Mathematics of Computational Imaging Systems

Plug-and-play (PnP)

Nelly Pustelnik

CNRS, ENSL, Laboratoire de physique, F-69342 Lyon, France

Towards deep learning: Plug-and-play

Deep learning – *General framework*

- Database : $\mathcal{S} = \left\{ (\overline{\mathrm{x}}_i, \mathrm{z}_i) \in \mathbb{R}^N \times \mathbb{R}^M \;\middle|\; i \in \mathbb{I} \cup \mathbb{J} \text{ such that } \mathrm{z}_i = \overline{\mathrm{x}}_i + \varepsilon \right\}$
- Prediction function : $d_\Theta(\mathrm{z}_i) = \eta^{[K]}\big(W^{[K]} \dots \eta^{[1]}(W^{[1]}\mathrm{z}_i + b^{[1]}) \dots + b^{[K]}\big)$

Variational formulation versus Plug and play

- **Forward-Backward**: $\mathrm{x}^{[k+1]} = \mathrm{prox}_{\tau\theta f \circ \mathrm{D}}(\ \mathrm{x}^{[k]} - \tau \mathrm{A}^\top(\mathrm{A}\mathrm{x}^{[k]} - \mathrm{z})\ )$
- **FB-PnP**: $\mathrm{x}^{[k+1]} = \mathrm{d}_\Theta(\ \mathrm{x}^{[k]} - \tau \mathrm{A}^\top(\mathrm{A}\mathrm{x}^{[k]} - \mathrm{z})\ )$

Remark

- Principle that can be applied to other algorithmic schemes PnP-ADMM, PnP-DR, HQS,...

Towards deep learning: Plug-and-play

- **FB-PnP**: $$\mathrm{x}^{[k+1]} = \mathrm{d}_\Theta\left(\, \mathrm{x}^{[k]} - \tau \mathrm{A}^\top(\mathrm{A}\mathrm{x}^{[k]} - \mathrm{z})\,\right)$$

Build a denoiser: MAP denoiser

- Defining the estimate $\mathrm{d}_\Theta(\mathrm{z})$ of $\overline{\mathrm{x}}$ from z as the maximum of $\pi(\mathrm{x}|\mathrm{z})$:

$$\mathrm{d}_\Theta(\mathrm{z}) \in \operatorname*{Argmin}_{\mathrm{x}} - \log \pi(\mathrm{z}|\mathrm{x}) - \log \pi(\mathrm{x}).$$

For white Gaussian noise, $-\log \pi(\mathrm{z}|\mathrm{x}) = \frac{1}{2\sigma^2}\|\mathrm{z} - \mathrm{x}\|_2^2$.
Assume that $-\log \pi(\mathrm{x}) = g(\mathrm{x})$. Then,

$$\mathrm{d}_\Theta(\mathrm{z}) = \mathrm{prox}_{\sigma^2 g}(\mathrm{z}).$$

Remark

- g often unknown. This can explain the desire of replacing proximity operator by more powerful denoisers.

Towards deep learning: Plug-and-play

- **FB-PnP**: $\mathrm{x}^{[k+1]} = \mathrm{d}_\Theta(\, \mathrm{x}^{[k]} - \tau \mathrm{A}^\top(\mathrm{A}\mathrm{x}^{[k]} - \mathrm{z})\,)$

Build a denoiser: MAP denoiser

- Defining the estimate $\mathrm{d}_\Theta(\mathrm{z})$ of $\overline{\mathrm{x}}$ from z as the maximum of $\pi(\mathrm{x}|\mathrm{z})$:

$$\mathrm{d}_\Theta(\mathrm{z}) = \underset{\mathrm{x}}{\mathrm{Argmin}} - \log \pi(\mathrm{z}|\mathrm{x}) - \log \pi(\mathrm{x}).$$

For white Gaussian noise, $-\log \pi(\mathrm{z}|\mathrm{x}) = \frac{1}{2\sigma^2}\|\mathrm{z} - \mathrm{x}\|_2^2$.
Assume that $-\log \pi(\mathrm{x}) = g(\mathrm{x})$. Then,

$$\mathrm{d}_\Theta(\mathrm{z}) = \mathrm{prox}_{\sigma^2 g}(\mathrm{z}).$$

Remark

- g often unknown. This can exmplain the desire of replacing proximity operator by more powerful denoisers.

