params}$	$\propto N_{params}$
Convergence	Not in practice (N_{unroll})	?	Yes

PET model for reconstruction





Statistical forward model (physics, geometry of scanner, potentially pharmacodynamics) for quantitative imaging

$$y_{it} = \mathcal{P}\left(\langle \mathbf{h}_i, \mathbf{x}_t \rangle + \underbrace{\bar{s}_{it} + \bar{r}_{it}}_{\bar{b}_{it}} \right)$$

y_{it}: data in LOR i frame t

$\mathbf{h}_i = [h_{ij}]_{j \in [1,J]}$: line i of $\mathbf{H}$

$\mathbf{x}_t \in \mathbb{R}^J$ activity for frame t

$\bar{s}_{it}$: scatter, $\bar{r}_{it}$: randoms expectations

Applications in:

- Oncology
- Neurology
- Pharmacology

Model-based reconstruction (e.g. ML-EM)

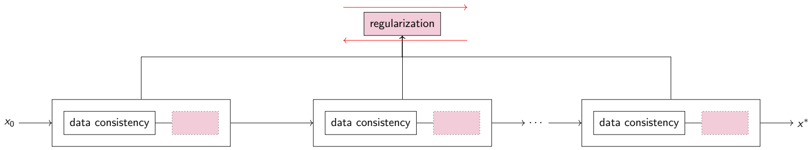
$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in[0,+\infty[^N}{\operatorname{argmin}} \underbrace{f(\mathbf{x};\mathbf{y},\mathbf{b}) + R(\mathbf{x})}_{C(\mathbf{x})} \iff 0 \in \partial(f + \iota_{[0,+\infty[^N} + R)(\hat{\mathbf{x}})$$

where $f(\mathbf{x};\mathbf{y},\mathbf{b}) = \sum_{m=1}^{M} [\mathbf{y}]_m \log(\frac{[\mathbf{y}]_m}{[\mathbf{Hx}+\mathbf{b}]_m}) + [\mathbf{Hx}+\mathbf{b}]_m - [\mathbf{y}]_m$.

Many algorithms T_C such that $\hat{\mathbf{x}} = T_C(\hat{\mathbf{x}}, \mathbf{y})$.

Choice for T_C: common reconstruction algorithms are based on

- with smooth priors: Majorization-Minimization with Bregman majorants (Rossignol'22), Forward-Backward
- with non smooth priors: ADMM/Douglas-Rachford

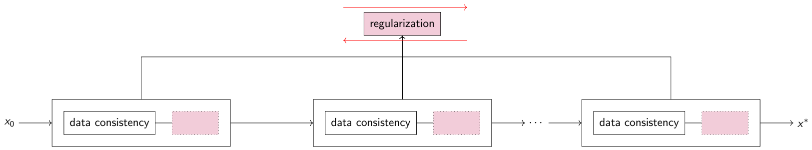


PnP iterations

$$(\forall n \in \mathbb{N}) \qquad \mathbf{x}^{n+1} = T_C(\mathbf{x}^n, \partial R, \mathbf{y}) \qquad \mathbf{x}^{n+1} = T_C(\mathbf{x}^n, \mathrm{prox}_R, \mathbf{y})$$

$$(\forall n \in \mathbb{N}) \qquad \mathbf{x}^{n+1} = T_C(\mathbf{x}^n, NN, \mathbf{y})$$

1. How to choose T_C?
2. How to choose NN such that $(\mathbf{x}^n)_{n\in\mathbb{N}}$ converges to some $\bar{\mathbf{x}}$?
3. Can we characterize $\bar{\mathbf{x}}$?
4. Can we control to which $\bar{\mathbf{x}}$ the sequence $(\mathbf{x}^n)_{n\in\mathbb{N}}$ converges?



