Unrolled proximal algorithms

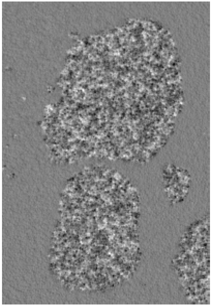
for piecewise homogeneous fractal images analysis

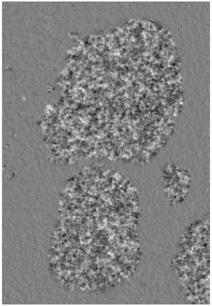
Machine Learning Reading Group

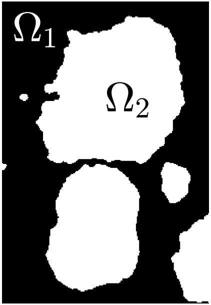
December 2022

Barbara Pascal

Joint work with H. T. V. Le (PhD student), N. Pustelnik (CNRS researcher)
P. Abry (CNRS research director) *Laboratoire de Physique, ENS Lyon*
M. Foare (Assistant professor), *CPE Lyon, LIP*



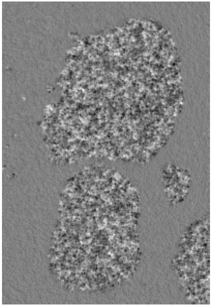


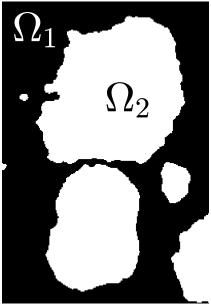


Goal: obtain a partition of the image into K homogeneous textures

$$\Omega = \Omega_1 \bigsqcup \ldots \bigsqcup \Omega_K$$

Textured image segmentation





Goal: obtain a partition of the image into K **homogeneous textures**

$$\Omega = \Omega_1 \bigsqcup \ldots \bigsqcup \Omega_K$$

Two paradigms for inverse problems

Goal: Recover quantity of interest $\overline{\boldsymbol{x}}$ from observation $\boldsymbol{y}$.

Goal: Recover quantity of interest $\overline{\boldsymbol{x}}$ from observation $\boldsymbol{y}$.

Unsupervised

1. propose a model:

$$\boldsymbol{y} = \mathcal{M}(\overline{\boldsymbol{x}}) + \zeta;$$

2. balance model & constraints

$\rightarrow$ variational approach:

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \ \mathcal{D}(\boldsymbol{y}, \mathcal{M}(\boldsymbol{x})) + \lambda \mathcal{R}(\boldsymbol{x});$$

3. fine-tune hyperparameters

$\rightarrow$ statistical arguments:

$$\lambda^\dagger \approx \underset{\lambda}{\operatorname{argmin}} \ \mathbb{E}\left[\mathcal{D}(\widehat{\boldsymbol{x}}, \overline{\boldsymbol{x}})\right]$$

Goal: Recover quantity of interest $\overline{\boldsymbol{x}}$ from observation $\boldsymbol{y}$.

Unsupervised

1. propose a model:

$$\boldsymbol{y} = \mathcal{M}(\overline{\boldsymbol{x}}) + \boldsymbol{\zeta};$$

2. balance model & constraints

$\rightarrow$ variational approach:

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}}\ \mathcal{D}(\boldsymbol{y}, \mathcal{M}(\boldsymbol{x})) + \lambda \mathcal{R}(\boldsymbol{x});$$

3. fine-tune hyperparameters

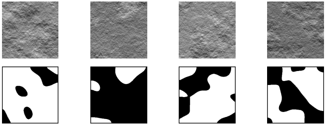
$\rightarrow$ statistical arguments:

$$\lambda^{\dagger} \approx \underset{\lambda}{\operatorname{argmin}}\ \mathbb{E}\left[\mathcal{D}(\widehat{\boldsymbol{x}}, \overline{\boldsymbol{x}})\right]$$

Supervised

1. build an annotated database:

$$\{(\overline{\boldsymbol{x}}_s, \boldsymbol{y}_s)\}_{s=1}^{S}$$



2. design of a parametric estimator

$$\widehat{\boldsymbol{x}} = NN(\boldsymbol{y}; \boldsymbol{\vartheta})$$

3. minimize the empirical risk

$$\underset{\vartheta}{\operatorname{minimize}}\ \frac{1}{S} \sum_{s=1}^{S} \left\| NN(\boldsymbol{y}_s; \vartheta) - \overline{\boldsymbol{x}}_s \right\|^2$$

Two paradigms for inverse problems

Goal: Recover quantity of interest $\overline{\boldsymbol{x}}$ from observation $\boldsymbol{y}$.

Unsupervised

1. propose a model:

$$\boldsymbol{y} = \mathcal{M}(\overline{\boldsymbol{x}}) + \zeta;$$

2. balance model & constraints

$\rightarrow$ variational approach:

$$\widehat{\boldsymbol{x}} = \operatorname*{argmin}_{\boldsymbol{x}} \ \mathcal{D}(\boldsymbol{y}, \mathcal{M}(\boldsymbol{x})) + \lambda \mathcal{R}(\boldsymbol{x});$$

3. fine-tune hyperparameters

$\rightarrow$ statistical arguments:

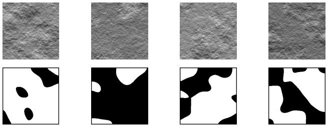
$$\lambda^{\dagger} \approx \operatorname*{argmin}_{\lambda} \mathbb{E}\left[\mathcal{D}(\widehat{\boldsymbol{x}}, \overline{\boldsymbol{x}})\right]$$

physics of the problem & standard methods

Supervised

1. build an annotated database:

$$\{(\overline{\boldsymbol{x}}_s, \boldsymbol{y}_s)\}_{s=1}^{S}$$



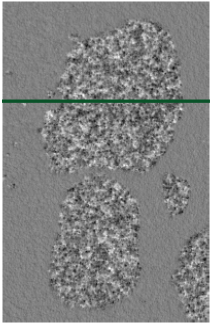
2. design of a parametric estimator

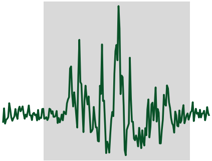
$$\widehat{\boldsymbol{x}} = NN(\boldsymbol{y}; \boldsymbol{\vartheta})$$

3. minimize the empirical risk

$$\operatorname*{minimize}_{\vartheta} \ \frac{1}{S} \sum_{s=1}^{S} \left\| NN(\boldsymbol{y}_s; \vartheta) - \overline{\boldsymbol{x}}_s \right\|^2$$

