Traitement des données massives et apprentissage
Ecole d'été, 5-9 juin 2023

Traitements d'images satellitaires : amélioration d'images et extraction d'information, de la modélisation à l'apprentissage avec illustrations en imagerie SAR

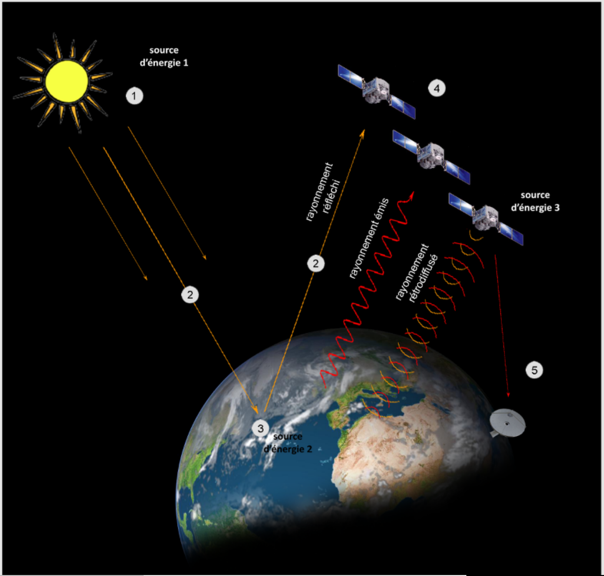
Florence Tupin (LTCI, Telecom Paris, Institut Polytechnique de Paris)
Emanuele Dalsasso (Cedric, CNAM, Paris)

June, 2023

Outline

Outline

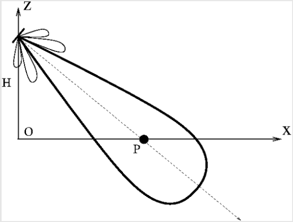
Remote Sensing systems (active and passive)

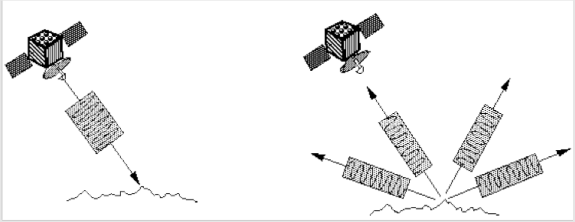


Principle of radar imaging

Radar imaging:

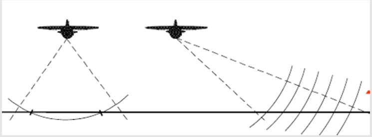
- Emission of an electro-magnetic wave by an antenna (GHz frequencies, cm wavelengths)
- Recording of the signal backscattered by the ground by the antenna





Charateristics:

- Lateral viewing
- Mono-static sensor



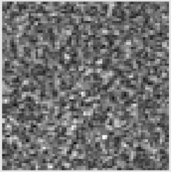
Introduction to SAR systems

Advantages

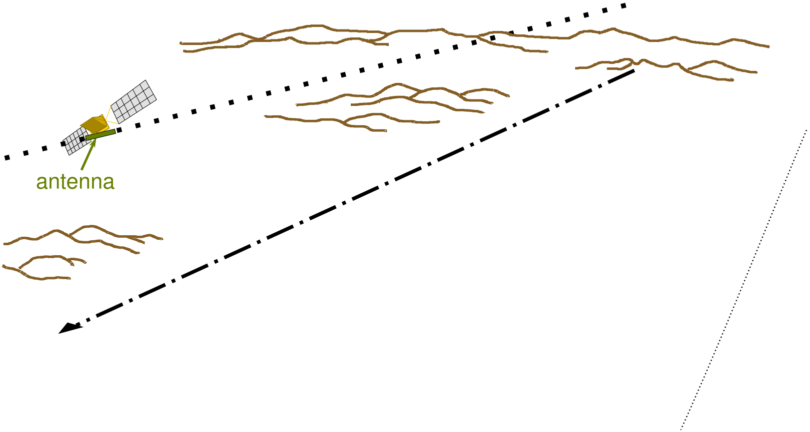
- All time sensor (own source of illumination)
- All weather sensor (electro-magnetic wave penetrating the clouds)
- Phase information linked to the distance from sensor to target ($\phi_{geom} = \frac{4\pi R}{\lambda}$)
- Complementary information of the optical images