PnP iterations

$$(\forall n \in \mathbb{N}) \qquad \mathbf{x}^{n+1} = T_C(\mathbf{x}^n, \partial R, \mathbf{y}) \qquad \mathbf{x}^{n+1} = T_C(\mathbf{x}^n, \mathrm{prox}_R, \mathbf{y})$$

$$(\forall n \in \mathbb{N}) \qquad \mathbf{x}^{n+1} = T_C(\mathbf{x}^n, NN, \mathbf{y})$$

1. How to choose T_C?
2. How to choose NN such that $(\mathbf{x}^n)_{n\in\mathbb{N}}$ converges to some $\overline{\mathbf{x}}$?
3. Can we characterize $\overline{\mathbf{x}}$?
4. Can we control to which $\overline{\mathbf{x}}$ the sequence $(\mathbf{x}^n)_{n\in\mathbb{N}}$ converges?

Algorithm PnP ADMM (Pesquet'21)

Require: $\mathbf{D}_\Theta : \mathbb{R}^N \mapsto [0, +\infty[^N$

for $n = 0$ to $N-1$ **do**

$\mathbf{x}^{n+1} = \text{prox}_{\mu\lambda f}(\mathbf{z}^n - \mathbf{u}^n)$

$\mathbf{z}^{n+1} = \text{prox}_{\mu R}(\mathbf{x}^{n+1} + \mathbf{u}^n)$

$\mathbf{z}^{n+1} = \mathbf{D}_\Theta(\mathbf{x}^{n+1} + \mathbf{u}^n)$

$\mathbf{u}^{n+1} = \mathbf{u}^n + \mathbf{x}^{n+1} - \mathbf{z}^{n+1}$

end for

Algorithm PnP FB (Hurault'22)

for $n = 0$ to $N-1$ **do**

$\mathbf{x}^{n+1} = \text{prox}_{\tau f}(1 - \tau\nabla R(\mathbf{x}^n))$

Backtracking on τ given C, $\mathbf{x}^{n+1}$ and $\mathbf{x}^n$

$\mathbf{x}^{n+1} = \text{prox}_{\tau f}(\frac{\tau}{\lambda}\mathbf{D}_{\Theta,\text{GS}}(\mathbf{x}^n) + (1 - \frac{\tau}{\lambda})\mathbf{x}^n)$

end for

When $\mathbf{D}_\Theta$ is the resolvent of a maximal monotone operator (MMO) i.e. $\mathbf{D}_\Theta$ is FNE, $\mathbf{x}^n$ and $\mathbf{z}^n$ converge to $\bar{\mathbf{x}}$

⚠ provided there exists at least 1 fixed point

If $\mathbf{Id} - \mathbf{D}_{\Theta,\text{GS}} = \nabla R_\Theta$ is L-Lipschitz, $\lambda L > \tau > 0$

- $\mathbf{x}^n$ converges to $\bar{\mathbf{x}}$, such that for $C = f + \iota_{[0,+\infty[^N} + R_\Theta/\lambda$, $\partial C(\bar{\mathbf{x}})/\partial\bar{\mathbf{x}} = 0$
- $C(\mathbf{x}^n)$ is non-increasing.

⚠ no uniqueness of stationary points

Algorithm PnP ADMM (Pesquet'21)

Require: $\mathbf{D}_\Theta : \mathbb{R}^N \mapsto [0, +\infty[^N$
  for $n = 0$ to $N-1$ **do**
    $\mathbf{x}^{n+1} = \mathrm{prox}_{\mu\lambda f}(\mathbf{z}^n - \mathbf{u}^n)$
    $\mathbf{z}^{n+1} = \mathrm{prox}_{\mu R}(\mathbf{x}^{n+1} + \mathbf{u}^n)$
    $\mathbf{z}^{n+1} = \mathbf{D}_\Theta(\mathbf{x}^{n+1} + \mathbf{u}^n)$
    $\mathbf{u}^{n+1} = \mathbf{u}^n + \mathbf{x}^{n+1} - \mathbf{z}^{n+1}$
  end for

$2\mathbf{D}_\Theta - \mathbf{Id}$ is 1-Lipschitz:
$\longrightarrow$ Lipschitz regularization

$\beta \max\{|||J_{2\mathbf{D}_\Theta - \mathbf{Id}}(\tilde{\mathbf{x}})||| + \epsilon - 1, 0\}^{1+\alpha}$ with

