Computational Imaging Project

Bibliographic restitution

Sparsity and patch for image restoration.

Authors:

Mohamed El Baha
Emmanuelle Bodji
Hervé Silue
Paul Coste
Quentin Ribaud

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Outline

I - Sparsity Theory

II - Paper’s contribution

III - Paper’s results

Introduction

Hyperspectral imaging contains more information than RGB imaging. It is used in many domains.

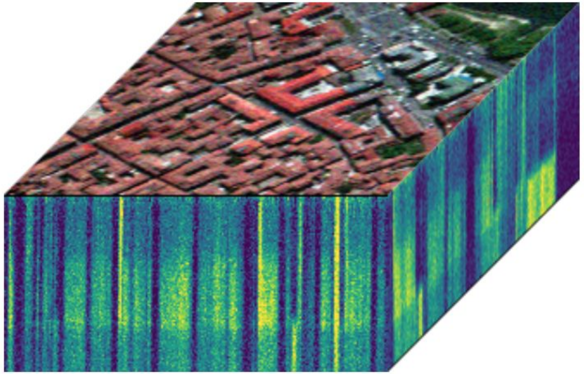


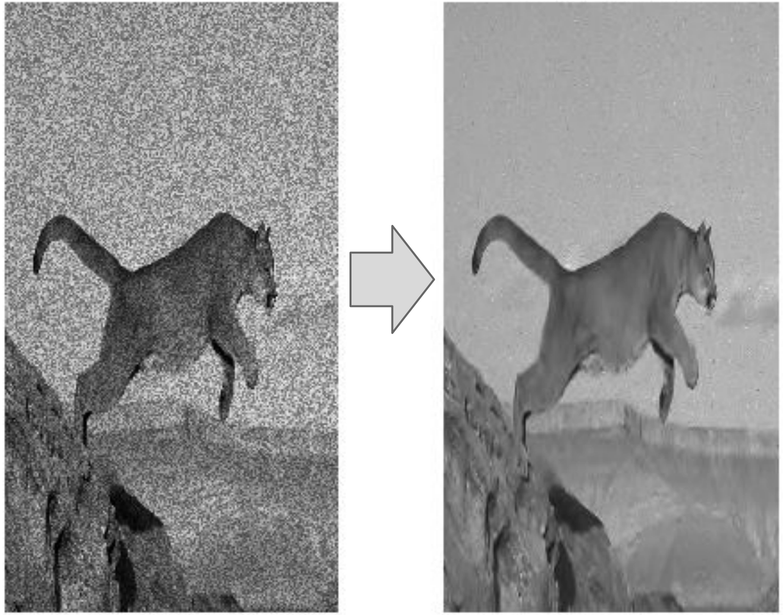
Image denoising could be difficult because of

- complex data
- lack of large-scale collection
- several types of sensors, which typically admit different numbers of spectral bands

Introduction

In our case, we want to denoise hyperspectral imaging. It is an inverse problem.

- Inverse problem: start from an observation and we try to reconstruct what we had at the beginning
- Denoising : find x the clean image from y the noisy image which is the observation



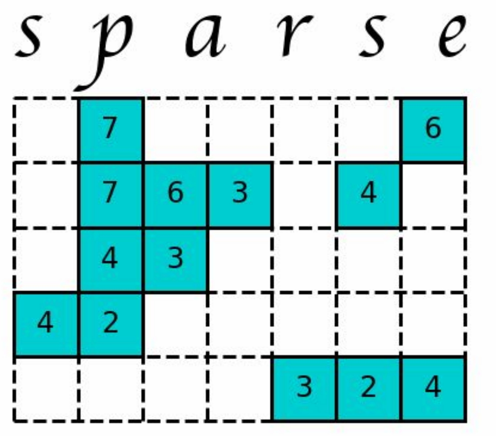
noisy image   image

I - Sparsity theory

What is Sparsity?