Difficulties

- Speckle (strong radiometric fluctuations)
- Sensitivity to geometric distorsions (lateral viewing and high incidence angle)

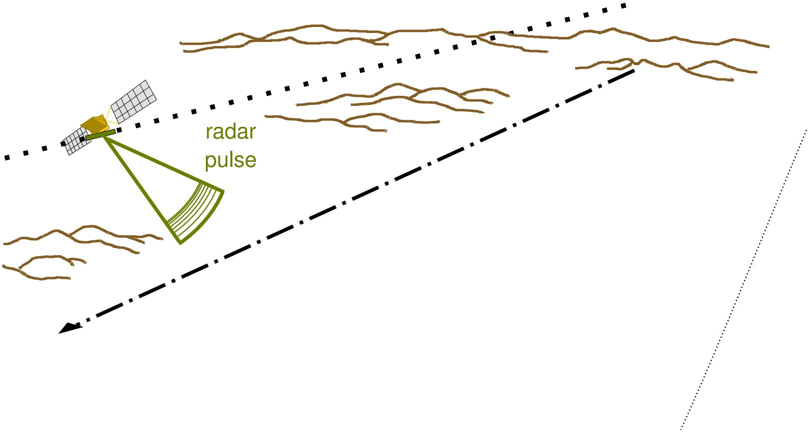


Principles of SAR image acquisition



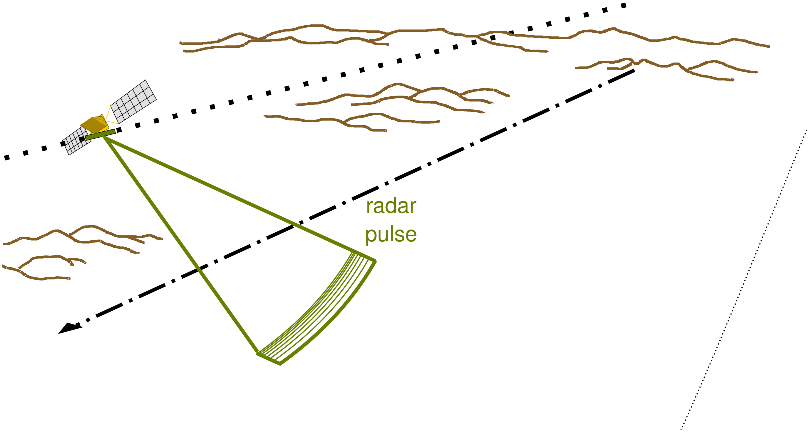
SAR imaging is an active imaging technique. . .

Principles of SAR image acquisition



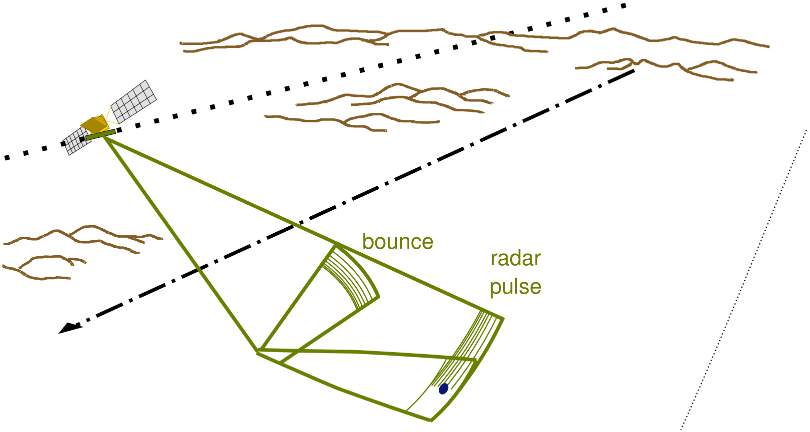
. . . based on the emission of an electromagnetic wave (typ. 10GHz).

