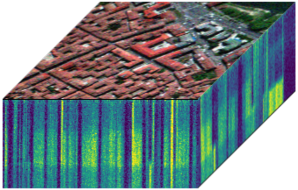
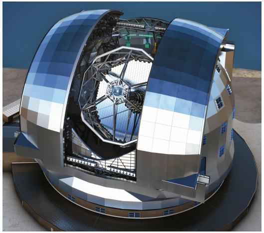
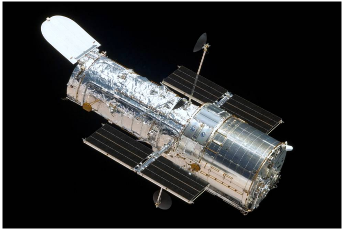
Physical Models and Machine Learning for Scientific Imaging

Julien Mairal

Inria Grenoble

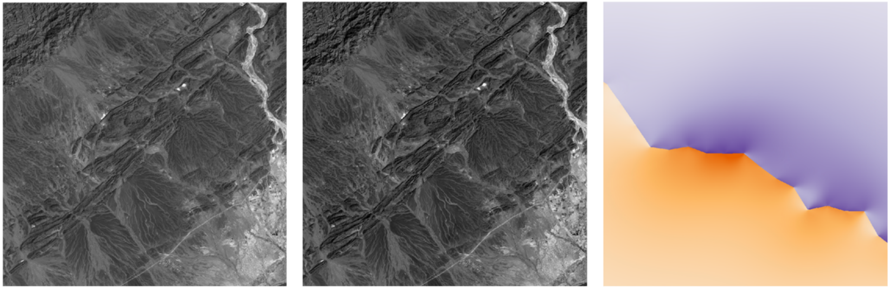
with B. Lecouat, T. Bodrito, T. Eboli, O. Flasseur
J. Ponce, M. Langlois, A.-M. Lagrange

Scientific imaging

- Extracting or recovering information from (raw) data image sources.
- Scientific applications where interpretation is important (astronomy, Earth observation).
- Dealing with various data sensors beyond RGB imaging.

Scientific imaging

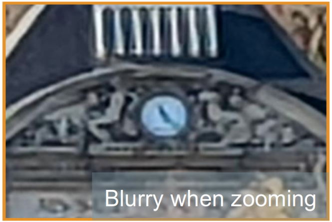


Picture from [Montagnon, Hollingsworth, Pathier, Marchandon, Dalla Mura, Giffard-Roisin, 2022].

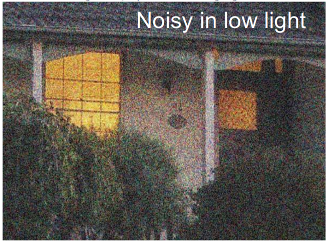
- Example: image alignment for ground deformation estimation.

Scientific imaging

Google Pixel 6, long range tele



Google Pixel 6, crop: limited detail at long range



Small aperture and sensor

- Even for personal photography, dealing with **real-world** degradations is hard.

Classical modeling for image/signal restoration

Find a reasonable model of degradation

For instance: $$\underbrace{y}_{\text{observations}} = A \underbrace{x}_{\text{true signal}} + \underbrace{\varepsilon}_{\text{noise}}.$$

Estimate the true signal by optimizing a reasonable cost function

For instance: $$\min_x \underbrace{\|y - Ax\|^2}_{\text{data fitting term}} + \underbrace{\lambda\phi(x)}_{\text{prior information}}.$$

Some classical priors

- Smoothness $\|\mathcal{L}x\|^2$.
- Total variation $\|\nabla x\|_1$.
- Wavelet sparsity $\|Wx\|_1$.

Use of the ℓ_1-norm in geophysics (Taylor et al., 1979)

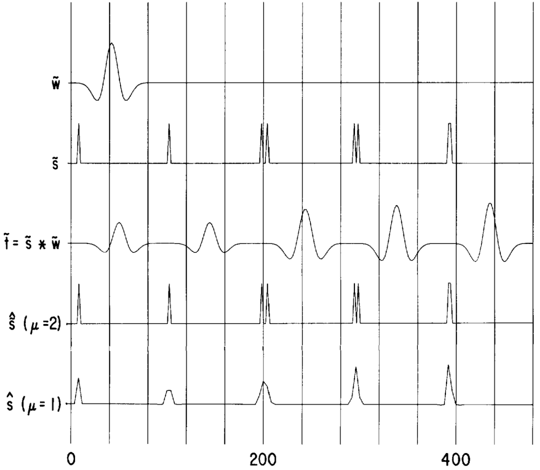


FIG. 8. High precision spike train extraction. A spike train $\tilde{\mathbf{s}}$ and wavelet $\tilde{\mathbf{w}}$ sampled at 2 msec are convolved to give $\tilde{\mathbf{t}} = \tilde{\mathbf{s}} * \tilde{\mathbf{w}}$ which is then resampled at 4 msec. Two spike trains are extracted from the 4 msec trace: $\hat{\tilde{\mathbf{s}}}$ at a sample rate of 2 msec and $\mathbf{s}$ at 4 msec, both with $\lambda = 30$.

Use of the ℓ_1-norm in geophysics (Taylor et al., 1979)

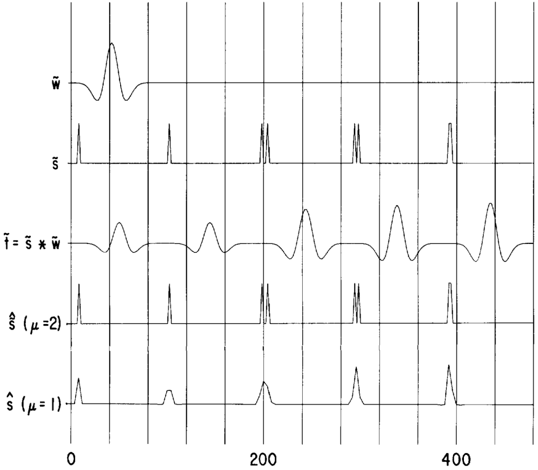
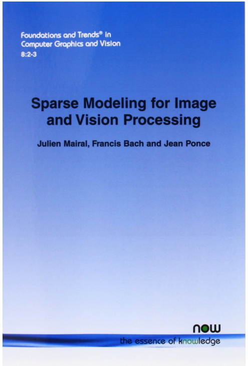


FIG. 8. High precision spike train extraction. A spike train $\tilde{\mathbf{s}}$ and wavelet $\tilde{\mathbf{w}}$ sampled at 2 msec are convolved to give $\tilde{\mathbf{t}} = \tilde{\mathbf{s}} * \tilde{\mathbf{w}}$ which is then resampled at 4 msec. Two spike trains are extracted from the 4 msec trace: $\hat{\tilde{\mathbf{s}}}$ at a sample rate of 2 msec and $\mathbf{s}$ at 4 msec, both with $\lambda = 30$.