Towards deep learning: Plug-and-play

FB-PnP: $$\mathrm{x}^{[k+1]} = \mathrm{d}_\Theta\left(\, \mathrm{x}^{[k]} - \tau \mathrm{A}^\top(\mathrm{A}\mathrm{x}^{[k]} - \mathrm{z})\,\right)$$

Build a denoiser: MMSE denoiser $\mathrm{d}_\Theta(\mathrm{z}) = \mathbb{E}(\mathrm{x}|\mathrm{z})$

- $\widetilde{\pi}(\mathrm{z}) \propto \exp(-\widetilde{g}_{\sigma^2}(\mathrm{z}))$: probability distribution for the noisy signal.
- Tweedie's formula: $\frac{\mathrm{z}-\mathrm{d}_\Theta(\mathrm{z})}{\sigma^2} = \nabla \widetilde{g}_{\sigma^2}(\mathrm{z})$,
- $\nabla \widetilde{g}_{\sigma^2}$: score of this distribution.
- $\widetilde{\pi}(\mathrm{z}) \neq \pi(\mathrm{x})$: the former is the convolution of the latter with a Gaussian smoothing kernel of bandwidth σ.

(Gradient)-FB-PnP

$$\mathrm{x}^{[k+1]} = \mathrm{prox}_{\frac{\tau}{2}\|\mathrm{A}\cdot - \mathrm{z}\|^2}\left(\, \mathrm{x}^{[k]} - \tau\sigma^{-2}(\mathrm{x}^{[k]} - \mathrm{d}_\Theta(\mathrm{x}^{[k]}))\,\right)$$

Towards deep learning: Plug-and-play

- **FB-PnP**: $\mathrm{x}^{[k+1]} = \mathrm{d}_\Theta\left(\mathrm{x}^{[k]} - \tau \mathrm{A}^\top(\mathrm{A}\mathrm{x}^{[k]} - \mathrm{z}) \right)$

Build a denoiser

- If $\mathrm{d}_\theta = \mathrm{prox}_{\tau\theta f}$ with f **convex** and $\tau < 2/\|\mathrm{A}\|^2$, then convergence to $\widehat{\mathrm{x}} \in \underset{\mathrm{x}}{\mathrm{Argmin}}\ \frac{1}{2}\|\mathrm{A}\mathrm{x} - \mathrm{z}\|_2^2 + \theta f(\mathrm{x})$.

Towards deep learning: Plug-and-play

- **FB-PnP**: $$x^{[k+1]} = d_\Theta\left(x^{[k]} - \tau A^\top (Ax^{[k]} - z) \right)$$

Build a denoiser

- If $d_\theta = \mathrm{prox}_{\tau\theta f}$ with f **convex** and $\tau < 2/\|A\|^2$, then convergence to $\widehat{x} \in \underset{x}{\mathrm{Argmin}}\ \frac{1}{2}\|Ax - z\|_2^2 + \theta f(x)$.
- If $d_\theta = \mathrm{prox}_{\tau\theta f}$ with f **nonconvex** and $\tau < 1/\|A\|^2$, then convergence to $\widehat{x} \in \underset{x}{\mathrm{Argmin}}\ \frac{1}{2}\|Ax - z\|_2^2 + \theta f(x)$.[Bolte, Sabach, Teboulle, 2014]

Towards deep learning: Plug-and-play

• **FB-PnP**: $$\mathrm{x}^{[k+1]} = \mathrm{d}_\Theta\left(\mathrm{x}^{[k]} - \tau \mathrm{A}^\top(\mathrm{A}\mathrm{x}^{[k]} - \mathrm{z}) \right)$$

Build a denoiser

• If $\mathrm{d}_\theta = \mathrm{prox}_{\tau\theta f}$ with f **convex** and $\tau < 2/\|\mathrm{A}\|^2$, then convergence to $\widehat{\mathrm{x}} \in \underset{\mathrm{x}}{\mathrm{Argmin}}\ \frac{1}{2}\|\mathrm{A}\mathrm{x} - \mathrm{z}\|_2^2 + \theta f(\mathrm{x})$.