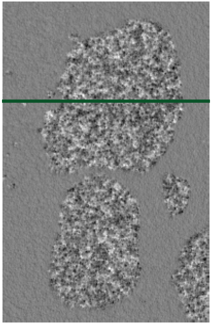
Piecewise monofractal model

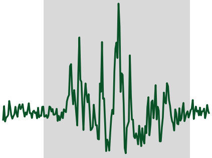




Fractals attributes

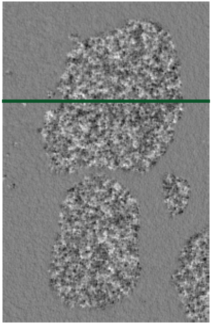
- variance σ^2 *amplitude of variations*

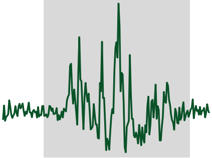




Fractals attributes

- variance σ^2 — *amplitude of variations*
- local regularity h — *scale invariance*

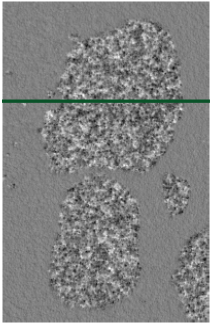


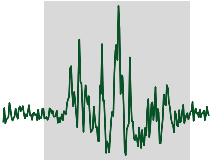


Fractals attributes

- variance σ^2 *amplitude of variations*
- local regularity h *scale invariance*

$$|f(x) - f(y)| \leq \sigma(x)|x - y|^{h(x)}$$

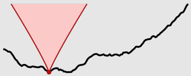




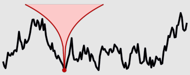
Fractals attributes

- variance σ^2 *amplitude of variations*
- local regularity h *scale invariance*

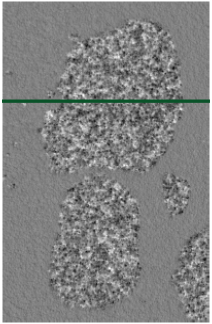
$$|f(x) - f(y)| \leq \sigma(x)|x - y|^{h(x)}$$

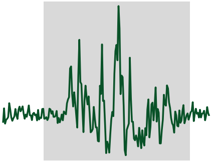


$h(x) \equiv h_1 = 0.9$



$h(x) \equiv h_2 = 0.3$





Piecewise monofractal model

Fractals attributes

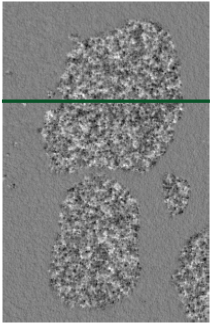
- variance σ^2 — *amplitude of variations*
- local regularity h — *scale invariance*

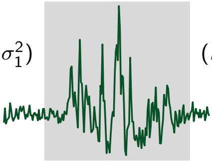
$$|f(x) - f(y)| \leq \sigma(x)|x - y|^{h(x)}$$

$h(x) \equiv h_1 = 0.9$ $\quad$ $h(x) \equiv h_2 = 0.3$

Segmentation

- σ^2 and h piecewise constant



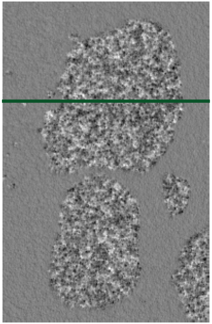


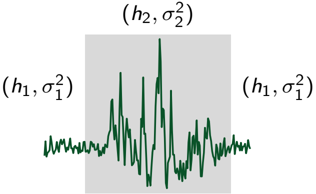
Fractals attributes

- variance σ^2 — *amplitude of variations*
- local regularity h — *scale invariance*

$$|f(x) - f(y)| \leq \sigma(x)|x - y|^{h(x)}$$

$h(x) \equiv h_1 = 0.9$ $\qquad$ $h(x) \equiv h_2 = 0.3$





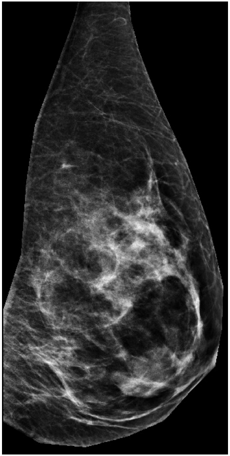
Segmentation

- σ^2 and h piecewise constant
- region Ω_k characterized by (σ_k^2, h_k)

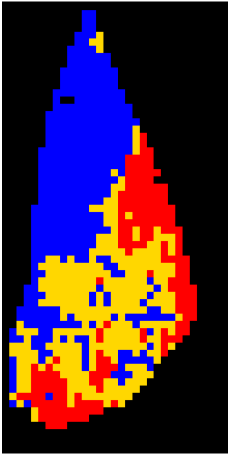
Observation: local regularity reflects health of breast tissues

(Marin et al., *Medical Physics*, 2017)

Mammogram



Local regularity (WTMM)



Microenvironment analysis via density fluctuations

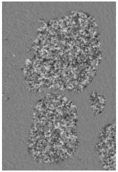
Healthy $\Leftrightarrow$ structured

- $h < 0.5$: negative correlations
- $h > 0.5$: positive correlations

Prone to cancer $\Leftrightarrow$ unstructured

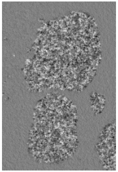
- $h = 0.5$: absence of correlations

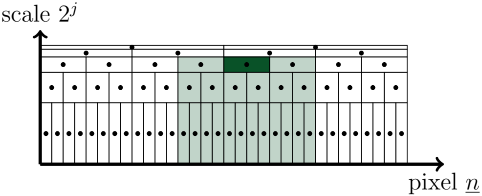
Textured image



Textured image

Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$



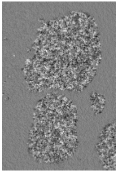


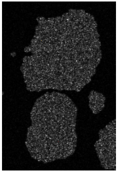
Multiscale analysis

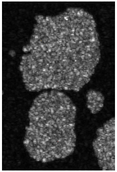
Textured image

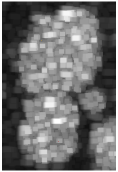
Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

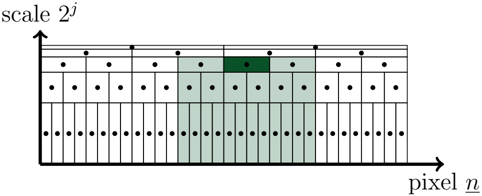
Scale $a = 2^1$ $a = 2^2$ … $a = 2^5$









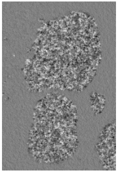


Multiscale analysis

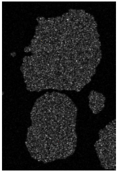
Textured image

Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

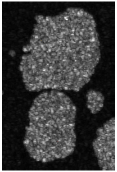
Scale $a = 2^1$ $a = 2^2$ $\dots$ $a = 2^5$