$\tilde{\mathbf{x}} = \kappa \bar{\mathbf{x}}_{\mathrm{EM}} + (1-\kappa)\mathbf{x}_{\mathrm{in}},\ \kappa \sim \mathcal{U}[0,1]$

Algorithm PnP FB (Hurault'22)

  for $n = 0$ to $N-1$ **do**
    $\mathbf{x}^{n+1} = \mathrm{prox}_{\tau f}(1 - \tau\nabla R(\mathbf{x}^n))$
    Backtracking on τ given C, $\mathbf{x}^{n+1}$ and $\mathbf{x}^n$
    $\mathbf{x}^{n+1} = \mathrm{prox}_{\tau f}(\frac{\tau}{\lambda}\mathbf{D}_{\Theta,\mathrm{GS}}(\mathbf{x}^n) + (1 - \frac{\tau}{\lambda})\mathbf{x}^n)$
  end for

$\mathbf{D}_{\Theta,\mathrm{GS}} = \mathbf{Id} - \nabla R_\Theta$: compose potential function (e.g. $\|\mathbf{Id} - \cdot\|^2$) with some $\mathbf{N}_\Theta$ and compute gradient ($\mathbf{D}_{\Theta,\mathrm{GS}} \neq \mathbf{N}_\Theta$)

$\mathbf{D}_{\Theta,\mathrm{GS}}(\mathbf{x}) = \mathbf{N}_\Theta(\mathbf{x}) + J^\top_{\mathbf{N}_\Theta(x)}(\mathbf{x} - \mathbf{N}_\Theta(\mathbf{x}))$

Learning the prior in PnP

Model agnostic

- Off-the-shelf non deep denoisers (Heide'14)
- Gaussian deep denoisers as prox surrogates: Bayesian interpretation (Meinhardt'17, Pesquet'21)
- Denoising score matching for learning ∇R

Model dependent

- In PET: Prox surrogate mapping low-to-standard dose images (Sureau'21)
- Adversarial regularization noisy and "clean" images (Cohen'21, Chand'24)

Model agnostic

- Off-the-shelf non deep denoisers (Heide'14)
- Gaussian deep denoisers as prox surrogates: Bayesian interpretation (Meinhardt'17, Pesquet'21)
- Denoising score matching for learning ∇R

Model dependent

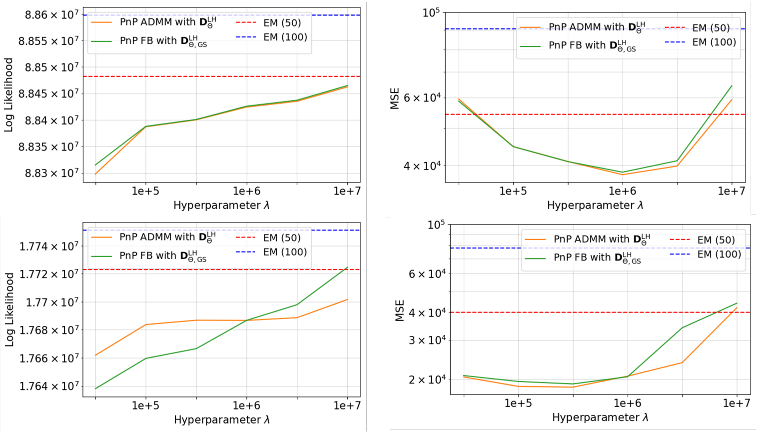
- In PET: Prox surrogate mapping low-to-standard dose images (Sureau'21)
- Adversarial regularization noisy and "clean" images (Cohen'21, Chand'24)

Training $\mathbf{D}_\Theta$ and $\mathbf{D}_{\Theta,\mathrm{GS}}$ as low- to high-dose denoisers

- **Brain simulation [^{18}F]-FDG (Biograph 6 TrueP/TrueV)**
- MRI/PET on 14 patients (11 for training)
- PET piecewise constant phantoms (100 anatomical regions)
- Simulations with normalization, attenuation, scatter, randoms, resolution modeling (4mm)
- Augmentation with dose variations (11 patients x 10 doses)
- References = CASToR reconstructions
- Inputs = CASToR reconstructions with fewer counts (/ 5)
- Differentiable U-net like architectures

When $\mathbf{D}_\Theta^{\mathrm{LH}}$ and $\mathbf{D}_{\Theta,\mathrm{GS}}^{\mathrm{LH}}$ are trained on the same task, winner between PnP FB and PnP ADMM?