• If $\mathrm{d}_\theta = \mathrm{prox}_{\tau\theta f}$ with f **nonconvex** and $\tau < 1/\|\mathrm{A}\|^2$, then convergence to $\widehat{\mathrm{x}} \in \underset{\mathrm{x}}{\mathrm{Argmin}}\ \frac{1}{2}\|\mathrm{A}\mathrm{x} - \mathrm{z}\|_2^2 + \theta f(\mathrm{x})$.[Bolte, Sabach, Teboulle, 2014]

• If d_Θ is built to be **firmly non-expansive** by regularizing the training loss. For instance $\mathrm{d}_\Theta = \frac{\mathrm{Id}+Q}{2}$ with Q a non-expansive operator, then convergence to an inclusion problem [Pesquet, Repetti, Terris, Wiaux, 2021].

Towards deep learning: Plug-and-play

• **FB-PnP**: $x^{[k+1]} = d_\Theta\left(x^{[k]} - \tau A^\top (Ax^{[k]} - z) \right)$

Build a denoiser

• If $d_\theta = \text{prox}_{\tau\theta f}$ with f **convex** and $\tau < 2/\|A\|^2$, then convergence to $\widehat{x} \in \underset{x}{\text{Argmin}}\ \frac{1}{2}\|Ax - z\|_2^2 + \theta f(x)$.

• If $d_\theta = \text{prox}_{\tau\theta f}$ with f **nonconvex** and $\tau < 1/\|A\|^2$, then convergence to $\widehat{x} \in \underset{x}{\text{Argmin}}\ \frac{1}{2}\|Ax - z\|_2^2 + \theta f(x)$.[Bolte, Sabach, Teboulle, 2014]

• If d_Θ is built to be **firmly non-expansive** by regularizing the training loss. For instance $d_\Theta = \frac{\text{Id}+Q}{2}$ with Q a non-expansive operator, then convergence to an inclusion problem [Pesquet, Repetti, Terris, Wiaux, 2021].

• If d_Θ is **built to be the proximal operator** of a nonconvex (weakly convex) functional, then convergence to a minimization problem. [Hurault, Leclaire, Papadakis, 2022]

Towards deep learning: Plug-and-play

• **FB-PnP**: $$\mathrm{x}^{[k+1]} = \mathrm{d}_\Theta\left(\mathrm{x}^{[k]} - \tau \mathrm{A}^\top(\mathrm{A}\mathrm{x}^{[k]} - \mathrm{z}) \right)$$

Build a denoiser

• If $\mathrm{d}_\theta = \mathrm{prox}_{\tau\theta f}$ with f **convex** and $\tau < 2/\|\mathrm{A}\|^2$, then convergence to $\widehat{\mathrm{x}} \in \underset{\mathrm{x}}{\mathrm{Argmin}}\ \frac{1}{2}\|\mathrm{A}\mathrm{x} - \mathrm{z}\|_2^2 + \theta f(\mathrm{x})$.

• If $\mathrm{d}_\theta = \mathrm{prox}_{\tau\theta f}$ with f **nonconvex** and $\tau < 1/\|\mathrm{A}\|^2$, then convergence to $\widehat{\mathrm{x}} \in \underset{\mathrm{x}}{\mathrm{Argmin}}\ \frac{1}{2}\|\mathrm{A}\mathrm{x} - \mathrm{z}\|_2^2 + \theta f(\mathrm{x})$.[Bolte, Sabach, Teboulle, 2014]

• If d_Θ is built to be **firmly non-expansive** by regularizing the training loss. For instance $\mathrm{d}_\Theta = \frac{\mathrm{Id}+Q}{2}$ with Q a non-expansive operator, then convergence to an inclusion problem [Pesquet, Repetti, Terris, Wiaux, 2021].

• If d_Θ is **built to be the proximal operator** of a nonconvex (weakly convex) functional, then convergence to a minimization problem. [Hurault, Leclaire, Papadakis, 2022]

• If d_Θ is built from an unfolded strategy to compute $\mathrm{prox}_{\|\mathrm{L}\cdot\|_1 + \iota_C}$.

Stability

☛ Given an input z and a perturbation ϵ, the error on the output can be upper bounded :

$$\|\mathrm{d}_\Theta(\mathrm{z}+\epsilon) - \mathrm{d}_\Theta(\mathrm{z})\| \leq \chi\|\epsilon\|.$$

where χ certificated of the robustness.