$a = 2^1$

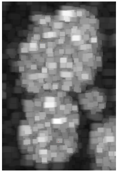


$a = 2^2$



$\dots$

$a = 2^5$

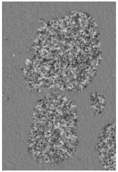


Proposition (Jaffard, 2004), (Wendt, 2008)

$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a \to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto \log(\boldsymbol{\sigma}^2) \\ \text{(variance)}}}$$

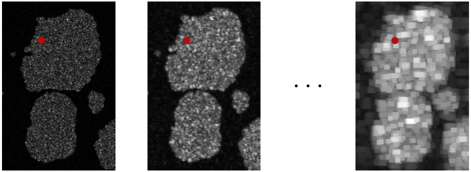
Multiscale analysis

Textured image



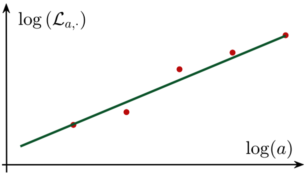
Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

Scale $a = 2^1$ $a = 2^2$ $\dots$ $a = 2^5$



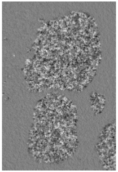
Proposition (Jaffard, 2004), (Wendt, 2008)

$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a\to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto \log(\boldsymbol{\sigma}^2) \\ \text{(variance)}}}$$



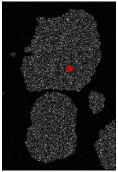
Multiscale analysis

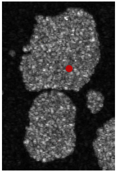
Textured image

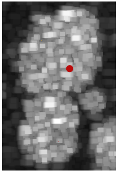


Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

Scale $a = 2^1$ $a = 2^2$ … $a = 2^5$

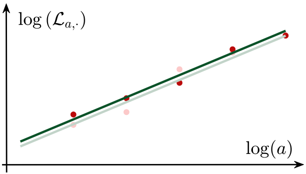




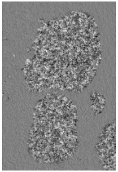


Proposition (Jaffard, 2004), (Wendt, 2008)

$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a \to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto \log(\boldsymbol{\sigma}^2) \\ \text{(variance)}}}$$

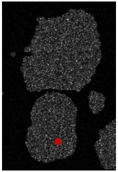


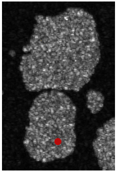
Textured image

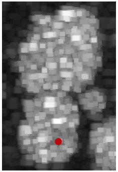


Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

Scale $a = 2^1$ $a = 2^2$ $\ldots$ $a = 2^5$

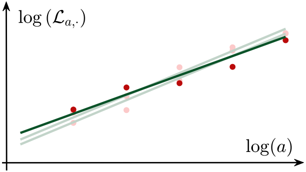






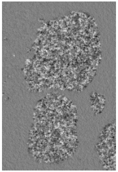
Proposition (Jaffard, 2004), (Wendt, 2008)

$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a\to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto\log(\boldsymbol{\sigma}^2)\\ \text{(variance)}}}$$



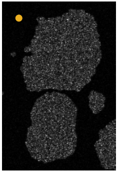
Multiscale analysis

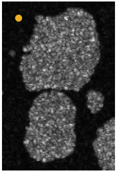
Textured image

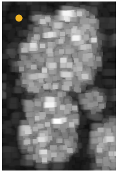


Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

Scale $a = 2^1$ $a = 2^2$ $\dots$ $a = 2^5$

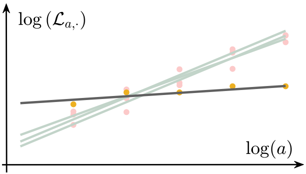






Proposition (Jaffard, 2004), (Wendt, 2008)

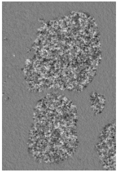
$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a \to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto \log(\boldsymbol{\sigma}^2) \\ \text{(variance)}}}$$

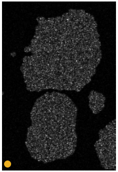


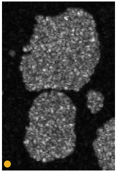
Textured image

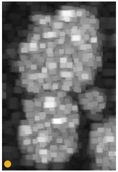
Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

Scale $a = 2^1$ $a = 2^2$ … $a = 2^5$



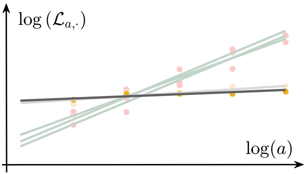






Proposition (Jaffard, 2004), (Wendt, 2008)

$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a\to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto\log(\boldsymbol{\sigma}^2)\\ \text{(variance)}}}$$

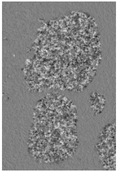


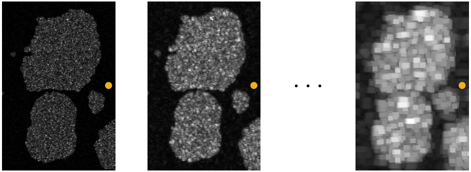
Multiscale analysis

Textured image

Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

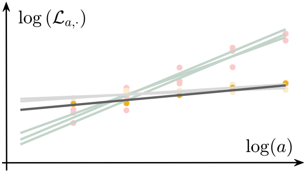
Scale $a = 2^1$ $a = 2^2$ ... $a = 2^5$





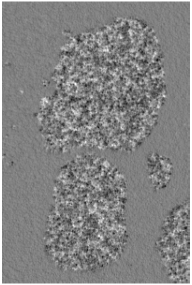
Proposition (Jaffard, 2004), (Wendt, 2008)

$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a\to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto \log(\boldsymbol{\sigma}^2) \\ \text{(variance)}}}$$



Linear regression $\log\left(\mathcal{L}_{a,\cdot}\right) \simeq \log(a) \underset{\text{regularity}}{\boldsymbol{h}} + \underset{\propto \log(\boldsymbol{\sigma}^2)}{\boldsymbol{v}}$

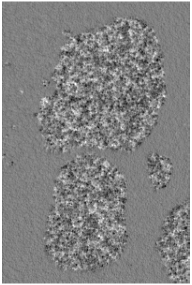
Textured image



Linear regression $\log\left(\mathcal{L}_{a,\cdot}\right) \simeq \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\propto \log(\boldsymbol{\sigma}^2)}$

$$\left(\widehat{\boldsymbol{h}}^{\text{LR}}, \widehat{\boldsymbol{v}}^{\text{LR}}\right) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a=a_{\min}}^{a_{\max}} \left\|\log\left(\mathcal{L}_{a,\cdot}\right) - \log(a)\boldsymbol{h} - \boldsymbol{v}\right\|_2^2$$