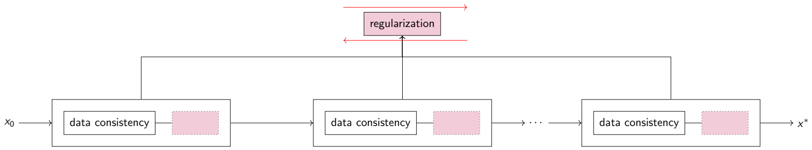
In principle, CNS on PnP FB lighter but...



$\longrightarrow$ Similar images with PnP ADMM and PnP FB

$\longrightarrow$ High sensitivity to hyperparameters (oversmooth)

$\longrightarrow$ What can be said about the fixed point?



PnP iterations

$$(\forall n \in \mathbb{N}) \qquad \mathbf{x}^{n+1} = T_C(\mathbf{x}^n, \partial R, \mathbf{y}) \qquad \mathbf{x}^{n+1} = T_C(\mathbf{x}^n, \mathrm{prox}_R, \mathbf{y})$$

$$(\forall n \in \mathbb{N}) \qquad \mathbf{x}^{n+1} = T_C(\mathbf{x}^n, NN, \mathbf{y})$$

1. How to choose T_C?
2. How to choose NN such that $(\mathbf{x}^n)_{n\in\mathbb{N}}$ converges to some $\overline{\mathbf{x}}$?
3. Can we characterize $\overline{\mathbf{x}}$?
4. Can we control to which $\overline{\mathbf{x}}$ the sequence $(\mathbf{x}^n)_{n\in\mathbb{N}}$ converges?

Characterizing the fixed points

- PnP FB with learned gradient $\mathbf{D}_{\Theta,\mathrm{GS}}$ (for $\overline{\mathbf{x}} \in \mathrm{int}([0,+\infty[^N))$

$$0 \in \nabla f(\overline{\mathbf{x}};\mathbf{y}) + (\overline{\mathbf{x}} - \mathbf{D}_{\Theta,\mathrm{GS}}(\overline{\mathbf{x}}))/\lambda$$

- PnP ADMM with learned prox surrogate $\mathbf{D}_{\Theta}$ (for $\overline{\mathbf{x}} \in \mathrm{int}([0,+\infty[^N))$

$$\overline{\mathbf{x}} = \mathbf{D}_{\Theta}(\overline{\mathbf{x}} - \lambda \nabla f(\overline{\mathbf{x}};\mathbf{y})) \qquad \text{(FP)}$$

$\longrightarrow$ Use (FP) for training $\mathbf{D}_{\Theta}$ in PnP ADMM

In practice

We want $\mathbf{x}_{\mathrm{HD}} = \mathbf{D}_\Theta(\mathbf{x}_{\mathrm{HD}} - \lambda_{\mathrm{Train}} \nabla f(\mathbf{x}_{\mathrm{HD}}; \mathbf{y}_{\mathrm{LD}}))$, we choose to minimize

$$\frac{\|\mathbf{D}_\Theta(\mathbf{x}_{\mathrm{HD}} - \lambda_{\mathrm{Train}} \nabla f(\mathbf{x}_{\mathrm{HD}}; \mathbf{y}_{\mathrm{LD}})) - \mathbf{x}_{\mathrm{HD}}\|^2}{\|\mathbf{x}_{\mathrm{HD}}\|^2}; \quad \lambda_{\mathrm{Train}} = \alpha_{\mathrm{Train}} \times \sqrt{\|\mathbf{H}\mathbf{x}_{\mathrm{HD}} + \mathbf{b}\|_1}$$

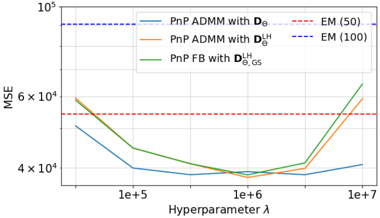
and we want the existence and uniqueness of the fixed point