Advertisment: More on the ℓ_1-norm in computer vision and image processing



Deep learning for image restoration

- **Engineer a realistic dataset**: Produce enough pairs (x_i, y_i) of clean/degraded images.
- **Choose a class of parametrized models** $\{f_\theta : \theta \in \Theta\}$.
- **Learn the parameters**:

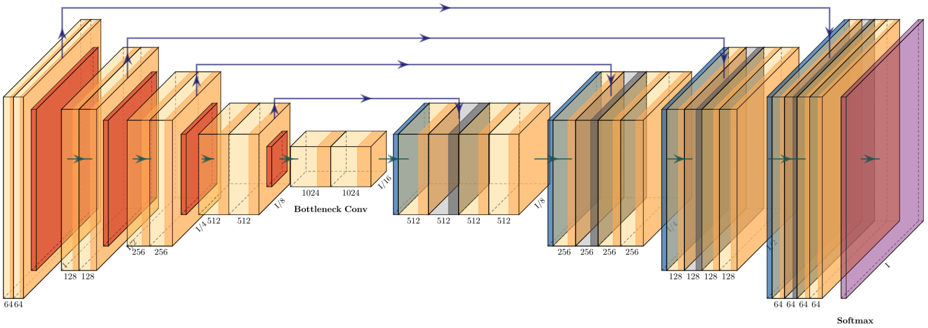
$$\min_\theta \frac{1}{n} \sum_{i=1}^{n} \|f_\theta(y_i) - x_i\|.$$

Deep learning for image restoration

- **Engineer a realistic dataset**: Produce enough pairs (x_i, y_i) of clean/degraded images.
- **Choose a class of parametrized models** $\{f_\theta : \theta \in \Theta\}$.
- **Learn the parameters**:

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \|f_\theta(y_i) - x_i\|.$$

Example: U-Net



Limitations/Drawbacks of both paradigms

"Classical approaches"

- **Advantage:** robustness and interpretability (solves what it is supposed to solve).
- **Limitation:** requires a reasonably well calibrated model.
- **Limitation:** designing a good image prior by hand is very hard.

Limitations/Drawbacks of both paradigms

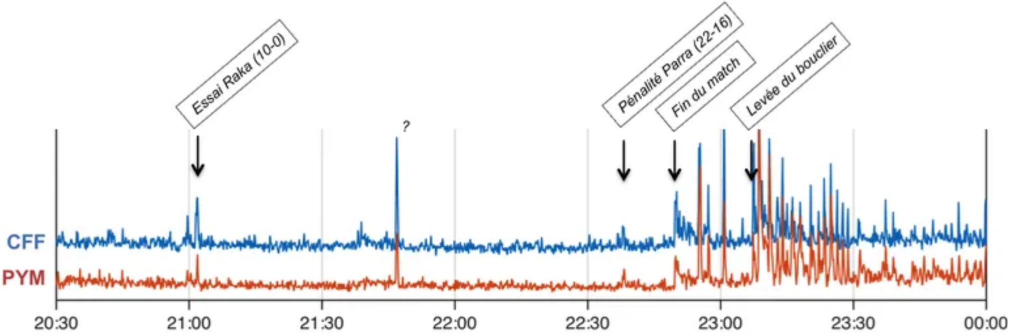
"Classical approaches"

- **Advantage:** robustness and interpretability (solves what it is supposed to solve).
- **Limitation:** requires a reasonably well calibrated model.
- **Limitation:** designing a good image prior by hand is very hard.

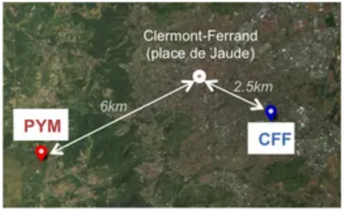
"Deep learning"

- **Advantage:** automatically adapts to the task and data.
- **Limitation**: tuned to a specific data distribution (data engineering is important).
- **Limitation**: very hard to acquire real pairs of clean/degraded images (see next slide).

One difficulty: unknown alignment due to floor vibrations



Energie sismique dans la bande de fréquences 1-3Hz enregistrée aux stations sismologiques du Puy de Manson (PYM) et de Clermont-Ferrand/Cézeaux (CFF) le 04 Juin 2017 lors de la finale du Top 14 de rugby entre l'ASM et le RCT.
Les pics d'énergie sont probablement liés aux mouvements de foule sur la place de Jaude.



@J. Vergne, EOST

What can we do?

Data engineering

- use real images with **realistic synthetic** degradations.
- take into account **the physics** of the sensor/task in the data generation process.

Use hybrid approches!

- combine model-based approaches with **trainable image priors** (plug-and-play, unrolled optimization, deep equilibrium. . .).

What can we do?

Data engineering

- use real images with **realistic synthetic** degradations.
- take into account **the physics** of the sensor/task in the data generation process.

Use hybrid approches!

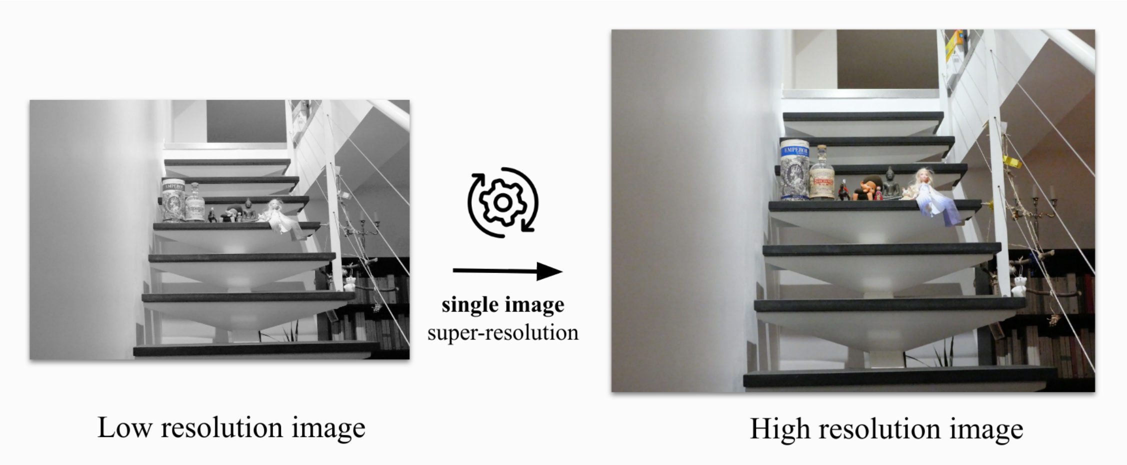
- combine model-based approaches with **trainable image priors** (plug-and-play, unrolled optimization, deep equilibrium. . .).

Let us see how to do that with a first concrete example.