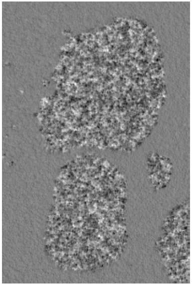
Textured image



Linear regression $\log\left(\mathcal{L}_{a,\cdot}\right) \simeq \log(a) \underset{\text{regularity}}{\boldsymbol{h}} + \underset{\propto \log(\boldsymbol{\sigma}^2)}{\boldsymbol{v}}$

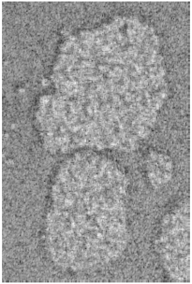
$$\left(\widehat{\boldsymbol{h}}^{\mathrm{LR}}, \widehat{\boldsymbol{v}}^{\mathrm{LR}}\right) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a=a_{\min}}^{a_{\max}} \left\| \log\left(\mathcal{L}_{a,\cdot}\right) - \log(a)\boldsymbol{h} - \boldsymbol{v} \right\|_2^2$$

Textured image



Local regularity $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$

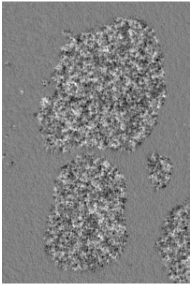
Local power $\widehat{\boldsymbol{v}}^{\mathrm{LR}}$



Linear regression $\underbrace{\mathbb{E}\log\left(\mathcal{L}_{a,\cdot}\right)}_{\text{expected value}} = \log(a)\underbrace{\bar{\boldsymbol{h}}}_{\text{regularity}} + \underbrace{\bar{\boldsymbol{v}}}_{\propto\log(\boldsymbol{\sigma}^2)}$

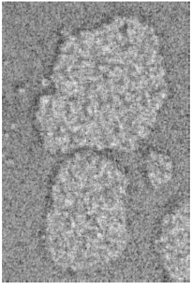
$$\left(\widehat{\boldsymbol{h}}^{\mathrm{LR}}, \widehat{\boldsymbol{v}}^{\mathrm{LR}}\right) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a=a_{\min}}^{a_{\max}} \left\|\log\left(\mathcal{L}_{a,\cdot}\right) - \log(a)\boldsymbol{h} - \boldsymbol{v}\right\|_2^2$$

Textured image



Local regularity $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$

Local power $\widehat{\boldsymbol{v}}^{\mathrm{LR}}$



$\longrightarrow$ large estimation variance

Proximal algorithms for image processing:

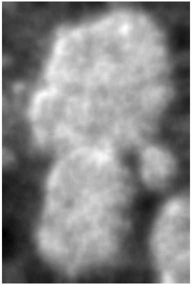
Texture segmentation based on fractal attributes in a variationnal framework

Smoothing filter (linear)

$$\left(\mathbf{I} + \lambda \mathbf{D}^\top \mathbf{D}\right)^{-1} \widehat{\boldsymbol{h}}^{\mathsf{LR}}$$

Linear regression $\widehat{\boldsymbol{h}}^{\mathsf{LR}}$

Smoothing



Smoothing filter (linear)

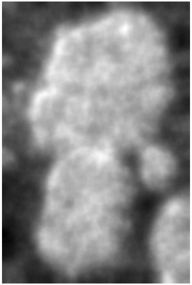
$$\left(\mathbf{I} + \lambda \mathbf{D}^\top \mathbf{D}\right)^{-1} \widehat{\boldsymbol{h}}^{\mathsf{LR}}$$

Regularization (nonlinear)

$$\operatorname*{argmin}_{\boldsymbol{h}} \|\boldsymbol{h} - \widehat{\boldsymbol{h}}^{\mathsf{LR}}\|_2^2 + \mathcal{R}(\boldsymbol{h})$$

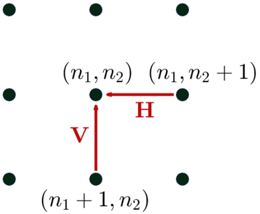
Linear regression $\widehat{\boldsymbol{h}}^{\mathsf{LR}}$

Smoothing



aim: enforce piecewise behavior of estimate

$$\underset{\boldsymbol{h}}{\operatorname{argmin}} \|\boldsymbol{h} - \widehat{\boldsymbol{h}}^{\mathsf{LR}}\|_2^2 + \lambda \mathsf{TV}(\boldsymbol{h})$$



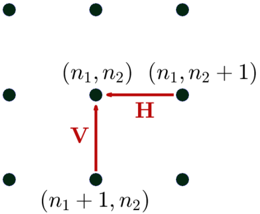
Discrete difference operator

$$(\mathbf{D}\boldsymbol{h})_{n_1,n_2} = \frac{1}{2}\begin{pmatrix} h_{n_1,n_2+1} - h_{n_1,n_2} \\ h_{n_1+1,n_2} - h_{n_1,n_2} \end{pmatrix}$$

---

aim: enforce piecewise behavior of estimate

$$\underset{\boldsymbol{h}}{\operatorname{argmin}} \; \|\boldsymbol{h} - \widehat{\boldsymbol{h}}^{\mathsf{LR}}\|_2^2 + \lambda \mathsf{TV}(\boldsymbol{h})$$

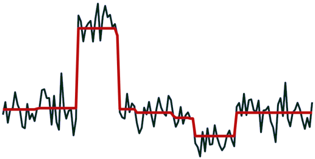


Discrete difference operator

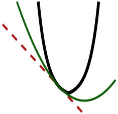
$$(\mathbf{D}\boldsymbol{h})_{n_1,n_2} = \frac{1}{2}\begin{pmatrix} h_{n_1,n_2+1} - h_{n_1,n_2} \\ h_{n_1+1,n_2} - h_{n_1,n_2} \end{pmatrix}$$

Total Variation penalization $\mathsf{TV}(\boldsymbol{h})$

$$\|\mathbf{D}\boldsymbol{h}\|_1 = \sum_{n_1=1}^{N-1}\sum_{n_2=1}^{N-1} |(\mathbf{H}\boldsymbol{h})_{n_1,n_2}| + |(\mathbf{V}\boldsymbol{h})_{n_1,n_2}|$$



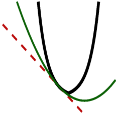
$$\underset{\boldsymbol{h}}{\text{minimize}} \quad \underbrace{\|\boldsymbol{h} - \widehat{\boldsymbol{h}}^{\mathsf{LR}}\|_2^2}_{\textbf{Quadratic norm}} \quad + \quad \underbrace{\lambda \, \|\mathbf{D}\boldsymbol{h}\|_1}_{\textbf{Total Variation}}$$



μ-strongly convex

Functionals minimization

$$\underset{\boldsymbol{h}}{\text{minimize}} \quad \underbrace{\|\boldsymbol{h} - \widehat{\boldsymbol{h}}^{\mathsf{LR}}\|_2^2}_{\textbf{Quadratic norm}} \quad + \quad \underbrace{\lambda \, \|\mathbf{D}\boldsymbol{h}\|_1}_{\textbf{Total Variation}}$$