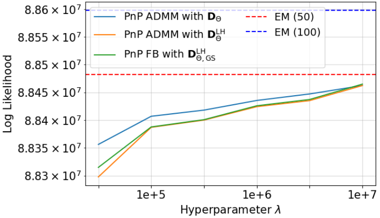
$$f \qquad \Longleftrightarrow \qquad f + \frac{\zeta}{2}\| \cdot - \mathbf{x}_{\mathrm{EM}}\|^2 \quad (\zeta = 10^{-6})$$

Féjer monotonicity of $(\mathbf{x}^{n+1} + \mathbf{u}^n) \longrightarrow \beta \max\{|||J_{2\mathbf{D}_\Theta - \mathbf{Id}}(\tilde{\mathbf{x}})||| + \epsilon - 1, 0\}^{1+\alpha}$ with $\tilde{\mathbf{x}} \in B_{\|\mathbf{x}_{\mathrm{EM}} - \mathbf{x}_{\mathrm{in}}\|}(\mathbf{x}_{\mathrm{in}})$ given $\mathbf{x}^0 = \mathbf{x}_{\mathrm{EM}}$ and $\mathbf{u}^0 = 0$

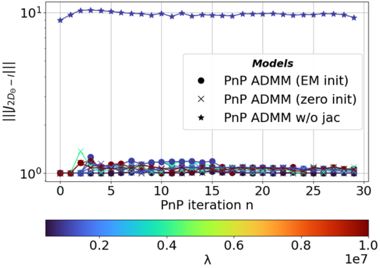
⚠ Only $\mathbf{D}_\Theta$ locally FNE needed

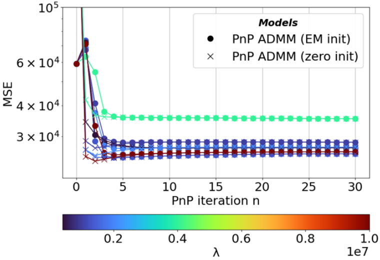
Evaluation with fixed point resolvent





$\longrightarrow$ lowest MSE similar with previous PnP-ADMM with $\mathbf{D}_\Theta^{\mathrm{LH}}$ and PnP-FB with $\mathbf{D}_{\Theta,\mathrm{GS}}^{\mathrm{LH}}$ but across a wider range of λ

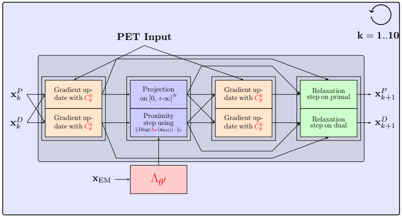




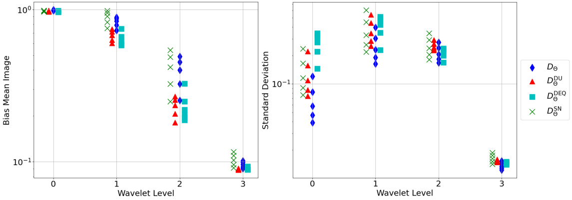
Evaluation of different FNE architectures

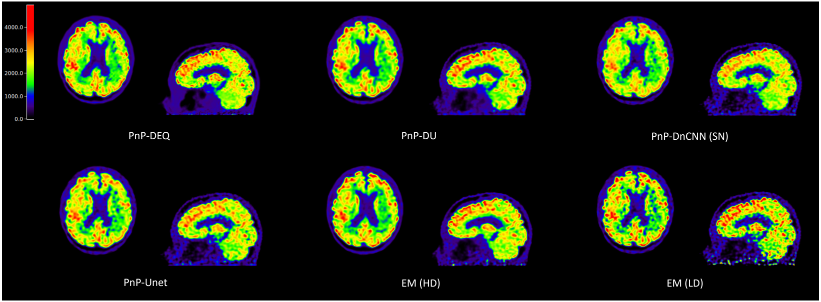
⚠ Jacobian regularization ($\beta \max\{|||J_{2\mathbf{D}_\Theta - \mathbf{Id}}(\tilde{\mathbf{x}})||| + \epsilon - 1, 0\}^{1+\alpha}$) is costly and very sensitive to HP