- Sparse means something that is in small number and spreaded out over an area (or distribution)
  - For example, sparse vectors and matrix are mostly composed of zeros and only a few non-zero valued elements

c1	c2	c3	c4	c5
0	0	0	5	0
0	0	0	0	0
0	0	1	0	0
0	0	0	0	0
3	0	0	0	0
0	0	0	0	0

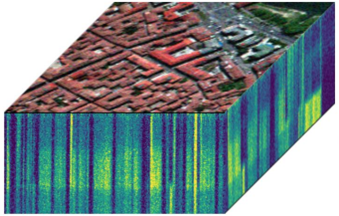


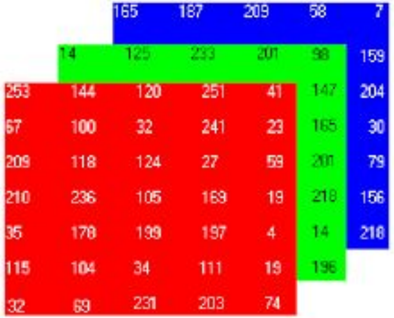
- Sparse vectors are lighter → faster computation

I - Sparsity theory

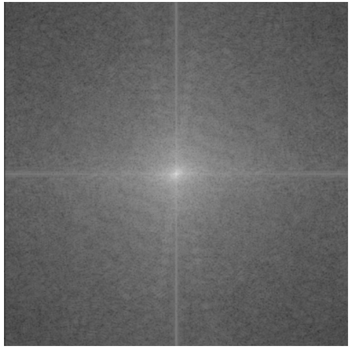
Image Representation

- ❏ 3D-Matrix (RGB, hyperspectral, …)





- ❏ Fourier coefficients (Lab 1)