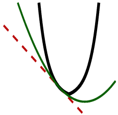


μ-strongly convex

- gradient descent $\boldsymbol{h}^{n+1} = \boldsymbol{h}^{n} - \tau \nabla \varphi(\boldsymbol{h}^{n})$

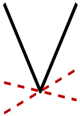
Functionals minimization

$$\underset{\boldsymbol{h}}{\text{minimize}} \quad \underbrace{\|\boldsymbol{h} - \widehat{\boldsymbol{h}}^{\mathrm{LR}}\|_2^2}_{\textbf{Quadratic norm}} \quad + \quad \underbrace{\lambda \, \|\mathbf{D}\boldsymbol{h}\|_1}_{\textbf{Total Variation}}$$



μ-strongly convex

nonsmooth

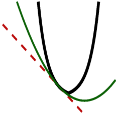


- gradient descent $\boldsymbol{h}^{n+1} = \boldsymbol{h}^n - \tau \nabla \varphi(\boldsymbol{h}^n)$
- implicit subgradient descent: proximal point algorithm

$$\boldsymbol{h}^{n+1} = \boldsymbol{h}^n - \tau \boldsymbol{u}^n, \ \boldsymbol{u}^n \in \partial\varphi(\boldsymbol{h}^{n+1})$$
$$\Leftrightarrow \ \boldsymbol{h}^{n+1} = (\mathbf{I} + \tau\partial\varphi)^{-1}(\boldsymbol{h}^n) := \mathrm{prox}_{\tau\varphi}(\boldsymbol{h}^n)$$

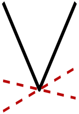
Functionals minimization

$$\underset{\boldsymbol{h}}{\text{minimize}} \quad \underbrace{\|\boldsymbol{h} - \widehat{\boldsymbol{h}}^{\mathsf{LR}}\|_2^2}_{\textbf{Quadratic norm}} \quad + \quad \underbrace{\lambda \, \|\mathbf{D}\boldsymbol{h}\|_1}_{\textbf{Total Variation}}$$



$\boldsymbol{\mu}$-strongly convex

nonsmooth



- gradient descent $\boldsymbol{h}^{n+1} = \boldsymbol{h}^n - \tau \nabla \varphi(\boldsymbol{h}^n)$
- implicit subgradient descent: proximal point algorithm

$$\boldsymbol{h}^{n+1} = \boldsymbol{h}^n - \tau \boldsymbol{u}^n, \; \boldsymbol{u}^n \in \partial \varphi(\boldsymbol{h}^{n+1})$$

$$\Leftrightarrow \quad \boldsymbol{h}^{n+1} = (\mathbf{I} + \tau \partial \varphi)^{-1}(\boldsymbol{h}^n) := \mathsf{prox}_{\tau\varphi}(\boldsymbol{h}^n)$$

- (accelerated) proximal splitting algorithms: split $\|\cdot\|_2^2$, $\|\cdot\|_1$ and $\mathbf{D}$

FISTA	**Chambolle - Pock**
proximity operator $\mathsf{prox}_{\tau\|\cdot\|_1}$	proximity operator $\mathsf{prox}_{\tau\|\cdot\|_1}$, $\mathbf{D}$ and $\mathbf{D}^\top$
gradient of $\|\mathbf{D}^\top \cdot - \widehat{\boldsymbol{h}}^{\mathsf{LR}}\|_2^2$	proximity operator $\mathsf{prox}_{\tau\|\cdot - \widehat{\boldsymbol{h}}^{\mathsf{LR}}\|_2^2}$

aim: enforce piecewise behavior of estimate

$$\underset{\boldsymbol{h}}{\operatorname{argmin}} \|\boldsymbol{h} - \widehat{\boldsymbol{h}}^{\mathrm{LR}}\|_2^2 + \lambda \mathrm{TV}(\boldsymbol{h})$$

FISTA in the dual
(Fast Iterative Soft Thresholding Algorithm)

init $\boldsymbol{g}, \bar{\boldsymbol{g}} \in \mathbb{R}^{2|\Omega|}$

for $n = 0, 1, \ldots$

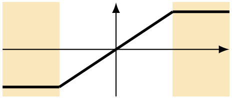
$$\boldsymbol{u}^{n+1} = (\mathbf{I} - \gamma \mathbf{D}\mathbf{D}^\top)\bar{\boldsymbol{g}}^n + \gamma \mathbf{D}\boldsymbol{h}^{\mathrm{LR}}$$

$$\boldsymbol{g}^{n+1} = \mathrm{prox}_{\gamma(\lambda\|\cdot\|_1)^*}\left(\boldsymbol{u}^{n+1}\right)$$

$$\tau_{n+1} = b^{-1}(n+b)$$

$$\bar{\boldsymbol{g}}^{n+1} = \boldsymbol{g}^n + \tau_{n+1}^{-1}(\tau_n - 1)\left(\boldsymbol{g}^{n+1} - \boldsymbol{g}^n\right)$$

$$\widehat{\boldsymbol{h}} = \widehat{\boldsymbol{h}^{\mathrm{LR}}} - \mathbf{D}^\top \boldsymbol{g}^\infty$$



$\mathrm{prox}_{\gamma(\lambda\|\cdot\|_1)^*}(\boldsymbol{x})$

Chambolle-Pock
(Accelerated thanks to strong-convexity)

init $\boldsymbol{h}^0, \bar{\boldsymbol{h}}^0 \in \mathbb{R}^{|\Omega|}; \boldsymbol{g} \in \mathbb{R}^{2|\Omega|}$

for $n = 0, 1, \ldots$

$$\boldsymbol{g}^{n+1} = \mathrm{prox}_{\sigma_n(\lambda\|\cdot\|_1)^*}\left(\boldsymbol{g}^n + \sigma_n \mathbf{D}\bar{\boldsymbol{h}}^n\right)$$

$$\boldsymbol{h}^{n+1} = \mathrm{prox}_{\tau_n\|\cdot - \widehat{\boldsymbol{h}}^{\mathrm{LR}}\|_2^2}\left(\boldsymbol{h}^n - \tau_n \mathbf{D}^\top \boldsymbol{g}^{n+1}\right)$$

$$\theta_n = \sqrt{1 + 2\mu\tau_n}$$

$$\tau_{n+1} = \tau_n/\theta_n, \quad \sigma_{n+1} = \theta_n \sigma_n$$

$$\bar{\boldsymbol{h}}^{n+1} = \boldsymbol{h}^n + \theta_n\left(\boldsymbol{h}^{n+1} - \boldsymbol{h}^n\right)$$