☛ [Combettes, Pesquet, 2020]: χ can be upper bounded by:

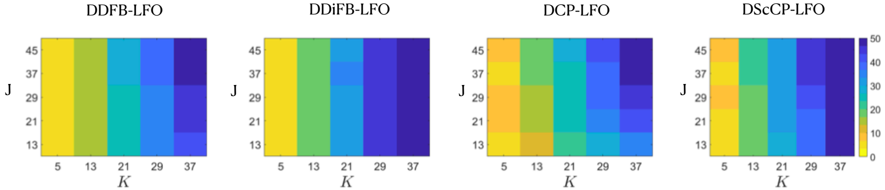
$$\chi \leq \prod_{k=1}^{K} \|W_k\|_S.$$

☛ [Pesquet, Repetti, Terris, Wiaux, 2020]: tighter bound by Lipschitz continuity:

$$\chi \approx \max_{(\mathrm{z}_s)_{s\in\mathbb{I}}} \|\,\mathrm{J}\,\mathrm{d}_\Theta(\mathrm{z}_s)\|_S.$$

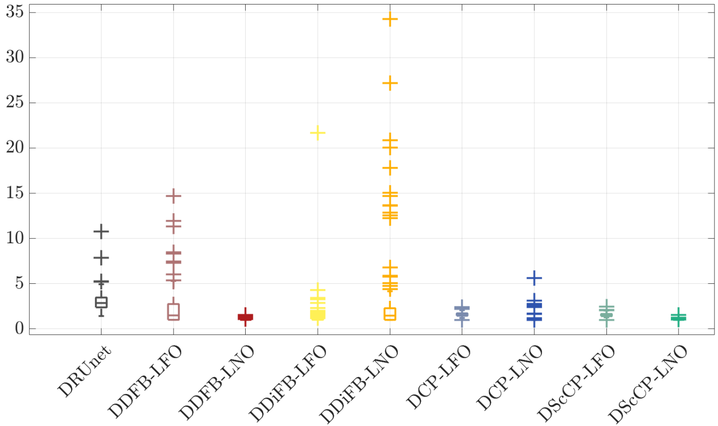
where J denotes the Jacobian operator.

Stability



Upper bound $\prod_{k=1}^{K} \|\mathrm{W}_k\|_S$ for different unfolded neural network configurations.

Stability



Distribution of $(\| \mathrm{J}\, f_\Theta(\mathrm{z}_s)\|_S)_{s\in\mathbb{J}}$ for 100 images extracted from BSDS500 validation dataset $\mathbb{J}$, for the proposed PNNs and DRUnet.

PnP based on unfolded denoiser

Ground truth

Noisy ($\sigma = 0.015$) – 20.11 dB

BM3D – **27.10** dB

DRUnet – 25.09 dB

DDFB-LNO – **27.23** dB

DScCP-LNO – 26.48 dB

Restoration performance.

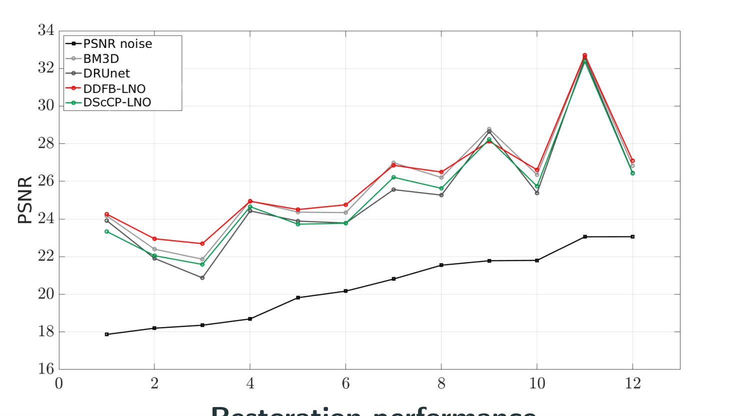
Restoration example for $\sigma = 0.015$, with parameters $\gamma = 1.99$ and β chosen optimally for each scheme.

PnP based on unfolded denoiser

DRUnet – 25.09 dB

DDFB-LNO – 27.23

PnP based on unfolded denoiser



Restoration performance.