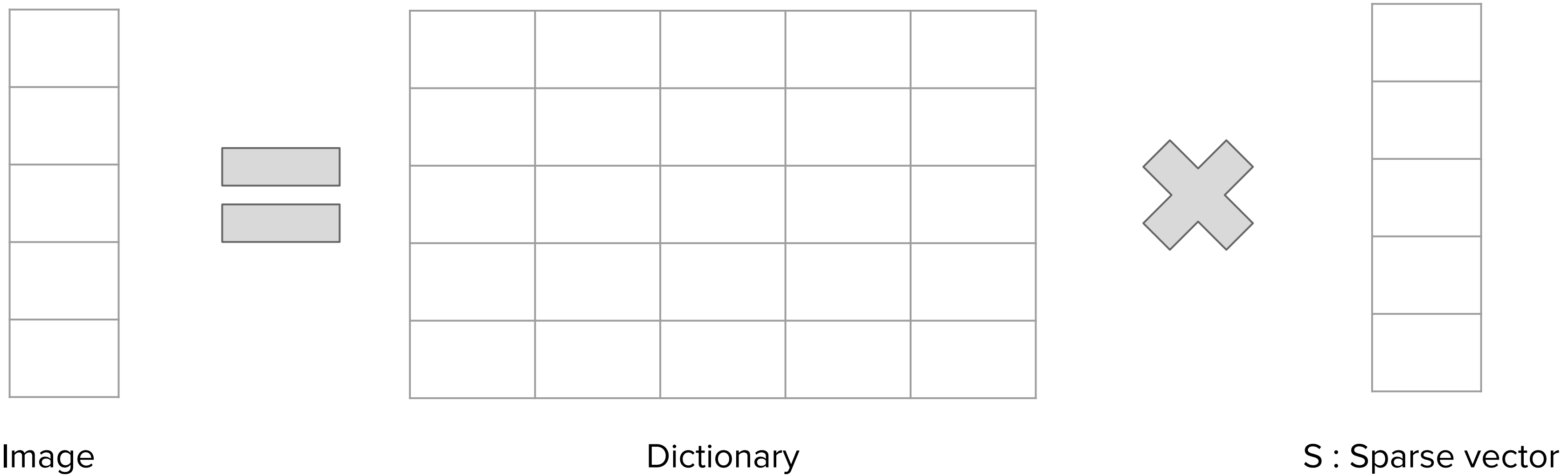


- ❏ Dictionaries

I - Sparsity theory

Dictionaries

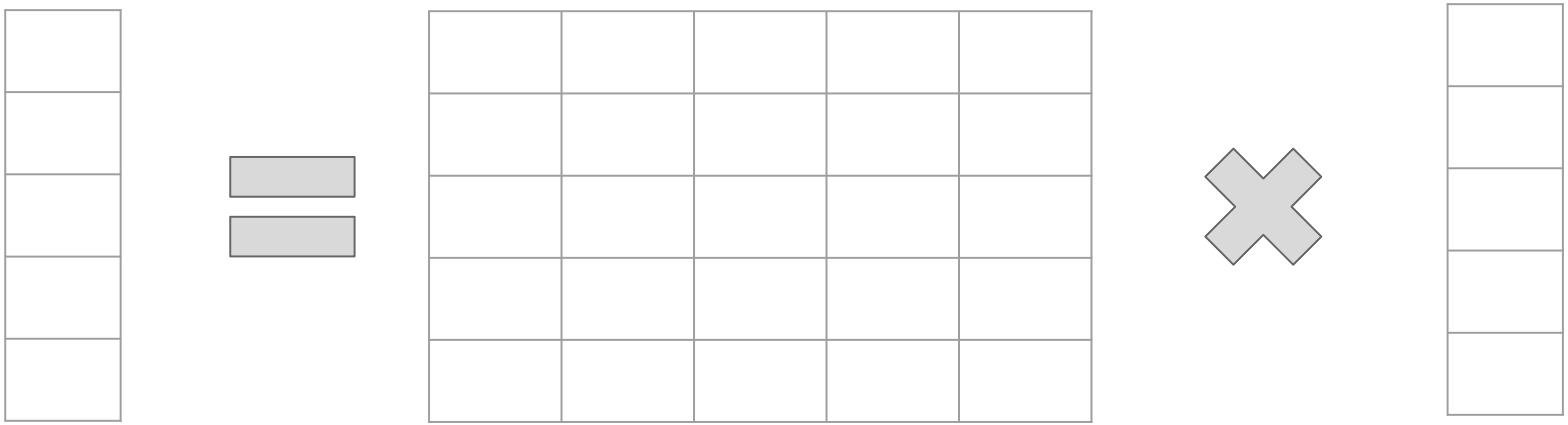
For instance, we can reconstruct an image from a dictionary D.
For a patch, the following calculation is made



- ❏ S : sparse vector for deciding which vector is used from D.
- ❏ D could be a universal basis or Tailored basis

I - Sparsity theory

How to use sparse vectors?



Objective : minimize the difference between the base image and the noisy one

$$\min_{\boldsymbol{\alpha}_i \in \mathbb{R}^p} \frac{1}{2} \underbrace{\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}_i\|^2}_{\text{Observation coefficient}} + \underbrace{\lambda \|\boldsymbol{\alpha}_i\|_1}_{\text{Regularization coefficient}}$$

The Regularization term allow us to not face an ill-posed problem

<u>Why are we using the L1 norm?</u>

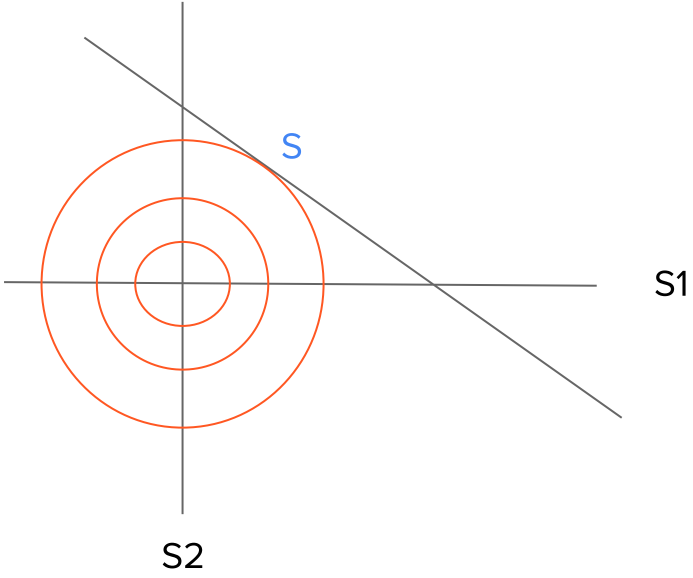
I - Sparsity theory

How to find sparse solution

Let’s suppose Y = XS where Y and X are known. Typically, Y can be a noisy image

Goal : find S sparse; There are an infinite number of S that satisfies this equation.