$$\widehat{\boldsymbol{h}} = \boldsymbol{h}^\infty$$

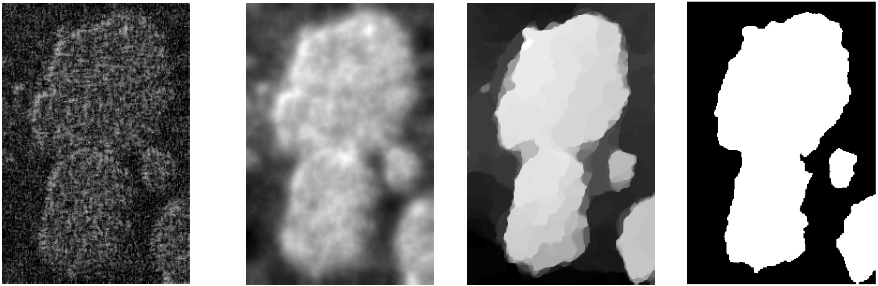
Smoothing filter (linear)

$$\left(\mathbf{I} + \lambda \mathbf{D}^\top \mathbf{D}\right)^{-1} \widehat{\boldsymbol{h}}^{\mathsf{LR}}$$

TV denoising (nonlinear)

$$\underset{\boldsymbol{h}}{\operatorname{argmin}} \|\boldsymbol{h} - \widehat{\boldsymbol{h}}^{\mathsf{LR}}\|_2^2 + \lambda \mathrm{TV}(\boldsymbol{h})$$

Linear regression $\widehat{\boldsymbol{h}}^{\mathsf{LR}}$ | Smoothing | TV denoising | Segmentation



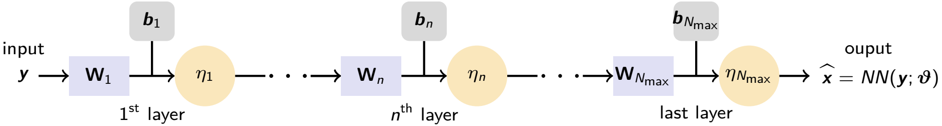
Stein-like **automated data driven** selection of regularization parameter λ

(Pascal et al., *J. Math. Imaging Vis.*, 2021)

Neural network based deep learning:

Unrolled proximal algorithm for supervised texture segmentation

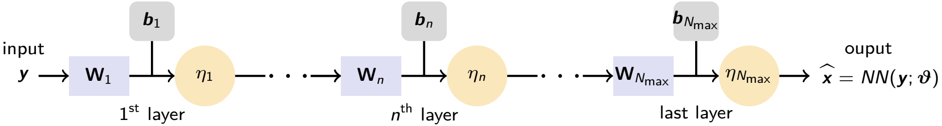
Deep neural network [Le Cun, 1990], [Goodfellow et col., 2016]



elementary layer: $\boldsymbol{x}^{[n]} = \eta_n \left(\mathbf{W}_n \boldsymbol{x}^{[n-1]} + \boldsymbol{b}_n \right)$, $\mathbf{W}_n$ linear, $\boldsymbol{b}_n$ bias and η_n nonlinear

learned weights: $\vartheta = \{(\mathbf{W}_n, \boldsymbol{b}_n)\}_{n=1}^{N_{\max}}$

Deep neural network [Le Cun, 1990], [Goodfellow et col., 2016]



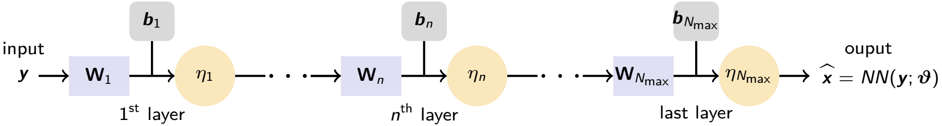
elementary layer: $\boldsymbol{x}^{[n]} = \eta_n\left(\mathbf{W}_n\boldsymbol{x}^{[n-1]} + \boldsymbol{b}_n\right)$, $\mathbf{W}_n$ linear, $\boldsymbol{b}_n$ bias and η_n nonlinear

learned weights: $\boldsymbol{\vartheta} = \{(\mathbf{W}_n, \boldsymbol{b}_n)\}_{n=1}^{N_{\max}}$

✓ ability to represent highly complex data; very good performance

Unrolled proximal algorithms

Deep neural network [Le Cun, 1990], [Goodfellow et col., 2016]



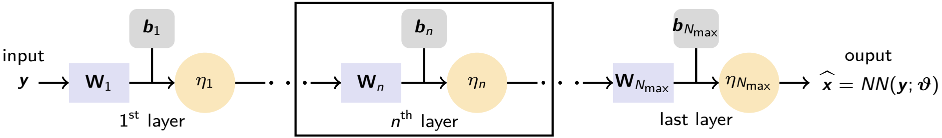
<u>elementary layer:</u> $\boldsymbol{x}^{[n]} = \eta_n \left(\mathbf{W}_n \boldsymbol{x}^{[n-1]} + \boldsymbol{b}_n \right)$, $\mathbf{W}_n$ linear, $\boldsymbol{b}_n$ bias and η_n nonlinear

learned weights: $\boldsymbol{\vartheta} = \{(\mathbf{W}_n, \boldsymbol{b}_n)\}_{n=1}^{N_{\max}}$

✓ ability to represent highly complex data; very good performance

✗ lack of interpretability and robustness; large training dataset required

Deep neural network [Le Cun, 1990], [Goodfellow et col., 2016]



<u>elementary layer:</u> $\boldsymbol{x}^{[n]} = \eta_n\left(\mathbf{W}_n \boldsymbol{x}^{[n-1]} + \boldsymbol{b}_n\right)$, $\mathbf{W}_n$ linear, $\boldsymbol{b}_n$ bias and η_n nonlinear

learned weights: $\boldsymbol{\vartheta} = \{(\mathbf{W}_n, \boldsymbol{b}_n)\}_{n=1}^{N_{\max}}$

✓ ability to represent highly complex data; very good performance

✗ lack of interpretability and robustness; large training dataset required

---

init $\boldsymbol{g} \in \mathbb{R}^{2|\Omega|}$

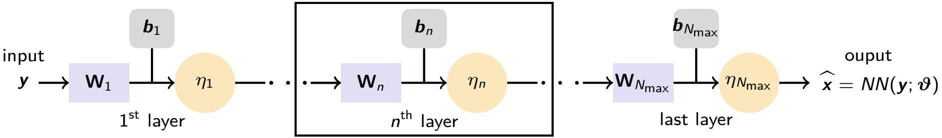
for $n = 0, 1, \ldots$

$\boldsymbol{u}^{n+1} = (\mathbf{I} - \gamma \mathbf{D}\mathbf{D}^\top)\boldsymbol{g}^n + \gamma \mathbf{D}\boldsymbol{h}^{\mathrm{LR}}$