Best PSNR values obtained with DDFB-LNO, DScCP-LNO, DRUnet and BM3D, on 12 images from BSDS500 validation set degraded, with $\sigma = 0.03$.

Conclusions for part 1

☛ Unified framework for several proximal unfolded NN schemes.

☛ Faster provides better unfolded strategy in this denoising framework.

☛ Proximal unfolded NN schemes: good compromise between number of parameters and performance.

☛ Proximal unfolded schemes may help to design stable neural networks.

2- Avoid new training with an equivariant strategy

Equivariant network

Equivariant network

- d_θ is **invariant** to a transformation D if the output remains unchanged:

$$\mathrm{d}_\Theta(\mathrm{Dz}) = \mathrm{d}_\Theta(\mathrm{z})$$

- d_θ is **equivariant** to a transformation D if the output changes in a corresponding way:

$$\mathrm{d}_\Theta(\mathrm{Dz}) = \mathrm{D}'\mathrm{d}_\Theta(\mathrm{z})$$

Particular case: If G acts on the same way in the output and input spaces, $\mathrm{d}_\Theta(\mathrm{Dz}) = \mathrm{D}\mathrm{d}_\Theta(\mathrm{z})$

Make d_Θ $\mathcal{G}$-equivariant

- Averaging over a group $\mathcal{G}$ of unitary matrix $\{\mathrm{D}_g\}_{g\in\mathcal{G}}$:

$$\mathrm{d}_{\theta,\mathcal{G}}(\mathrm{z}) = \frac{1}{|\mathcal{G}|}\sum_{g\in\mathcal{G}} \mathrm{D}_g^{-1}\mathrm{d}_\Theta(\mathrm{D}_g\mathrm{z}) \quad \rightarrow \quad \text{can be computationally demanding}$$

Equivariant Plug-and-play

- **FB-PnP**:

$$x^{[k+1]} = d_\Theta(\ x^{[k]} - \gamma A^\top (Ax^{[k]} - z)\ )$$

- **Equivariant FB-PnP** [Terris, Moreau, Pustelnik, Tachella, 2024]

Sample $g_k \in \mathcal{G}$

$$x^{[k+1]} = D_{g_k}^{-1} d_\Theta (D_{g_k}\ x^{[k]} - \gamma A^\top (Ax^{[k]} - z)\ )$$

Equivariant Plug-and-play

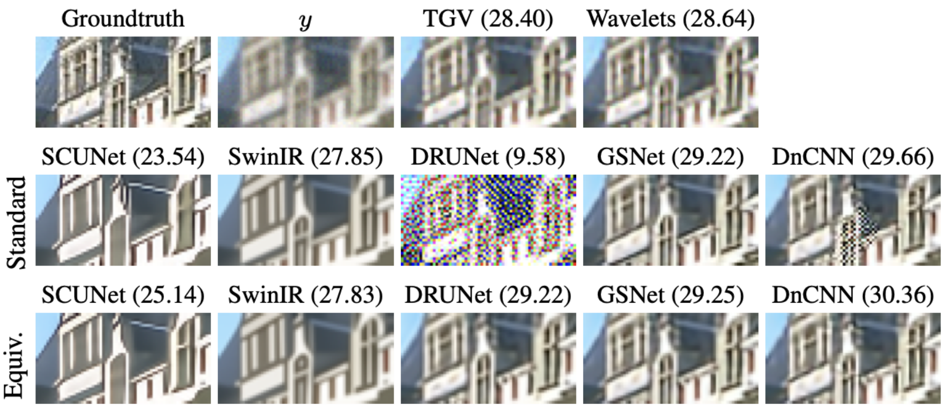
Proposition

$\mathrm{d}_\Theta \mathrm{z} = \mathrm{W}\mathrm{z}$ be a linear denoiser with singular value decomposition $W = \sum_{i=1}^{n} \lambda_i u_i v_i^\top$ and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$. If the principal component $u_1 v_1^\top$ is not $\mathcal{G}$-equivariant, the averaged denoiser $\mathrm{d}_{\Theta,\mathcal{G}}$ has a strictly smaller Lipschitz constant than d_Θ.