Principles of SAR image acquisition



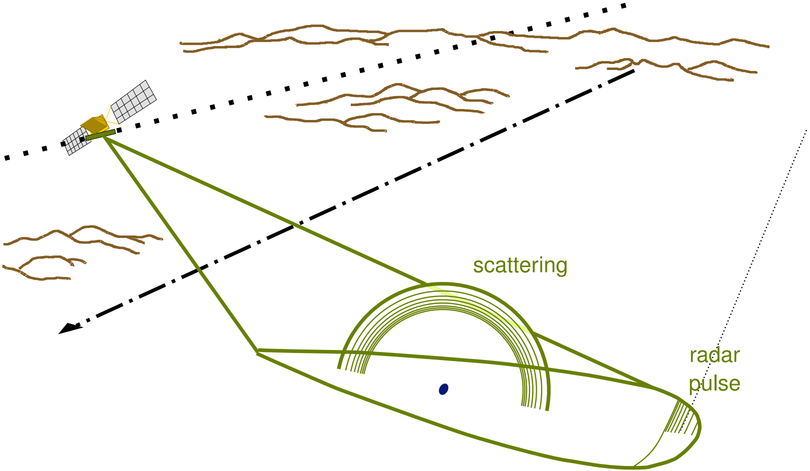
. . . based on the emission of an electromagnetic wave (typ. 10GHz).

Principles of SAR image acquisition



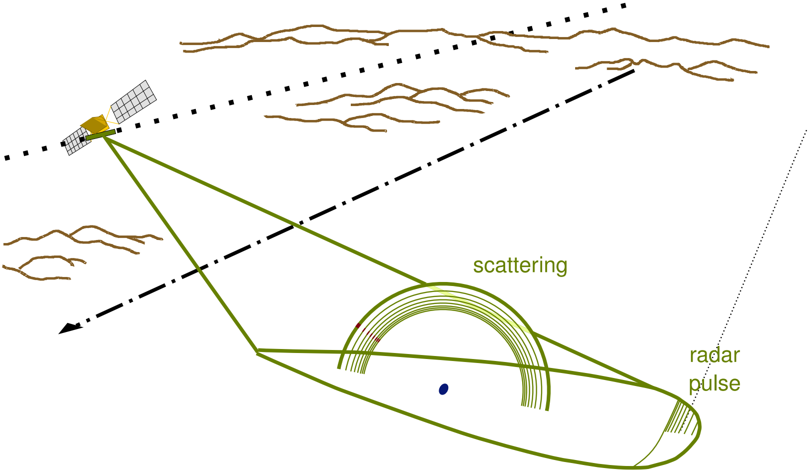
Depending on the scene geometry, the radar pulse is reflected...

Principles of SAR image acquisition



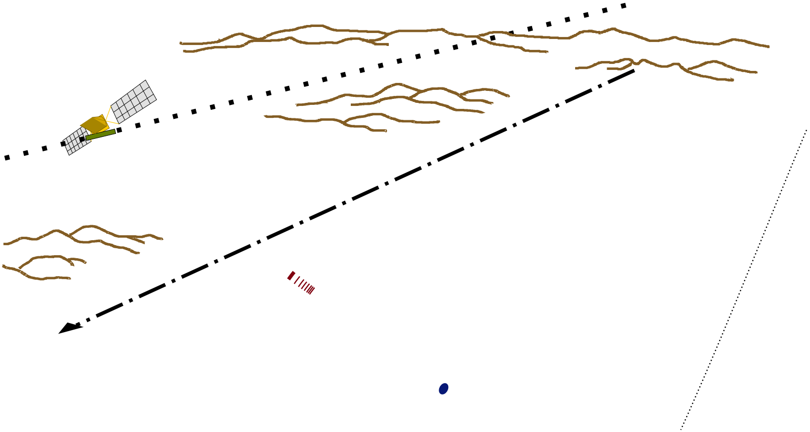
. . . or scattered. . .