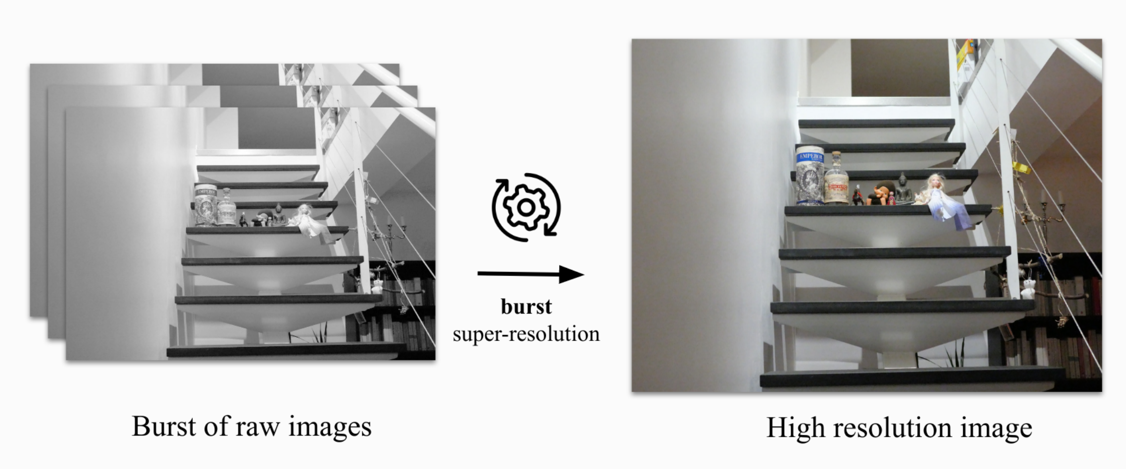
Raw Burst Image Processing

- B. Lecouat, J. Ponce, and J. Mairal. Lucas-Kanade Reloaded: End-to-End Super-resolution from Raw Image Bursts. *ICCV*. 2021.
- B. Lecouat, T. Eboli, J. Ponce, and J. Mairal. High Dynamic Range and Super-Resolution From Raw Image Bursts. *SIGGRAPH*. 2022.

Single-Image Super-Resolution vs...



Super-Resolution from Raw Bursts - Handheld Camera



[Tsai and Huang, 1984], [Farsiu et al., 2004], [Wronski et al., 2019], [Bhat et al., 2021], ...

Picture taken at high ISO with low exposure time

Left: high-quality jpg output of the camera ISP (one frame).
Right: $\times 4$ super-resolution from a burst of 30 raw images (handheld camera).

Picture taken at high ISO with low exposure time

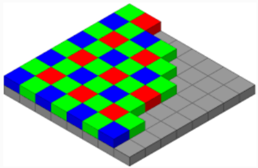
Left: high-quality jpg output of the camera ISP (one frame).
Right: $\times 4$ super-resolution from a burst of 30 raw images (handheld camera).

Picture taken at high ISO with low exposure time

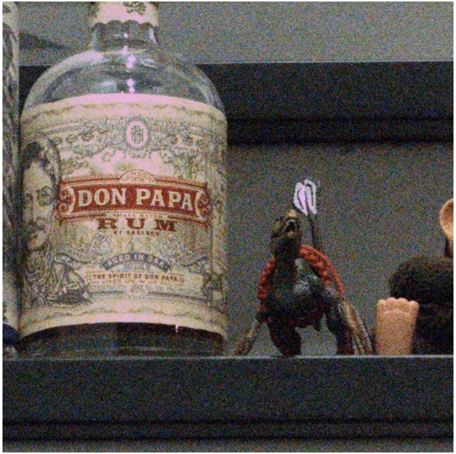
Left: high-quality jpg output of the camera ISP (one frame).
Right: $\times 4$ super-resolution from a burst of 30 raw images (handheld camera).

Challenges

- **Aligning** images with subpixel accuracy (for **super-resolution**).
- Dealing with noisy raw data (**blind denoising**).
- Reconstructing color images from raw data (**demosaicking**).

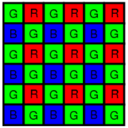


RGB sensor (bayer pattern)



The Camera raw processing pipeline (simplified view)

How does your camera process sensor data?



→ White balance. Black substraction. → **Denoising** ↓ **Demosaicking** → Tone mapping. Gamma correction. →

Working with raw data is important, before the camera ISP produces irremediable damage!

Aliasing is your ally [Vandewalle et al. 2006], [Wronski et al., 2019]

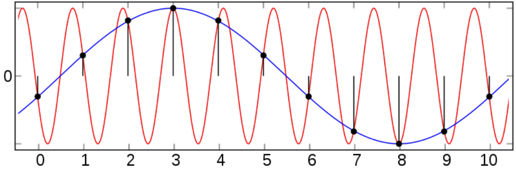


Figure: Example of aliasing: undersampled sinusoid causes confusion with a sinusoid with lower frequency. Picture from Wikipedia.

- Aliasing is usually mitigated with some optical / digital filters.
- **But anti-aliasing removes high frequency measurements**!

The "old" world of classical inverse problems.

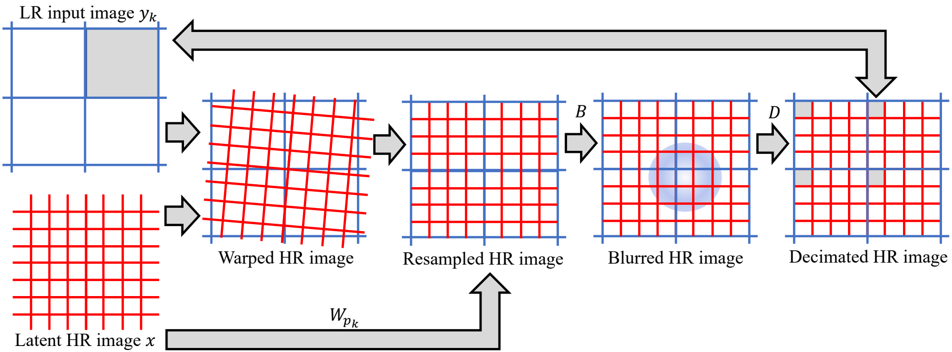


Image formation model for a burst $Y = y_1, \ldots, y_k$

$$y_k = DBW_{p_k}x + \varepsilon_k.$$

The "old" world of classical inverse problems.

Image formation model for a burst $Y = y_1, \ldots, y_k$

$$y_k = DBW_{p_k}x + \varepsilon_k.$$

Inverse problem

$$\min_{x,p_k} \frac{1}{K} \sum_{k=1}^{K} \|y_k - DBW_{p_k}x\|^2 + \lambda\phi(x).$$

A natural strategy

- define an appropriate prior $\phi(x)$ for natural images and optimize!

The "old" world of classical inverse problems.