- ❏ Find S that minimize L2 norm



$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{N} |x_i|^2\right)^{1/2} = \sqrt{x_1^2 + x_2^2 + \ldots + x_N^2}$$

NB : S is not sparse.

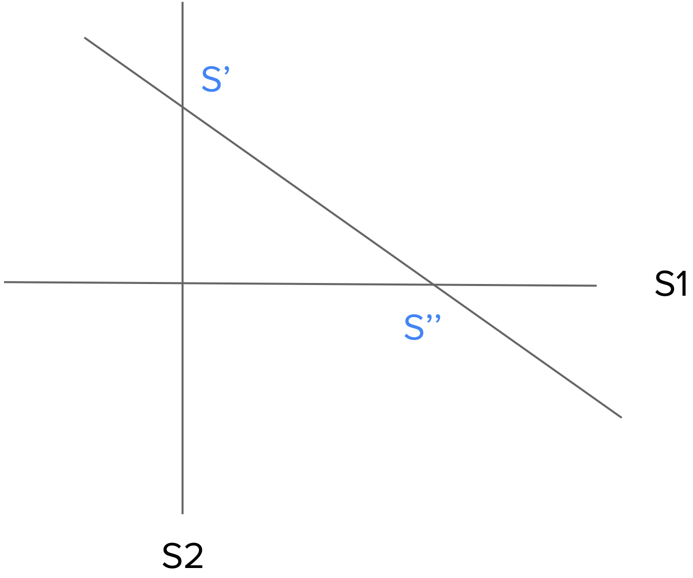
I - Sparsity theory

How to find sparse solution

Let's suppose Y = XS where Y and X are known.

Goal : find S sparse; There are an infinite number of S that satisfies this equation.

- ❑ Find S that minimize L0 norm => S will be sparse but it is N-P hard ➜ hard to compute



$$||x||_0 = \sharp i | x_i \neq 0$$

- ❑ How to choose the solution?
- ❑ This could be hard if the dimension is high

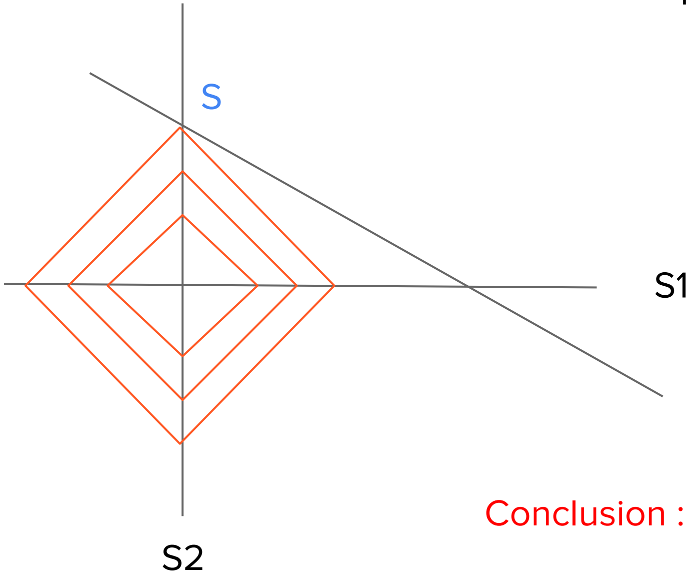
I - Sparsity theory

How to find sparse solution

Let's suppose Y = XS where Y and X are known.

Goal : find S ; There are an infinite number of S that satisfies this equation.

Find S that minimize L1 norm => S will be sparse and can be found S in a convex way.



$$\text{L1 norm: } \|\mathbf{w}\|_1 = \sum_i^n |w_i|$$

- Can be generalised to high dimensions

Conclusion : L1 norm is useful for promoting sparse solution

I - Sparsity theory

Advantages of sparsity

- Is enforced by the L1 norm
- storage and calculation are faster
  (since only non null values are stored and used for calculation)
- Only a few terms are needed to modelize the data
- Prevents overfitting