Principles of SAR image acquisition



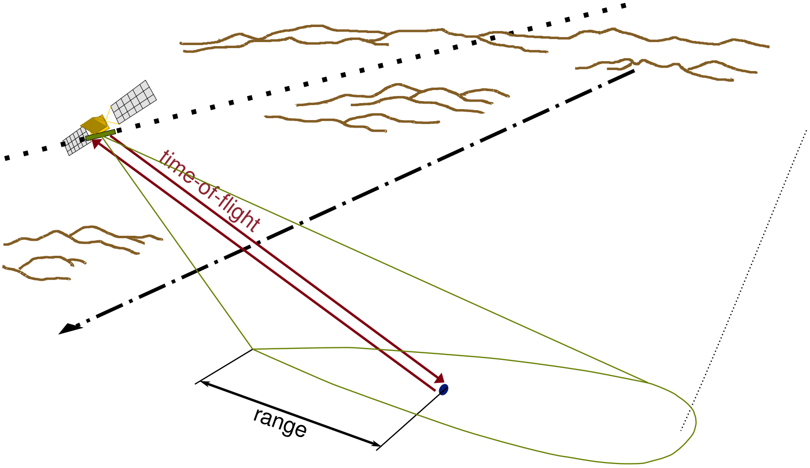
. . . and part of the incident energy is sent back to the antenna.

Principles of SAR image acquisition



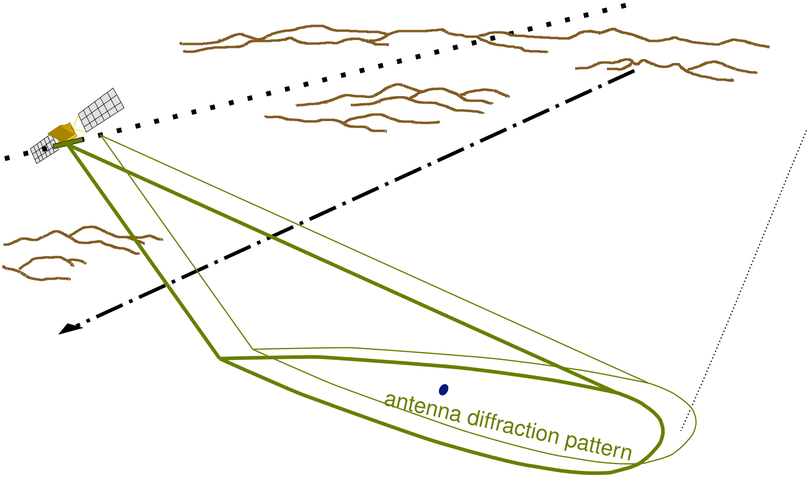
. . . and part of the incident energy is sent back to the antenna.

Principles of SAR image acquisition



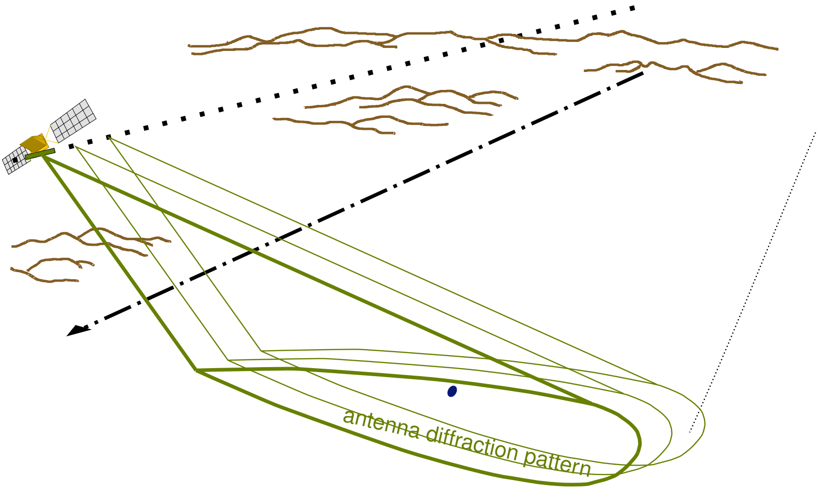
The location on the ground of the scatterer is deduced from the time-of-flight.