Image formation model for a burst $Y = y_1, \ldots, y_k$

$$y_k = DBW_{p_k}x + \varepsilon_k.$$

Inverse problem

$$\min_{x,p_k} \frac{1}{K}\sum_{k=1}^{K} \|y_k - DBW_{p_k}x\|^2 + \lambda\phi(x).$$

A natural strategy

- define an appropriate prior $\phi(x)$ for natural images and optimize!
- **What is a good ϕ?**

The deep learning world.

Inverse problem given $Y = y_1, \ldots, y_K$

$$\min_{x, p_k} \frac{1}{K} \sum_{k=1}^{K} \|y_k - DBW_{p_k} x\|^2 + \lambda \phi(x).$$

What a deep learning model would look like?

- Need pairs of (LR bursts Y / SR image x), then solve

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \|f_\theta(Y_i) - x_i\|.$$

- **How to obtain realistic training data**?
- **What is a good** f_θ?

The new world of trainable algorithms.

Inverse problem given $Y = y_1, \ldots, y_K$

$$\min_{x, p_k} \frac{1}{K} \sum_{k=1}^{K} \|y_k - DBW_{p_k} x\|^2 + \lambda \phi_{\theta}(x).$$

How to design a hybrid approach?

- **Idea 1 (unrolled optimization [Gregor and LeCun, 2010]):** For a given burst Y_i, call $f_\theta(Y_i)$ the output of the optimization procedure solving the inverse problem.

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \|f_\theta(Y_i) - x_i\|.$$

- This allows learning the parameters θ through **differentiable programming**.

The new world of trainable algorithms.

Inverse problem given $Y = y_1, \ldots, y_K$

$$\min_{x,p_k} \frac{1}{K} \sum_{k=1}^{K} \|y_k - DBW_{p_k} x\|^2 + \lambda \phi_{\theta}(x).$$

How to design a hybrid approach?

- **Idea 1 (unrolled optimization [Gregor and LeCun, 2010]):** For a given burst Y_i, call $f_\theta(Y_i)$ the output of the optimization procedure solving the inverse problem.

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \|f_\theta(Y_i) - x_i\|.$$

- This allows learning the parameters θ through **differentiable programming**.
- **What is a good $\phi_\theta(x)$**?

The new world of trainable algorithms.

Inverse problem given $Y = y_1, \dots, y_K$

$$\min_{x, p_k} \frac{1}{K} \sum_{k=1}^{K} \|y_k - DBW_{p_k} x\|^2 + \lambda \phi_\theta(x).$$

How to design a hybrid approach?

- **Sketch of Idea 2 (plug and play [Venkatakrishnan et al., 2013]):** In the optimization procedure, the effect of ϕ_θ is to denoise the estimate x.
- Replace the "denoising" steps involving ϕ_θ by a classical neural network parametrized by θ (U-net). Then, no need to define explicitly ϕ_θ.

The new world of trainable algorithms.

Inverse problem given $Y = y_1, \ldots, y_K$

$$\min_{x, p_k} \frac{1}{K} \sum_{k=1}^{K} \|y_k - DBW_{p_k} x\|^2 + \lambda \phi_\theta(x).$$

How to design a hybrid approach?

- **Sketch of Idea 2 (plug and play [Venkatakrishnan et al., 2013]):** In the optimization procedure, the effect of ϕ_θ is to denoise the estimate x.
- Replace the "denoising" steps involving ϕ_θ by a classical neural network parametrized by θ (U-net). Then, no need to define explicitly ϕ_θ.
- **How do we generate realistic training data?**

The new world of trainable algorithms.

Inverse problem given $Y = y_1, \ldots, y_K$

$$\min_{x, p_k} \frac{1}{K} \sum_{k=1}^{K} \|y_k - DBW_{p_k} x\|^2 + \lambda \phi_\theta(x).$$

How to design a hybrid approach?

- **Data engineering [Bhat et al., 2021]:** Consider an database of high-quality RGB images x_i and generate low resolution bursts Y_i with synthetic degradations (random motion, bayer pattern, noise).
- Finally, you are in shape to learn your model

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \|f_\theta(Y_i) - x_i\|.$$

Another Problem: Limited Range

Another Problem: Ghosts

Figure: Misalignements artefacts due to moving objects in the scene. Our implementation did not handle fast moving objects and then generated visual artefacts.

Solution: More Accurate Modeling

Inverse problem given $y_1, \ldots, y_K$

$$\min_{x, p_k} \frac{1}{K} \sum_{k=1}^{K} \|w_k \circ (y_k - DBW_{p_k} x)\|^2 + \lambda \phi_\theta(x),$$

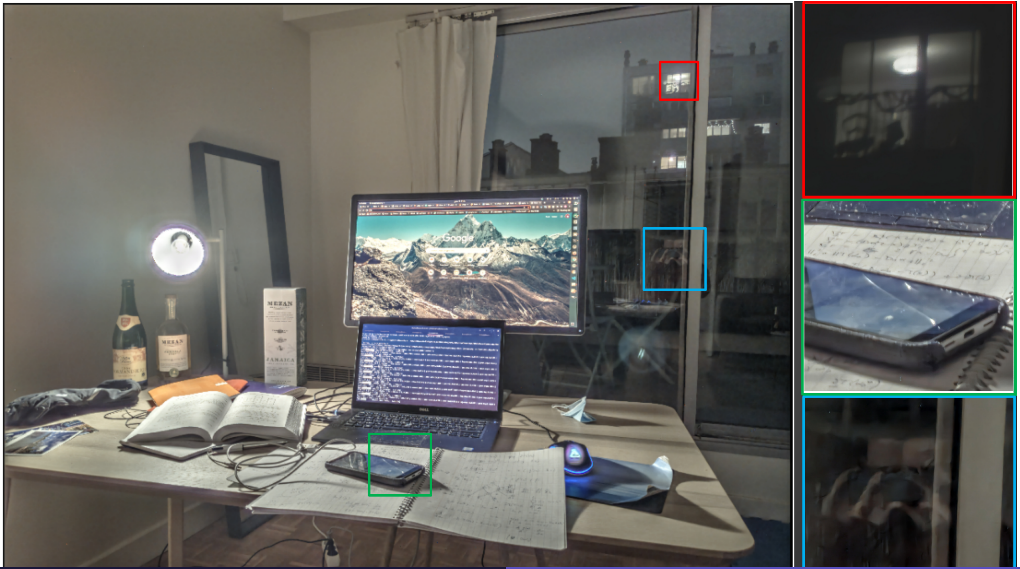
with

$$w_k = \frac{\Delta t_k m(y_k)}{\sum_{j=1}^{K} \Delta t_j m(y_j)} \circ g\left(y_k, W_k y_1\right),$$

- Δt_j: Duration of exposition for frame j;
- $m(y_j)$: Binary mask for saturated pixels;
- $g(y_k, W_k y_1)$: is frame y_k well aligned with y_1? (weight for each pixel).