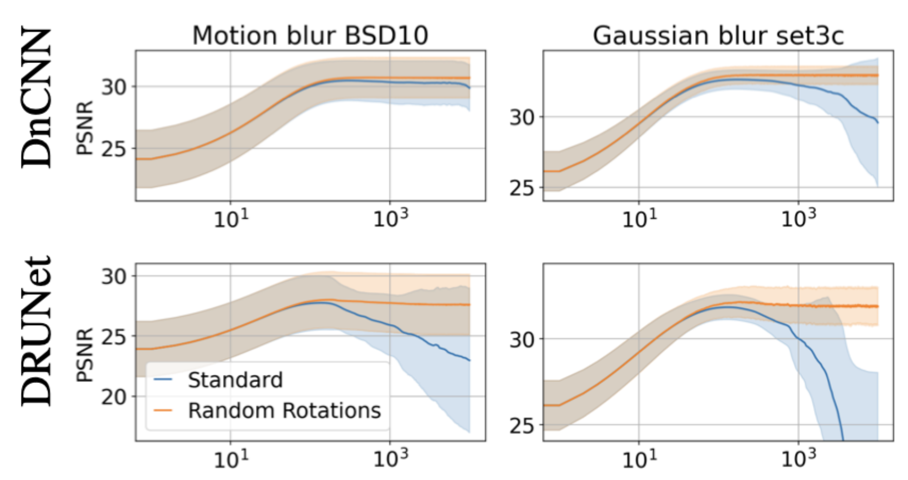
	Lipschitz DnCNN	DnCNN	SCUNet	SwinIR	DRUNet (σ_d=0.01)
Standard d_Θ	1.06	1.44	5.78	6.28	1.57
Equivariant $\mathrm{d}_{\Theta,\mathcal{G}}$	0.92	1.18	4.19	4.05	1.26

Lipschitz constant (i.e., the spectral norm of the Jacobian) of various denoisers averaged over 10 different patches of 64×64 pixels. Equivariant denoisers are obtained by averaging over the group of 90 degree rotations and reflections.

Numerical experiments Gaussian debluring



Numerical experiments debluring



References

- H.T.V. Le, A. Repetti, N. Pustelnik, Unfolded proximal neural networks for robust image Gaussian denoising, IEEE Transactions on Image Processing, 33, pp. 4475 - 4487, Aug. 2024.
- M. Terris, T. Moreau, N. Pustelnik, J. Tachella Equivariant plug-and-play image reconstruction, IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle WA, USA, 2024.
- N. Laurent, J. Tachella, E. Riccietti, N. Pustelnik, Multilevel plug-and-play image reconstruction, submitted, 2024.

Conclusions and perspectives

- Different types of convergence guarantees: convergence of the iterates, convergence rate for f or the sequence.
- Numerous proximal algorithms depending of the applicative context.
- Strong convexity: key tool to accelerate algorithms and prove convergence rates on iterates.
- Some algorithms stay difficult to analyze : FISTA, primal-dual schemes.
- Large scales stay difficult to handle : block, multilevel,...

Conclusions and perspectives

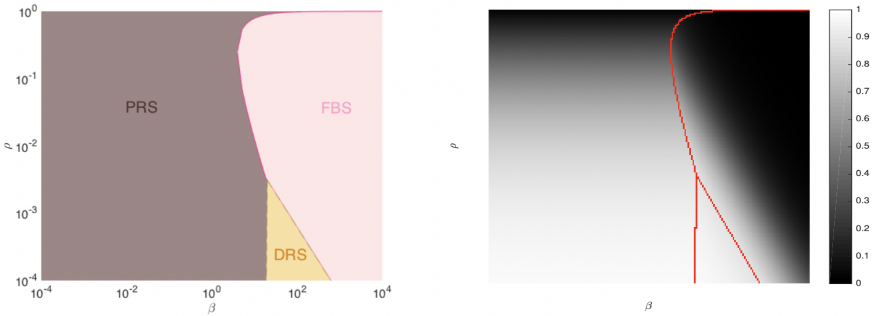


Figure 3: (Left) Regimes where PRS or FBS or DRS achieves a better rate according to Proposition 7 when $\alpha = 1$ as a function of (β, ρ). (Right) Optimal numerical rates and associated regions.

