Image Enhancement from Raw Image Bursts
Some Opportunities for Scientific Imaging

Julien Mairal
Inria Grenoble
The Thoth team at Inria: Who are we?

Team members
- 5 permanent researchers
- 2 post-docs
- 3 engineers
- 17 PhD students

Research topics
- representation learning in images and videos (self-supervised, incremental, few-shot...)
- image processing and scientific imaging (remote sensing, astrophysics)
- online learning and causality (bandits, counterfactual reasoning)
- large-scale optimization (stochastic, federated, distributed)
- high-dimensional sampling
- representation learning in graph data (molecules, crystallography).
The Thoth team at Inria: Highlight 1

**Self-supervised learning**: SwAV, DINO (Caron et al., NeurIPS 2020, ICCV 2021).
The Thoth team at Inria: Highlight 1 (joint with Meta)

**Self-supervised learning**: SwAV, DINO (Caron et al., NeurIPS 2020, ICCV 2021).
The Thoth team at Inria: Highlight 2 (joint with Criteo)

**Counterfactual learning**: (Zenati et al., ICML 2023).

- Contexts $x_i$ from $\mathcal{X}$ are generated from some data distribution (patients).
- Given a context $x_i$, a stochastic policy $\pi_0$ generates an action $a_i$ (drug dose).
- Given a pair of action/context $(x_i, a_i)$, we observe a loss $y_i$ (dead or alive).

Counterfactual risk minimization consists of optimizing $\pi$ given logged data:

$$L(\pi) = \mathbb{E}_{x} \left[ \mathbb{E}_{a \sim \pi(\cdot|x)} \left[ \mathbb{E}_{y|x,a} \left[ y \frac{\pi(a|x)}{\pi_0(a|x)} \right] \right] \right] + \lambda \Omega(\pi).$$

**Questions**

- which estimator? which regularizer $\Omega$? which parametrization of $\pi$?
- how to exploit multiple deployments (this paper)?
The Thoth team at Inria: Highlight 3

**Transformer models for graphs** (Mialon et al., 2021, Menegaux et al., 2023).

![Diagram of Transformer models for graphs](image)
Revenons à nos moutons
Collaborators

Bruno Lecouat  Jean Ponce  Thomas Eboli

Single-Image Super-Resolution vs...

Low resolution image

High resolution image
Super-Resolution from Raw Bursts - Handheld Camera

Burst of raw images

burst super-resolution

High resolution image

[Tsai and Huang, 1984], [Farsiu et al., 2004], [Wronskski et al., 2019], [Bhat et al., 2021], . . .
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**).
The Camera raw processing pipeline (simplified view)

How does your camera process sensor data?

- White balance.
- Black subtraction.
- Denoising
- Conversion to sRGB.
- 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!
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).

**Interesting for us**: semi-synthetic raw datasets from Bhat et al. [2021].
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).

**Interesting for us**: semi-synthetic raw datasets from Bhat et al. [2021].

**Our solution**: embedding physical image formation model in a trainable architecture
The “old” world of classical inverse problems.

Image formation model

\[ y_k = DBW_{p_k}x + \varepsilon_k. \]
The “old” world of classical inverse problems.

Image formation model

\[ y_k = DBW_{p_k} x + \varepsilon_k. \]

Inverse problem given \( y_1, \ldots, y_K \)

\[
\min_{x, p_k} \frac{1}{K} \sum_{k=1}^{K} \| y_k - DBW_{p_k} x \|_2^2 + \lambda \phi_\theta(x).
\]

A natural strategy

- define an appropriate prior \( \phi_\theta(x) \) for natural images and optimize!
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 \( \phi_\theta \).
- 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 \((x_i, Y_i)_{i=1,...,n}\), consider

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

such that \(\hat{x}_\theta(Y) \in \arg \min_x \min_{p_k} \frac{1}{K} \sum_{k=1}^{K} \|y_k - DBW_{p_k} x\|^2 + \lambda \phi_\theta(x)\).
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 $(x_i, Y_i)_{i=1,...,n}$, 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.
Another Problem: Limited Range
Another Problem: Ghosts

![Image: Misalignments artefacts due to moving objects in the scene. Our implementation did not handle fast moving objects and then generated visual artefacts.](image)

**Figure:** Misalignments artefacts due to moving objects in the scene. Our implementation did not handle fast moving objects and then generated visual artefacts.
Result with Bracketing
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

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

- \( \Delta t_j \): 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).
The method now works with dynamic scenes!

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

Pendant plus de deux ans, équipes de la maîtrise d’ouvrage et de la maîtrise d’œuvre, artisans, compagnons, chercheurs ont sécurisé la cathédrale, étape préalable à sa future restauration. Cette phase s’est achevée à l’été 2021. Venus de toute la France, de nombreux...
Extension to High-Dynamic Range Imaging (HDR)
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).


References II