Result with Bracketing

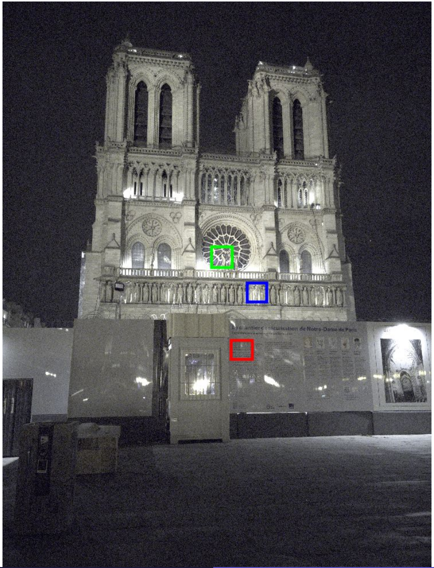
Result with Bracketing



The method now works with dynamic scenes!

Low resolution | [Bhat et al. 2021a] | [Lecouat et al. 2021] | [Luo et al. 2021] | Ours

Joint (blind) denoising, demosaicking, super-resolution and HDR.

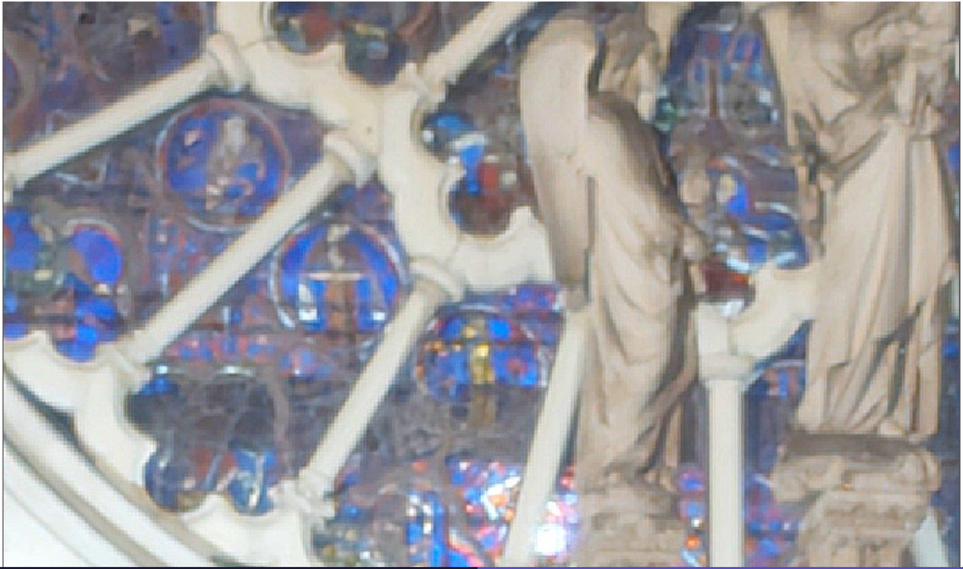


Extension to High-Dynamic Range Imaging (HDR)

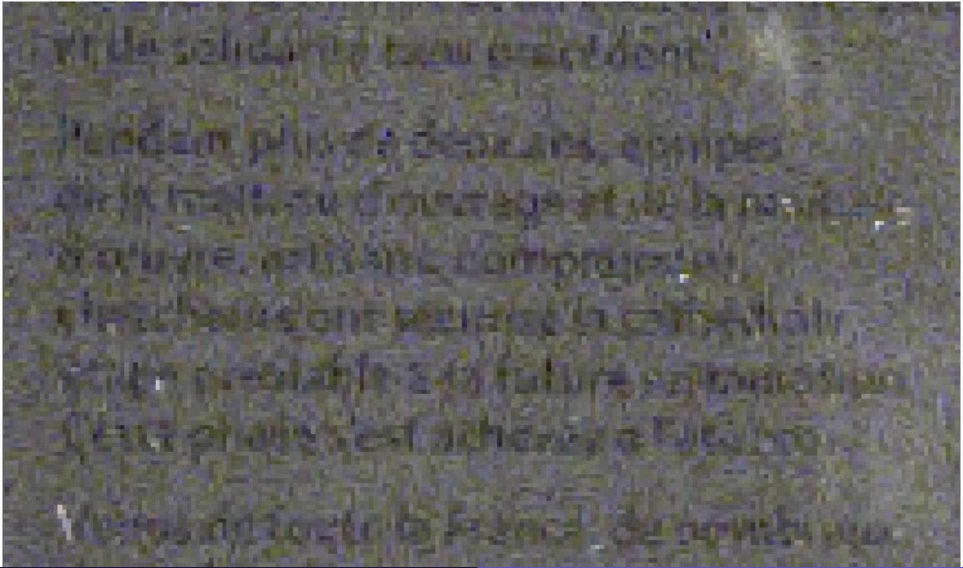
Extension to High-Dynamic Range Imaging (HDR)

Extension to High-Dynamic Range Imaging (HDR)

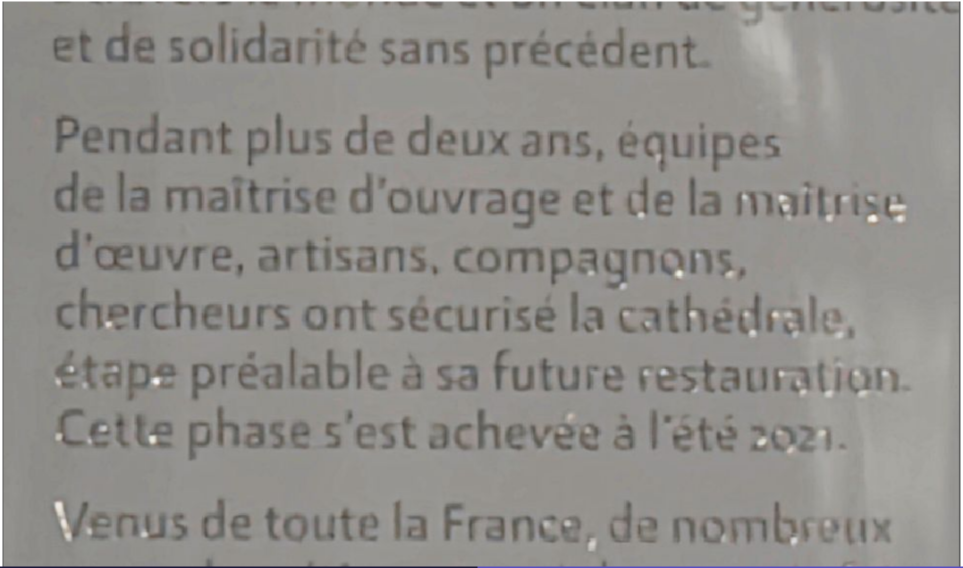
Extension to High-Dynamic Range Imaging (HDR)



Extension to High-Dynamic Range Imaging (HDR)



Extension to High-Dynamic Range Imaging (HDR)



Extension to High-Dynamic Range Imaging (HDR)

Multiframe super resolution: prior work

and, among many others:

- **interpolation-based methods**: [Hardie, 2007], [Takeda et al., 2007];
- **iterative approaches**: [Irani and Peleg, 1991], [Elad and Feuer, 1997],[Farsiu et al., 2004];
- **(deep) learning-based approaches**: [Bhat et al., 2021], [Molini et al., 2019], [Deudon et al., 2019], [Luo et al., 2021];
- and also the literature on video super-resolution (typically not dealing with raw data).