	Architecture	FNE by design	# params	Approx. time/epoch	# epochs with jac	Runtime
$\mathbf{D}_\Theta$	U-net	No	1 079 000	2h	20	Fast
$\mathbf{D}_\Theta^{\mathrm{SN}}$	DnCNN with spectral norm	Yes	167 620	1min	-	Fast
$\mathbf{D}_\Theta^{\mathrm{DU}}$	Unfolding	No but close (Pustelnik'23)	59 689	3h	3	$\propto N_{\mathrm{layers}} = 10$
$\mathbf{D}_\Theta^{\mathrm{DEQ}}$	Deep equilibrium	Yes	18 298	2h	-	$\propto N_{\mathrm{iter}} = 1000$

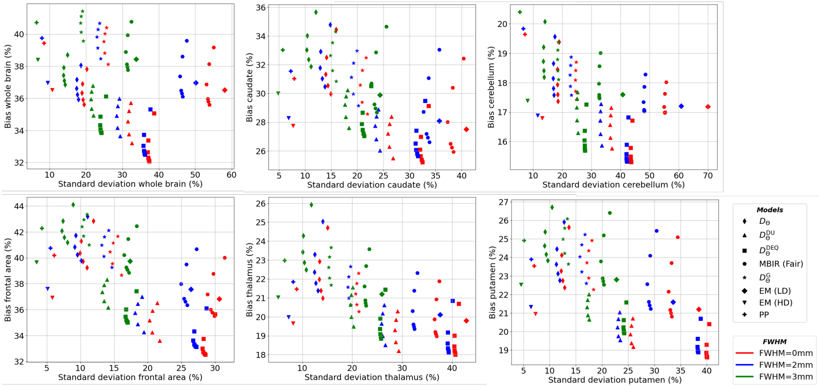


⟶ Evaluation on 50 data replicates from the same phantom (multiscale/ROI-based analysis)