Conclusions and perspectives

- Unfolded = truncated iterations.

- Unfolded is certainly the best compromise between end-to-end NN and variational approaches but theoretical guarantees still lack.

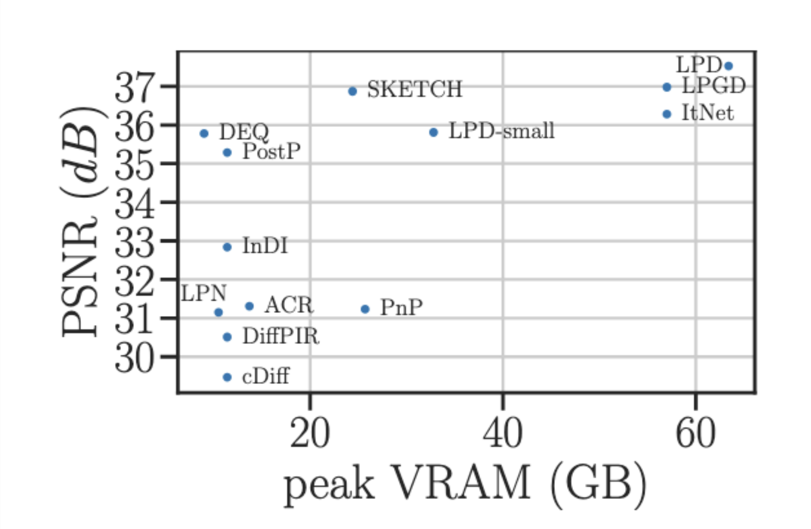
Conclusions and perspectives

Table 2.1: Reconstruction performances on the **LoDoPaB-CT** dataset, with *best* to *medium*, *low* and *worst* results highlighted. We compare state-of-the-art approaches and vary the number of views. ℓ/m is the normalized data-fidelity.

LoDoPaB-CT - **2D** [TEST]	**SSIM** ↑	**PSNR** ↑	$\ell/m \downarrow (.10^5)$	[train] **VRAM** ↓	**time/epoch** (s)↓	[test] **VRAM** ↓	**time/sample** (s)↓
FBP	0.392	24.05	5.930	✗	✗	0.0482	0.0035
TV	0.814	33.03	6.174	✗	✗	0.0492	22.3596
Post-processing	0.793	33.41	6.156	4.5	50	0.1199	0.0159
cDiffusion [T=1000]	0.735	32.46	6.089	4.6	66	0.1339	13.71
InDI [T=100]	0.793	33.33	6.171	4.5	50	0.1312	1.3368
RED [$\sigma = 10.$]	0.759	30.55	6.181	3.5 / 14.8	14 / 272	0.0842	16.0887
PnP-PGD [$\sigma = 10.$]	0.759	30.57	6.181	3.5 / 14.8	14 / 272	0.0842	20.2022
LPN [$\sigma = 10.$]	0.782	32.26	6.169	4.2	120	0.8547	16.9200
DiffPIR	0.747	32.50	6.176	4.6	66	0.1344	19.2262
Adversarial Regularizer	0.778	31.21	6.170	5.6	323	0.2355	20.1532
Adversarial Convex Regularizer	0.786	31.53	6.172	6.0	253	0.1455	52.4104
ItNet	0.845	36.05	6.175	32.2	918	0.1232	0.2632
LPGD	0.844	36.04	6.175	32.2	923	0.1243	0.2573
LPD	0.851	36.39	6.174	150	2910	0.4566	0.5888

@ Extracted from Romain Vo PhD

Conclusions and perspectives



© Extracted from Romain Vo PhD

Conclusions and perspectives

- Unfolded = truncated iterations.
- Unfolded is certainly the best compromise between end-to-end NN and variational approaches but theoretical guarantees still lack.
- Bilevel framework to derive learning schemes.
- Applied to other problems such as edges detection.

Conclusions and perspectives

Clean image

Ground-truth edges

BZ

CE $= 2.30$

BDCN

3.36

DiffusionEdge

3.34

BZ-PNN

0.96

Conclusions and perspectives

- PnP: good solution to avoid training on new dataset.
- Stronger convergence guarantees than unfolded schems but still very sensitive. Equivariant-PnP, Multilevel-PnP are solution to improve it.