Perspectives for Scientific Imaging

We develop trainable algorithm that encode prior knowledge about the problem. The goal is to recover true signals and not hallucinate details.

Scientific applications

- astronomical images and microscopy.
- software-based adaptive optics.
- remote sensing.

Technological challenges

- data fusion from heterogeneous sensors.
- focus stacking.
- depth estimation and 3D reconstruction (ongoing).

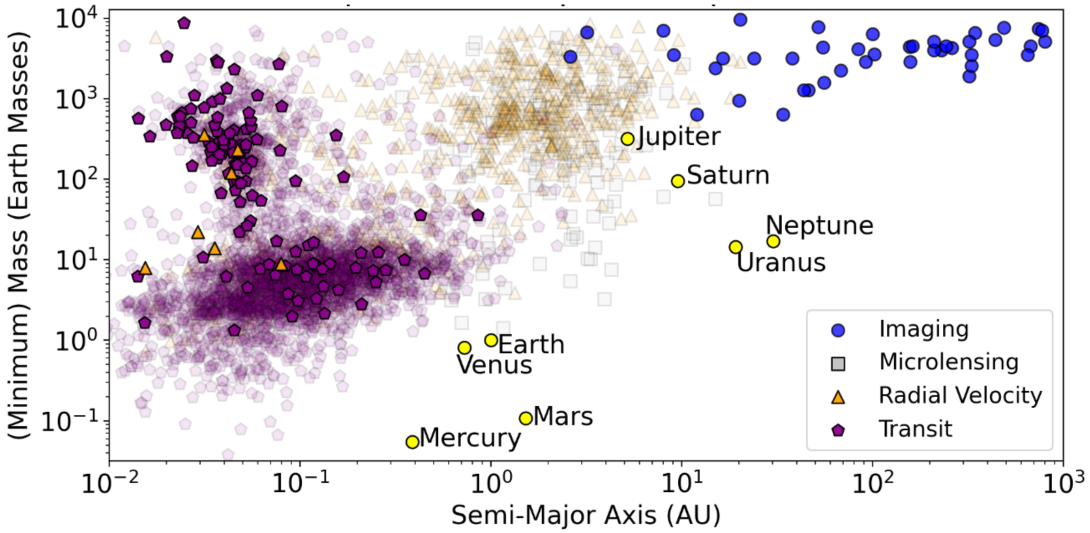
Machine Learning for Astronomy

Very short version

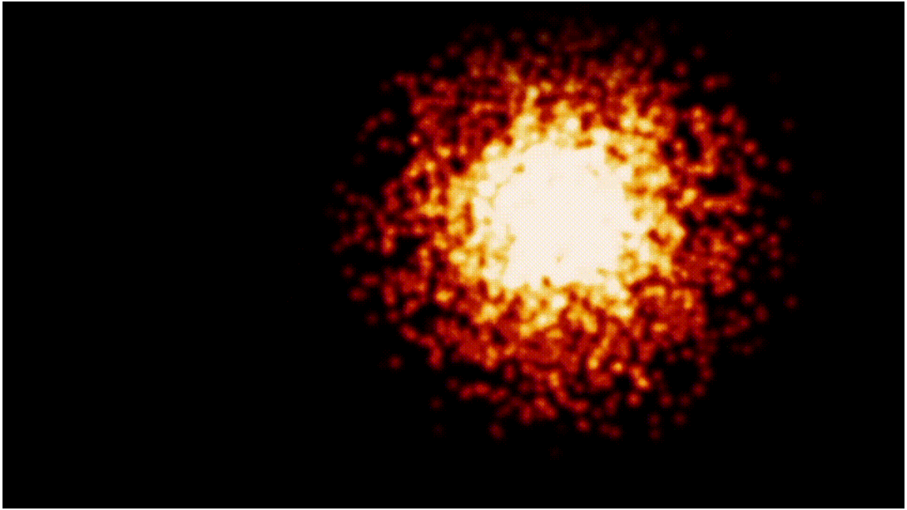
- O. Flasseur, T. Bodrito, J. Mairal, J. Ponce, M. Langlois and A.-M. Lagrange. Deep PACO: Combining Statistical Models with Deep Learning for Exoplanet Detection and Characterization in Direct Imaging at High Contrast. to appear in Monthly Notices of the Royal Astronomical Society (MNRAS). 2023.

Slides courtesy of Théo Bodrito

Exoplanet detection: current progress

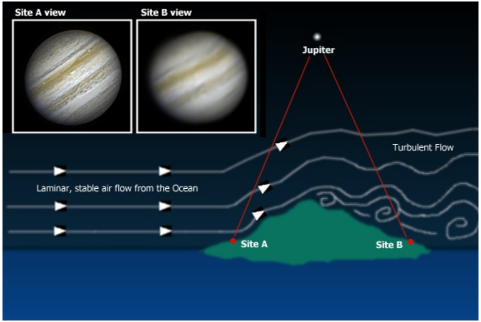


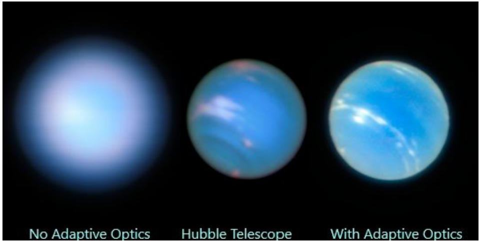
Challenge: contrast $10^6 \sim 10^5$



A coronagraph blocks light emitted by the star. Image credit: NASA

Challenge: atmospheric disturbances - adaptive optics





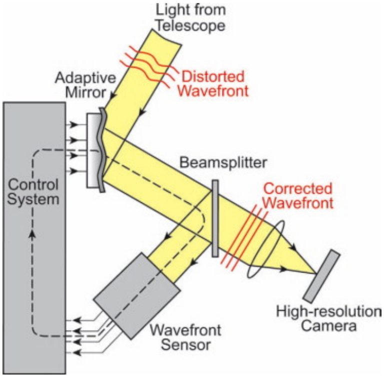


Image credit: Damian Peach and ESO

Angular Differential Imaging (ADI)