$\boldsymbol{g}^{n+1} = \mathrm{prox}_{\gamma(\lambda\|\cdot\|_1)^*}\left(\boldsymbol{u}^{n+1}\right)$

$\widehat{\boldsymbol{h}} = \widehat{\boldsymbol{h}^{\mathrm{LR}}} - \mathbf{D}^\top \boldsymbol{g}^\infty$

Deep neural network [Le Cun, 1990], [Goodfellow et col., 2016]



elementary layer: $\boldsymbol{x}^{[n]} = \eta_n \left(\mathbf{W}_n \boldsymbol{x}^{[n-1]} + \boldsymbol{b}_n \right)$, $\mathbf{W}_n$ linear, $\boldsymbol{b}_n$ bias and η_n nonlinear

learned weights: $\boldsymbol{\vartheta} = \{(\mathbf{W}_n, \boldsymbol{b}_n)\}_{n=1}^{N_{\max}}$

✓ ability to represent highly complex data; very good performance

✗ lack of interpretability and robustness; large training dataset required

---

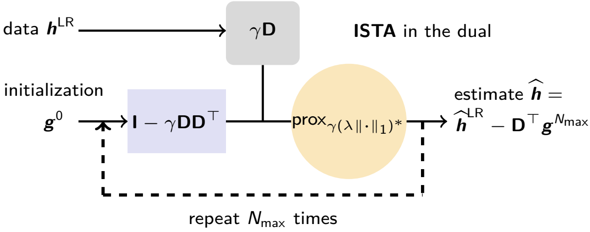
init $\boldsymbol{g} \in \mathbb{R}^{2|\Omega|}$

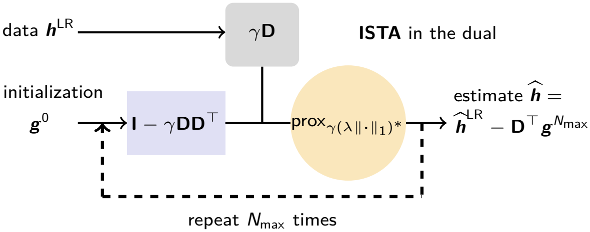
for $n = 0, 1, \dots$

$\boldsymbol{u}^{n+1} = (\mathbf{I} - \gamma \mathbf{D}\mathbf{D}^\top) \boldsymbol{g}^n + \gamma \mathbf{D} \boldsymbol{h}^{\mathrm{LR}}$

$\boldsymbol{g}^{n+1} = \operatorname{prox}_{\gamma(\lambda \|\cdot\|_1)^*} \left(\boldsymbol{u}^{n+1} \right)$

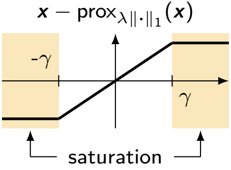
$\widehat{\boldsymbol{h}} = \widehat{\boldsymbol{h}^{\mathrm{LR}}} - \mathbf{D}^\top \boldsymbol{g}^\infty$





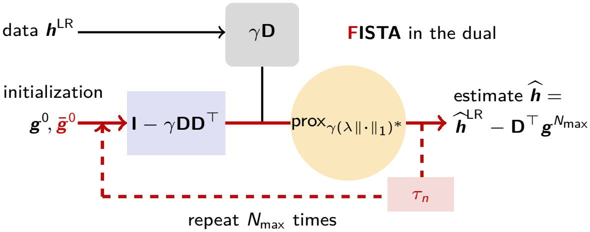
Neural network from unrolling of ISTA: Deep ISTA Hloc

- linear operator $\mathbf{W}_n$: $\mathbf{I} - \mathbf{D}_n^{[1]}\mathbf{D}_n^{[2]}$:
  - decouple filters $\mathbf{D}_n^{[1]}$ and $\mathbf{D}_n^{[2]}$ in a given layer;
  - decouple filters $\mathbf{D}_n^{[\cdot]}$ at different layer;
  - γ_n implicitly learn in $\mathbf{D}_n^{[1]}$;
- bias $\boldsymbol{b}_n$: $\mathbf{D}_n^{[1]}\widehat{\boldsymbol{h}}^{\mathrm{LR}}$ unorthodox “input"
- η_n : $\mathrm{prox}_{\lambda\|\cdot\|_1} = \mathrm{HardTanh}_\lambda$



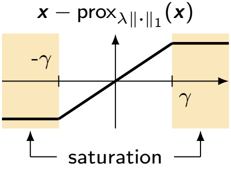
(Le et al., *EUSIPCO*, 2022)

Unrolled proximal algorithms

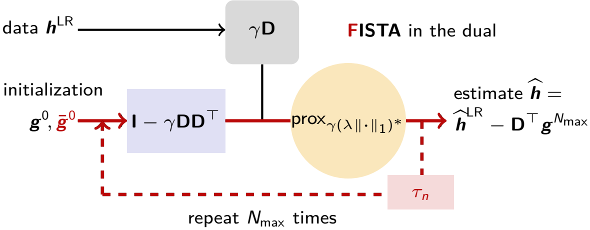


Neural network from unrolling of FISTA: Deep FISTA Hloc (DFH)

- linear operator $\mathbf{W}_n$: $\mathbf{I} - \mathbf{D}_n^{[1]}\mathbf{D}_n^{[2]}$:
  - decouple filters $\mathbf{D}_n^{[1]}$ and $\mathbf{D}_n^{[2]}$ in a given layer;
  - decouple filters $\mathbf{D}_n^{[\cdot]}$ at different layer;
  - γ_n implicitly learn in $\mathbf{D}_n^{[1]}$; inertia parameter τ_n;
- bias $\boldsymbol{b}_n$: $\mathbf{D}_n^{[1]}\widehat{\boldsymbol{h}}^{\mathrm{LR}}$ unorthodox “input"
- η_n : $\mathrm{prox}_{\lambda\|\cdot\|_1} = \mathrm{HardTanh}_\lambda$

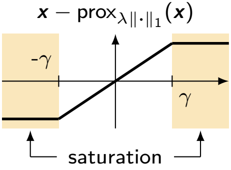


(Le et al., *EUSIPCO*, 2022)



Neural network from unrolling of FISTA: Deep FISTA Hloc (DFH)

- linear operator $\mathbf{W}_n$: $\mathbf{I} - \mathbf{D}_n^{[1]}\mathbf{D}_n^{[2]}$:
  - decouple filters $\mathbf{D}_n^{[1]}$ and $\mathbf{D}_n^{[2]}$ in a given layer;
  - decouple filters $\mathbf{D}_n^{[\cdot]}$ at different layer;
  - γ_n implicitly learn in $\mathbf{D}_n^{[1]}$; inertia parameter τ_n;
- bias $\boldsymbol{b}_n$: $\mathbf{D}_n^{[1]}\widehat{\boldsymbol{h}}^{\mathrm{LR}}$ unorthodox "input"
- η_n : $\mathrm{prox}_{\lambda\|\cdot\|_1} = \mathrm{HardTanh}_\lambda$