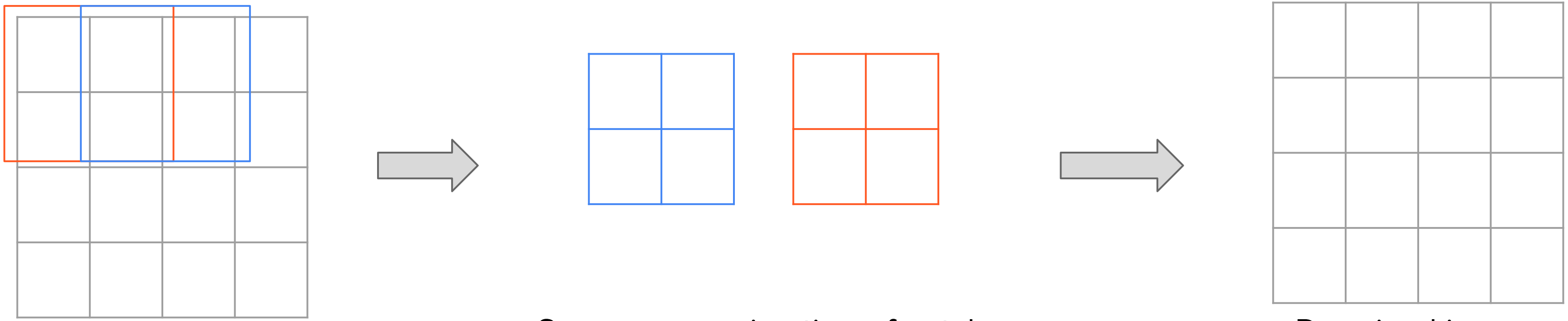
II - Paper's contribution

Classical approach

$$\min_{\boldsymbol{\alpha}_i \in \mathbb{R}^p} \frac{1}{2}\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}_i\|^2 + \lambda\|\boldsymbol{\alpha}_i\|_1$$

- ❑ yi the patch of noisy image
- ❑ D the dictionary
- ❑ α_i the sparse matrix
- ❑ λ controls the amount of regularization

- ❑ L1 norm to introduce sparsity
- ❑ We assume that D is given



Noise image

Sparse approximation of patches

Denoised image

Denoised image is obtained by averaging these estimates since each pixel belongs to several patches

II - Paper's contribution

Paper approach

With a kind of gradient descent method we obtain:

$$\boldsymbol{\alpha}_i^{(t+1)} = S_\lambda \left[\boldsymbol{\alpha}_i^{(t)} + \eta \mathbf{D}^\top \left(\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}_i^{(t)} \right) \right] \quad (1)$$

trans(D)(y_i – $D\alpha_i$) the direction

$\eta > 0$ is a step-size

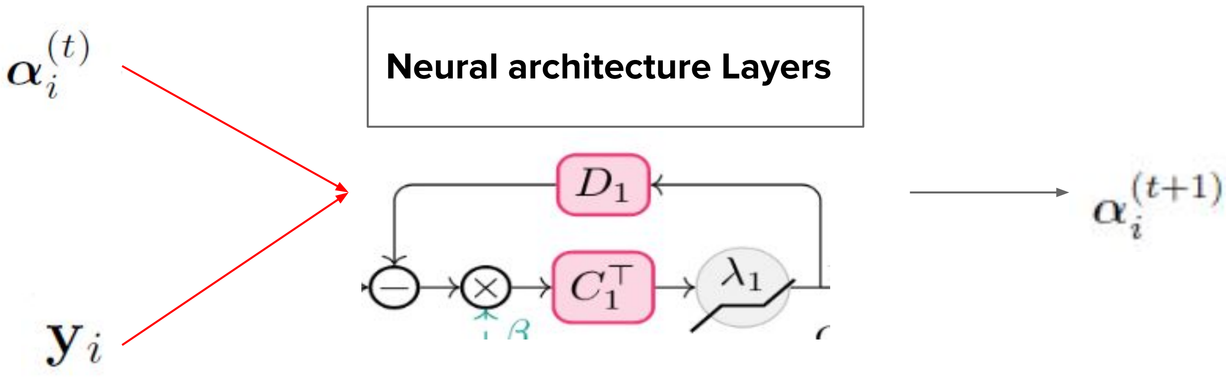
S_λ activation function

II - Paper’s contribution

$$\alpha_i^{(t+1)} = \boxed{S_\lambda}\left[\alpha_i^{(t)} + \eta \mathbf{D}^\top \left(\mathbf{y}_i - \mathbf{D}\alpha_i^{(t)}\right)\right]$$