Principles of SAR image acquisition



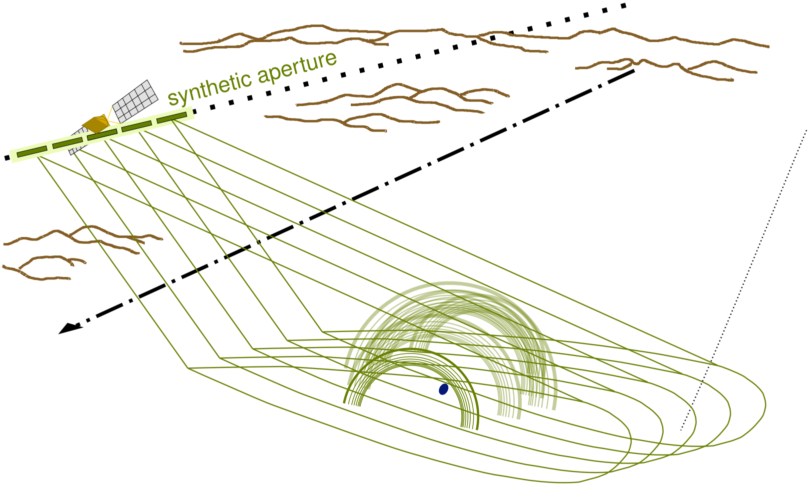
As the satellite moves, the antenna diffraction pattern covers another area. . .

Principles of SAR image acquisition

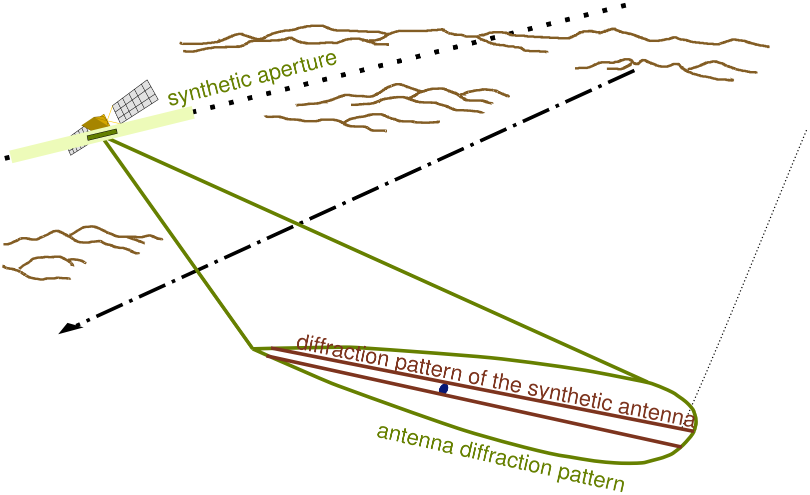


. . . thereby forming a 2D image.