(Le et al., *EUSIPCO*, 2022)

✓ interpretable operations

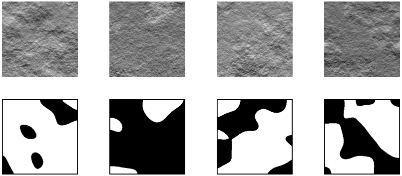
✓ $\boldsymbol{y} \mapsto NN(\boldsymbol{y};\vartheta)$ L_ϑ-lipschitzien (Combettes et al., *SIAM J. Math. Data Science*, 2020)

✓ versatile: similarly for Chambolle-Pock ⇒ Deep Strong-convexity Hloc (DSH)

End-to-end learning

textured image $\boldsymbol{y}$

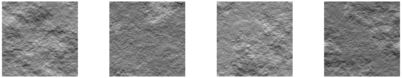
underlying segmentation $\boldsymbol{x}$



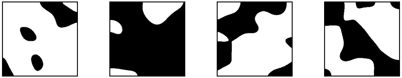
Synthetic textures: design of the database

End-to-end learning

textured image $\boldsymbol{y}$

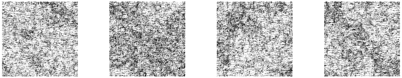


underlying segmentation $\boldsymbol{x}$



Fractal texture segmentation

linear regression $\boldsymbol{h}^{\mathrm{LR}}$

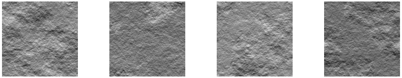


true regularity $\overline{\boldsymbol{h}}$

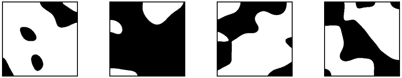
Synthetic textures: design of the database

End-to-end learning

textured image $\boldsymbol{y}$

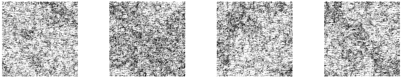


underlying segmentation $\boldsymbol{x}$



Fractal texture segmentation

linear regression $\boldsymbol{h}^{\mathrm{LR}}$

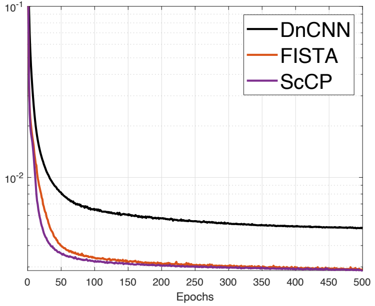


true regularity $\overline{\boldsymbol{h}}$

Config. I: $H_1 = 0.5$, $H_2 = 0.8$ **easy** Config. II: $H_1 = 0.5$, $H_2 = 0.65$ **difficult**

- training on 2000 images in Config. I
- implementation and training with PyTorch
- minimization of the quadratic empirical loss with ADAM

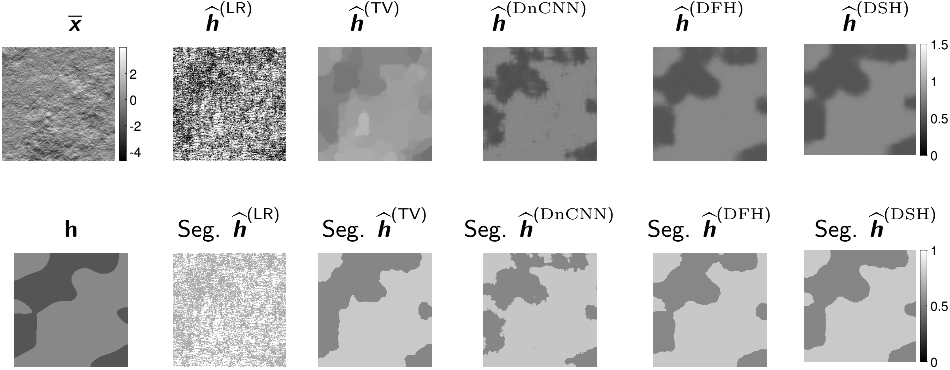
  500 epochs; learning rate 10^{-4}



training loss — testing loss

$\Longrightarrow$ more stable during training; faster training; possible to reduce the database size?

Variational approach vs.unrolled algorithms



Two complexities

- $\mathbf{5.10^3}$: $N_{\max} = 13$ layers; $\mathbf{D}_n^{[\cdot]}$ containing $|\mathbb{F}| = 21$ filters
- $\mathbf{3.10^4}$: $N_{\max} = 45$ layers; $\mathbf{D}_n^{[\cdot]}$ containing $|\mathbb{F}| = 37$ filters

State-of-the-art denoising neural network **DnCNN**

- $\mathbf{5.10^3} \Leftrightarrow 9$ layers of DnCNN;
- $\mathbf{3.10^4} \Leftrightarrow 10$ layers of DnCNN;

Training on Config. I; testing on 100 images in Config. I and 100 images in Config. II

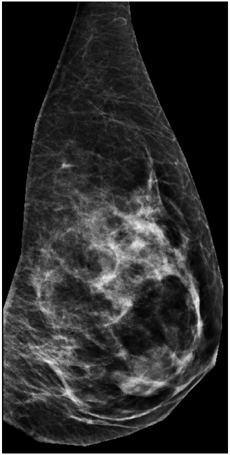
	TV-Stein	DnCNN 5.10^3	DFH 5.10^3	DFH 3.10^4	DSH 5.10^3	DSH 3.10^4
Estimation error						
Config. I	0.339±-0.048	0.113±0.011	0.073±0.008	0.069±0.007	0.072±0.007	0.069±0.007
Config. II	0.306±0.029	0.145±0.014	0.115±0.011	0.120±0.012	0.116±0.012	0.119±0.012
Segmentation score						
Config. I	83.7±2.68	81.0±1.73	94.8±0.45	95.2±0.41	94.8±0.42	95.3±0.41
Config. II	73.2±.51	68.4±1.36	70.7±2.65	70.9±2.67	68.3±2.74	70.6±2.71

Application to breast cancer detection

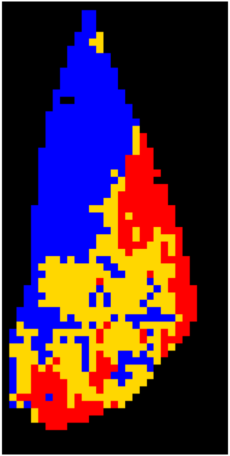
Observation: local regularity reflects health of breast tissues

(Marin et al., *Medical Physics*, 2017)

Mammogram



Local regularity (WTMM)



Total Variation (TV-Stein)