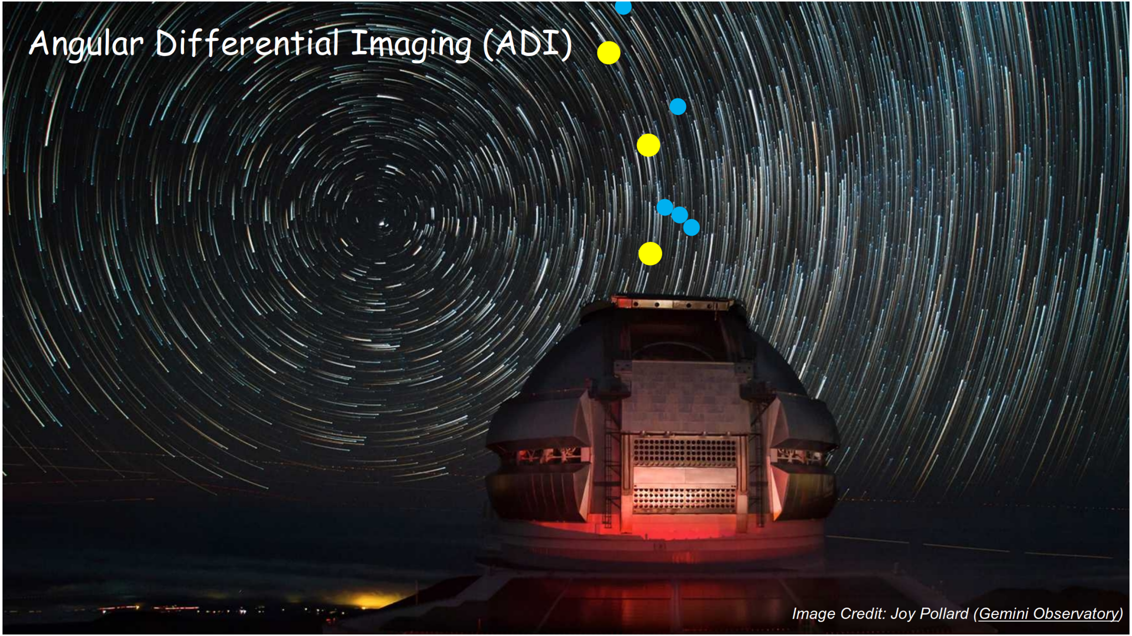
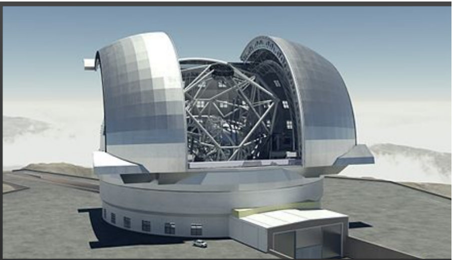


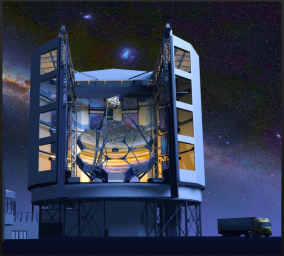
Image Credit: Joy Pollard (Gemini Observatory)

Direct imaging: ongoing ground-based projects

Very Large Telescope 1999
(50m2)

Extremely Large Telescope 2027
(978m2)

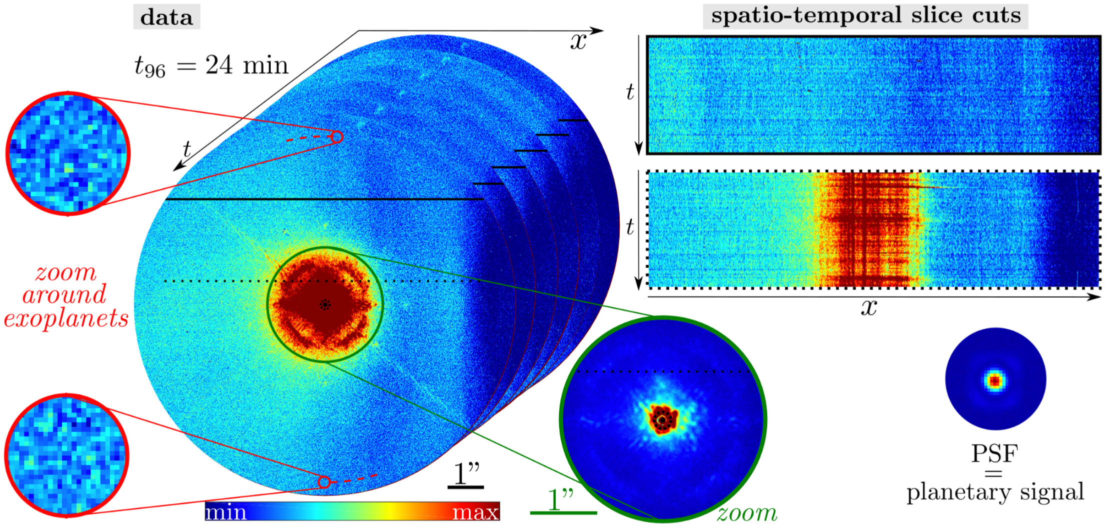




Thirty Meter Telescope 2027
(655m2)

Giant Magellan Telescope 2029
(368m2)

Image data in practice



Speckles are temporally quasi-static but spatially non-stationary. Image credit: Olivier Flasseur

How to train a discriminative machine learning model?

Problem: very few positive samples

- use real data and inject synthetic sources.
- we have very good models of planetary PSF.
- shuffling trick during training to “remove” unknown sources.

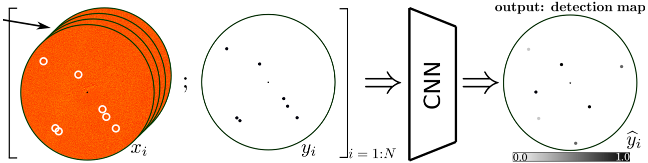
How to train a discriminative machine learning model?

Problem: very few positive samples

- use real data and inject synthetic sources.
- we have very good models of planetary PSF.
- shuffling trick during training to “remove” unknown sources.

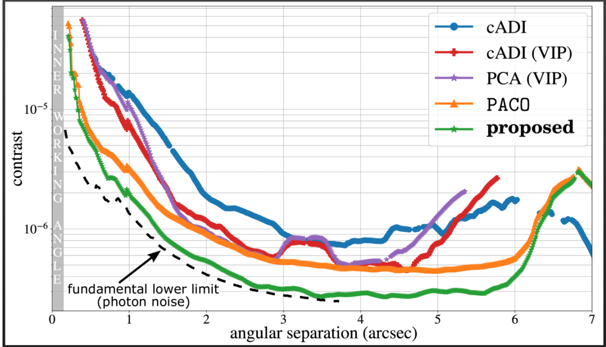
A glimpse at the deep PACO algorithm (Flasseur et al., 2023)

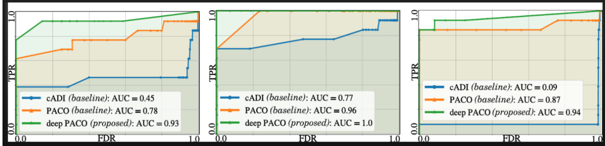
- Whitening: remove speckles structure using PACO (Flasseur et al., 2020).
- Derotation: align the sources.
- Supervised learning (U-net) with Dice loss.