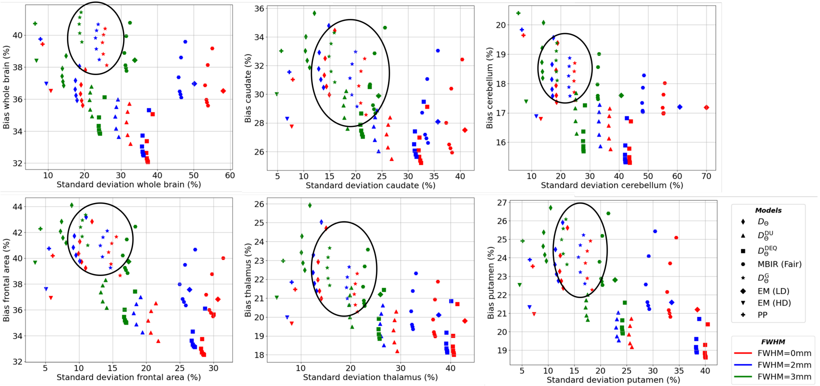




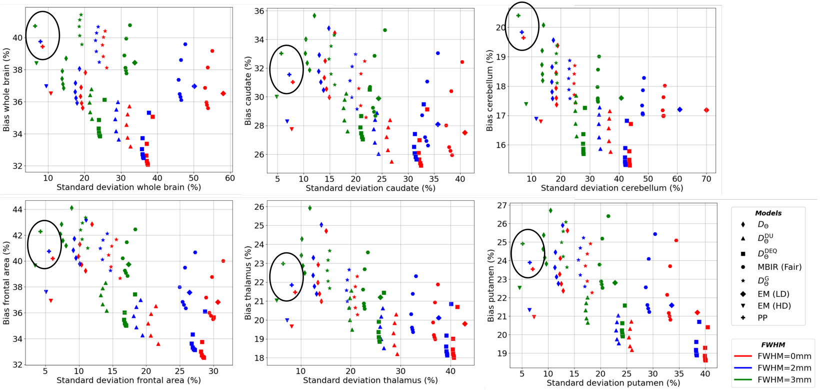
⟶ SN not competitive



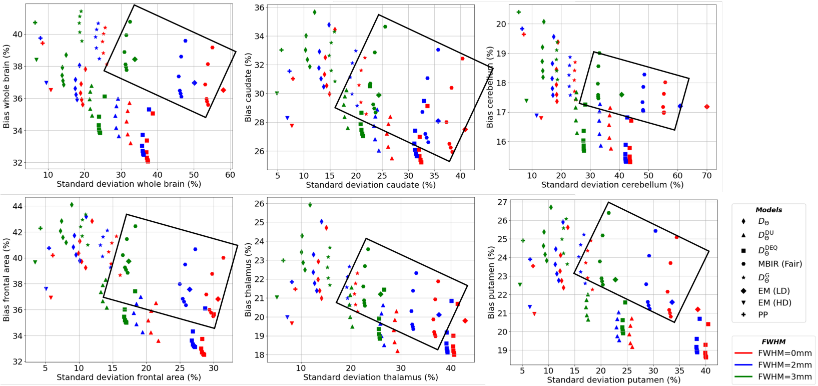
$\longrightarrow$ Outperforms MBIR and PP + PnP with Gaussian denoiser



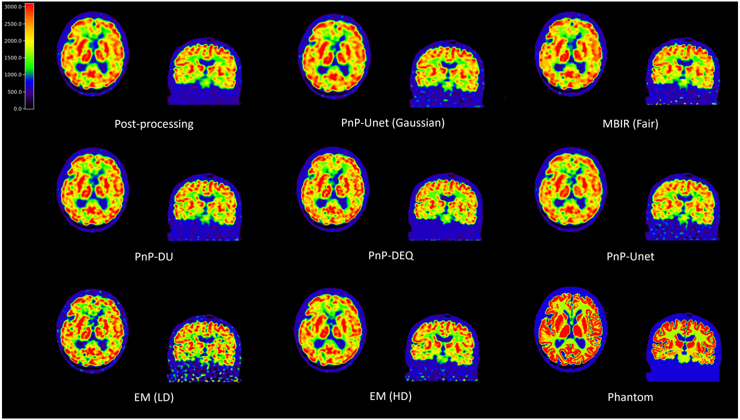
$\longrightarrow$ Outperforms MBIR and PP + PnP with Gaussian denoiser

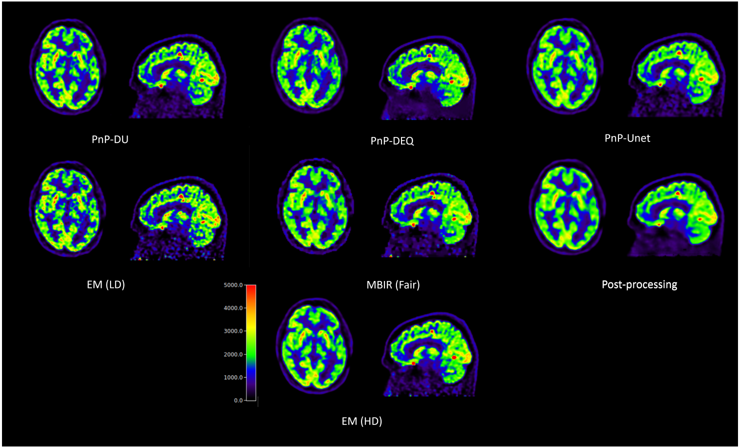


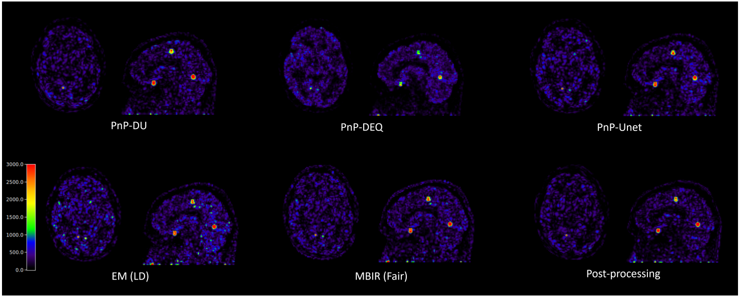
$\longrightarrow$ Outperforms MBIR and PP + PnP with Gaussian denoiser