<u>For each iteration :</u>

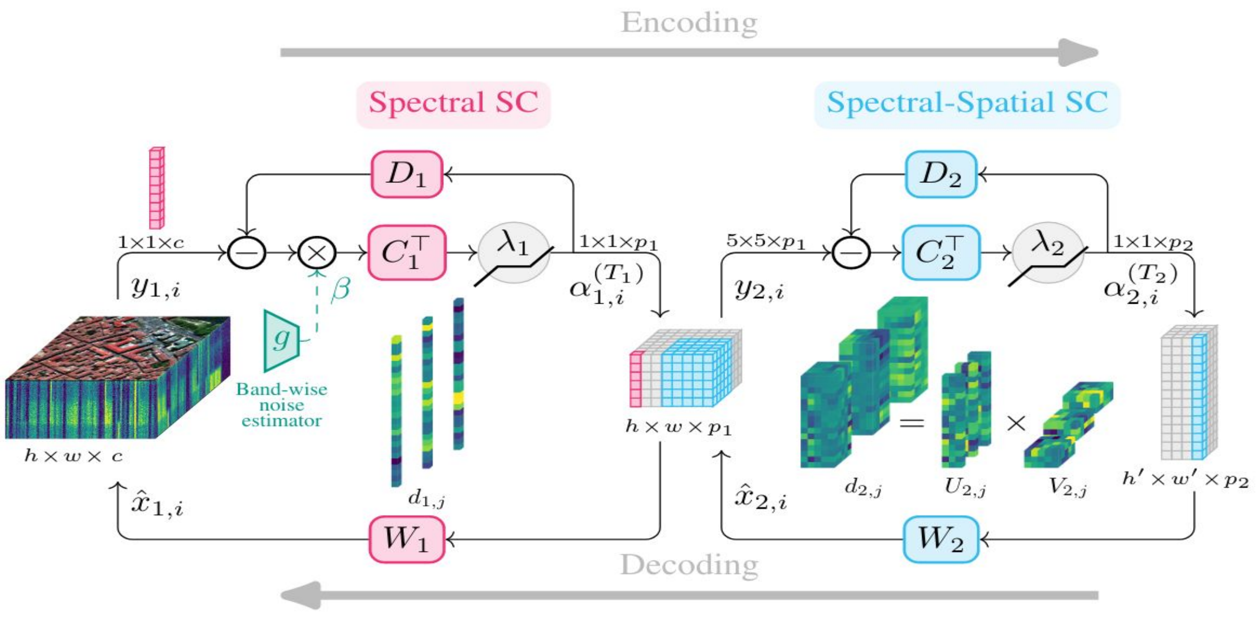
A non linear activation function



parameters

$\mathbf{D}$ η S_λ

II - Paper's contribution



Cost function

$$\min_{\{\alpha_i \in \mathbb{R}^p\}_{i=1,\dots,n}} \frac{1}{2}\left\| \mathbf{y} - \boxed{\frac{1}{m}\sum_{i=1}^{n} \mathbf{R}_i \mathbf{D} \alpha_i \right\|^2 + \lambda \sum_{i=1}^{n} \|\alpha_i\|_1.}$$

III - Paper's results

- **Learn-based approach** : Using dictionary.
- **Learn-free approach** :No dictionary, Ex: BM3D.

	Data req.	*training*	*inference*	*adapt. to new data*	*complex noise*
learning-free	no req.	no training	slow	easy	poor
learning-based	clean data	slow	fast	complicated	good perf.

Table 1 : Comparison between learning-free and learning-based approaches

III - Paper's results

- **Dataset** : ICVL that consists of 204 images of size 1392 × 1300 with 31 bands.

- **Evaluation metrics** :
  - **MPSNR** (Mean Peak Signal-to-Noise Ratio) which is the classical PSNR metric averaged across bands.