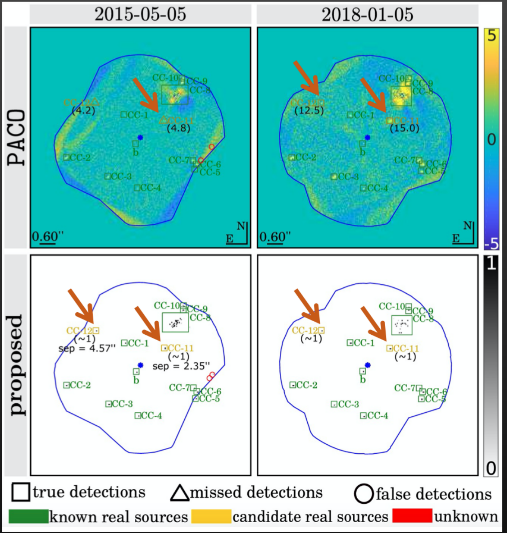
Deep PACO results

Synthetic data benchmarks





Real data: (HD 95086) One new source !



Perspectives and on-going work

Extensions

- Multi-spectral extension (done).
- Flux estimation (done).
- Better uncertainty estimation (desired).
- Exploitation of huge databases of past observations (upcoming).

Appendix

The "old" world of classical inverse problems.

Simple relaxation with "half quadratic splitting" + block coordinate descent

$$\min_{x,z,p_k} \frac{1}{K} \sum_{k=1}^{K} \|y_k - U_{p_k} z\|^2 + \frac{\mu_t}{2} \|z - x\|^2 + \lambda \phi_\theta(x).$$

- minimizing with respect to p_k (parameters of an affine transformation) is performed by Gauss-Newton steps. This is the algorithm of **Lucas and Kanade [1981]**.
- minimizing with respect to x requires computing the **proximal operator** of ϕ_θ.
- minimizing w.r.t. z can be done by gradient descent steps.

Bridging the two worlds with trainable algorithms.

Idea 1: plug-and-play priors [Venkatakrishnan et al., 2013]

Replace proximal operator

$$\arg\min_{x} \frac{1}{2}\|z - x\|^2 + \lambda \phi_\theta(x),$$

by a convolutional neural network $f_\theta(z)$.

Bridging the two worlds with trainable algorithms.

Idea 1: plug-and-play priors [Venkatakrishnan et al., 2013]

Replace proximal operator

$$\arg\min_x \frac{1}{2}\|z - x\|^2 + \lambda\phi_\theta(x),$$

by a convolutional neural network $f_\theta(z)$.

Idea 2: bi-level optimization

Given a dataset of training pairs $(x_i, Y_i)_{i=1,\ldots,n}$, consider

$$\min_\theta \frac{1}{n}\sum_{i=1}^{n} \|\hat{x}_\theta(Y_i) - x_i\|_1$$

such that $\quad \hat{x}_\theta(Y) \in \arg\min_x \min_{p_k} \frac{1}{K}\sum_{k=1}^{K} \|y_k - DBW_{p_k}x\|^2 + \lambda\phi_\theta(x).$

Bridging the two worlds with trainable algorithms.

Idea 1: plug-and-play priors [Venkatakrishnan et al., 2013]

Replace proximal operator

$$\arg\min_x \frac{1}{2}\|z - x\|^2 + \lambda\phi_\theta(x),$$

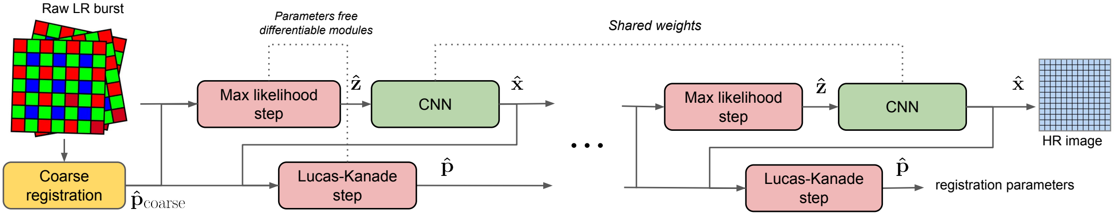
by a convolutional neural network $f_\theta(z)$.

Idea 3: unrolled optimization [Gregor and LeCun, 2010]

- Consider the previous optimization procedure with T steps, producing an estimate $\hat{x}_{\theta,T}(Y)$, given a burst $Y = y_1, \ldots, y_K$.
- Given a dataset of training pairs $(x_i, Y_i)_{i=1,\ldots,n}$, minimize

$$\min_\theta \frac{1}{n}\sum_{i=1}^{n} \|\hat{x}_{\theta,T}(Y_i) - x_i\|_1.$$

Schematic view of our method.



- we keep the interpretability of the classical inverse problem formulation.
- we benefit from a data-driven image prior.

References I

Goutam Bhat, Martin Danelljan, Luc Van Gool, and Radu Timofte. Deep burst super-resolution. *arXiv preprint arXiv:2101.10997*, 2021.

Michel Deudon, Alfredo Kalaitzis, Md Rifat Arefin, Israel Goytom, Zhichao Lin, Kris Sankaran, Vincent Michalski, Samira E Kahou, Julien Cornebise, and Yoshua Bengio. Highres-net: Multi-frame super-resolution by recursive fusion. 2019.

Michael Elad and Arie Feuer. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE transactions on image processing*, 6(12): 1646–1658, 1997.

Sina Farsiu, M Dirk Robinson, Michael Elad, and Peyman Milanfar. Fast and robust multiframe super resolution. *IEEE transactions on image processing*, 13(10):1327–1344, 2004.

Karol Gregor and Yann LeCun. Learning fast approximations of sparse coding. In *Proc. International Conference on Machine Learning (ICML)*, 2010.

Russell Hardie. A fast image super-resolution algorithm using an adaptive wiener filter. *IEEE Transactions on Image Processing*, 16(12):2953–2964, 2007.

References II

Michal Irani and Shmuel Peleg. Improving resolution by image registration. *CVGIP: Graphical models and image processing*, 53(3):231–239, 1991.

Bruce D Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of Imaging Understanding Workshop*, 1981.

Andrea Bordone Molini, Diego Valsesia, Giulia Fracastoro, and Enrico Magli. Deepsum: Deep neural network for super-resolution of unregistered multitemporal images. *IEEE Transactions on Geoscience and Remote Sensing*, 58(5):3644–3656, 2019.

Hiroyuki Takeda, Sina Farsiu, and Peyman Milanfar. Kernel regression for image processing and reconstruction. *IEEE Transactions on image processing*, 16(2):349–366, 2007.

Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *IEEE Global Conference on Signal and Information Processing*, pages 945–948. IEEE, 2013.