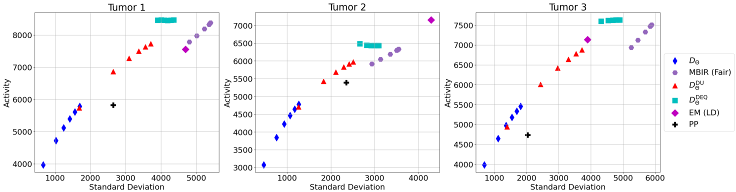


$\longrightarrow$ Outperforms MBIR and PP + PnP with Gaussian denoiser


3000.0
2500.0
2000.0
1500.0
1000.0
500.0
0.0
Post-processing
PnP-Unet (Gaussian)
MBIR (Fair)
PnP-DU
PnP-DEQ
PnP-Unet
EM (LD)
EM (HD)
Phantom


PnP-DU
PnP-DEQ
PnP-Unet
EM (LD)
MBIR (Fair)
Post-processing
5000.0
4000.0
3000.0
2000.0
1000.0
0.0
EM (HD)





$\longrightarrow$ DEQ (and DU) more robust to out-of-distribution examples

Conclusions

- PnP reconstruction for PET based on ADMM/DR with convergence guarantees and existence of a unique fixed point (available in CASToR - F. Sureau)
- interest of learning a resolvent of a MMO for reconstruction
- constrained network (DEQ, DU) embedding prior knowledge easier training/more robust
- No need for overparameterized operators

Perspectives

- investigate further relationship between optimization and Jacobian regularization
- MRI conditioning of the learned operator
- investigate further the robustness of our approach to data perturbation

[1] Emilie Chouzenoux et al. "Convergence Results for Primal-Dual Algorithms in the Presence of Adjoint Mismatch". In: *SIAM Journal on Imaging Sciences* (Jan. 2023). Publisher: Society for Industrial and Applied Mathematics University City, Philadelphia. DOI: 10.1137/22M1490223.

[2] Regev Cohen et al. "It Has Potential: Gradient-Driven Denoisers for Convergent Solutions to Inverse Problems". In: *Advances in Neural Information Processing Systems*. Vol. 34. 2021, pp. 18152–18164.

[3] Minh N. Dao and Hung M. Phan. "Adaptive Douglas–Rachford Splitting Algorithm for the Sum of Two Operators". In: *SIAM Journal on Optimization* 29.4 (2019), pp. 2697–2724. DOI: 10.1137/18M121160X.

[4] Alexis Goujon, Sebastian Neumayer, and Michael Unser. "Learning Weakly Convex Regularizers for Convergent Image-Reconstruction Algorithms". In: (2023). Publisher: arXiv Version Number: 1. DOI: 10.48550/ARXIV.2308.10542.

[5] Samuel Hurault, Arthur Leclaire, and Nicolas Papadakis. *Gradient Step Denoiser for convergent Plug-and-Play*. 2022. DOI: 10.48550/arXiv.2110.03220.

[6] Samuel Hurault et al. *Convergent Bregman Plug-and-Play Image Restoration for Poisson Inverse Problems*. en. June 2023. URL: http://arxiv.org/abs/2306.03466.

[7] Jean-Christophe Pesquet et al. "Learning Maximally Monotone Operators for Image Recovery". en. In: *SIAM Journal on Imaging Sciences* 14.3 (Jan. 2021), pp. 1206–1237. DOI: 10.1137/20M1387961.

[8] M. Savanier, Claude Comtat, and Florent Sureau. "Learning with fixed point condition for convergent PnP PET reconstruction". In: *ISBI 2024 - 21st IEEE International Symposium on Biomedical Imaging*. Athenes, Greece, May 2024.

[9] Florent Sureau et al. "Convergent ADMM Plug and Play PET Image Reconstruction". In: *Proceedings of the 17th International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*. 2023. DOI: 10.48550/arXiv.2310.16846.