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right)$$

MAX_I is the maximum possible pixel value of the image

  - **MSSIM** (Mean Structural Similarity Index Measurement), which is based on the SSIM metric.
    - SSIM is in [0, 1]. 1 indicates perfect structural similarity and 0 indicates no structural similarity

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

- μ_x the average of x;
- μ_y the average of y;
- σ_x^2 the variance of x;
- σ_y^2 the variance of y;
- σ_{xy} the covariance of x and y;
- $c_1 = (k_1L)^2$, $c_2 = (k_2L)^2$ two variables to stabilize the division with weak denominator;
- L the dynamic range of the pixel-values (typically this is $2^{\#bits\ per\ pixel} - 1$);
- $k_1 = 0.01$ and $k_2 = 0.03$ by default.

III - Paper's results

i.i.d Gaussian noise with fixed σ per band

Gaussian noise with different σ for each band.

Noise with correlated σ across band

Stripes noise

σ	Metrics	Noisy	BM3D	BM4D	GLF	LLRT	NGMeet	SMDS	QRNN3D	T3SC	T3SC-SSL
5	MPSNR	34.47	46.17	48.85	51.25	51.86	**52.74**	50.91	48.80	52.62	51.42
	MSSIM	0.7618	0.9843	0.9916	0.9949	0.9951	**0.9960**	0.9944	0.9918	0.9959	0.9952
25	MPSNR	21.44	37.86	39.89	43.16	43.43	44.74	42.83	44.20	**45.38**	44.73
	MSSIM	0.1548	0.9269	0.9510	0.9695	0.9746	0.9796	0.9700	0.9782	**0.9825**	0.9805
50	MPSNR	16.03	34.22	34.22	39.26	39.69	41.08	39.25	41.67	**42.16**	41.62
	MSSIM	0.0502	0.8654	0.8654	0.9197	0.9504	0.9602	0.9382	0.9655	**0.9677**	0.9646
100	MPSNR	10.85	30.43	32.47	34.79	36.39	37.55	35.64	37.19	**38.99**	38.50
	MSSIM	0.0144	0.7557	0.8155	0.7982	0.9182	0.9311	0.8815	0.9140	**0.9439**	0.9394
[0-15]	MPSNR	33.89	45.81	45.35	50.57	48.50	41.67	48.23	52.07	**53.31**	51.26
	MSSIM	0.6386	0.9767	0.9735	0.9948	0.9899	0.9078	0.9900	0.9957	**0.9967**	0.9955
[0-55]	MPSNR	23.36	39.06	38.43	44.22	41.13	32.94	41.76	47.13	**48.64**	46.82
	MSSIM	0.2601	0.9231	0.9074	0.9818	0.9580	0.7565	0.9620	0.9884	**0.9911**	0.9882
[0-95]	MPSNR	19.06	36.17	35.55	41.43	38.44	29.40	38.94	43.98	**46.30**	44.75
	MSSIM	0.1614	0.8760	0.8540	0.9674	0.9354	0.6609	0.9357	0.9753	**0.9859**	0.9822
Corr.	MPSNR	28.85	42.73	42.13	47.05	45.76	38.06	45.98	48.90	**49.89**	48.78
	MSSIM	0.4740	0.9599	0.9070	0.9881	0.9824	0.8536	0.9835	0.9911	**0.9923**	0.9911
Strip.	MPSNR	21.20	34.88	37.70	42.06	39.38	39.78	41.98	44.60	**44.74**	43.80
	MSSIM	0.1508	0.8641	0.9198	0.9628	0.9258	0.9333	0.9655	**0.9806**	0.9805	0.9773

Table 2 : Denoising performance on ICVL

Conclusion

THANK YOU
FOR YOUR ATTENTION

Slides

III - Our implementation

- **Normalize data** => gradient descent method is speed and efficient.

Different types of synthetic noise for models evaluation

Noise patterns	Explanation
i.i.d Gaussian noise with known variance	Same noise on all bands.
Gaussian noise with unknown band-dependent variance	different standard deviation for each band, which is uniformly drawn in a fixed interval
Noise with spectrally correlated variance	Gaussian noise with standard deviation varying continuously across bands
Stripes noise	we applied additive stripes noise to 33% of bands