Principles of SAR image acquisition

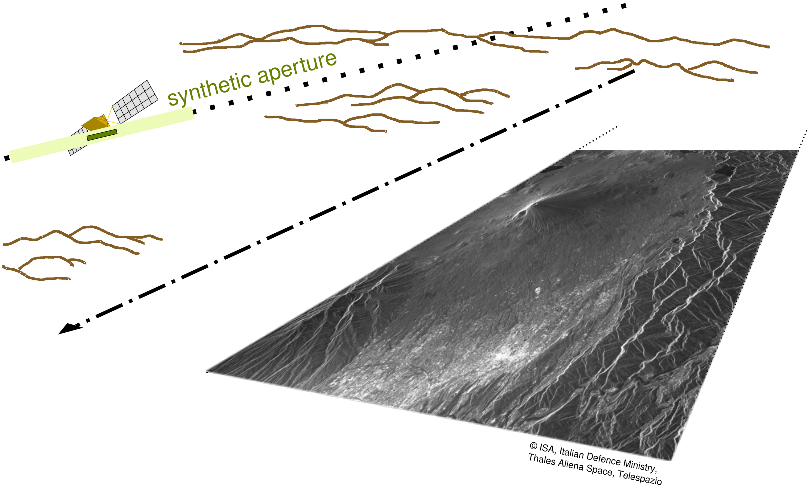


Aperture synthesis consists of numerically combining the echoes. . .



. . . which greatly improves the resolution.

Principles of SAR image acquisition



SAR data

Different kinds of "images"

- single SAR image

amplitude → **object classification**, ...

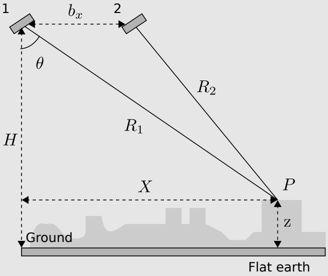
- interferometry: 2 SAR images

phase difference → **elevation**, ...

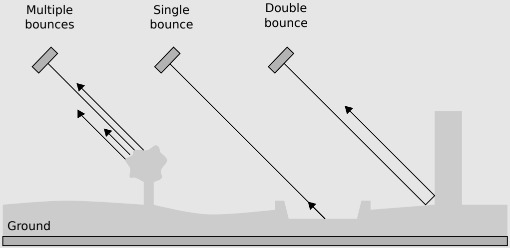
- polarimetry: 3 SAR channels

complex correlation → **geophysical properties**

SAR, InSAR, PolSAR, PolInSAR



(a) InSAR



(b) PolSAR

SAR data

Different kinds of "images"

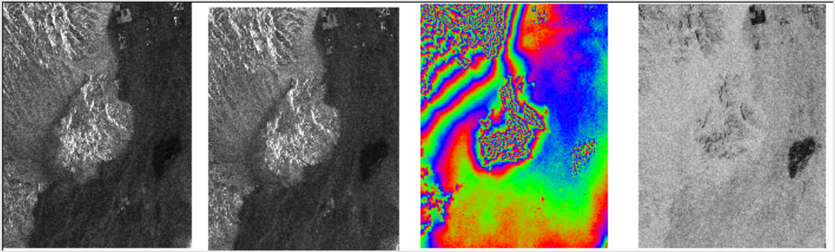
- single SAR image

  amplitude → **object classification**, ...
- interferometry: 2 SAR images

  phase difference → **elevation**, ...
- polarimetry: 3 SAR channels

  complex correlation → **geophysical properties**

SAR, InSAR, PolSAR, PolInSAR



Different kinds of "images"

- single SAR image

amplitude → **object classification**, ...

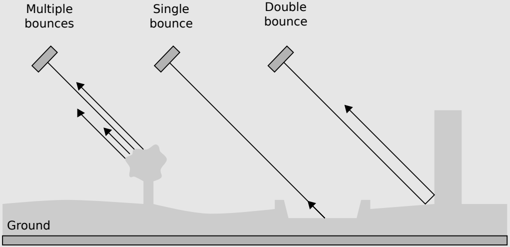
- interferometry: 2 SAR images

phase difference → **elevation**, ...

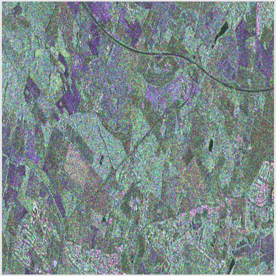
- polarimetry: 3 SAR channels

complex correlation → **geophysical properties**

SAR, InSAR, PolSAR, PolInSAR



(a) PolSAR mechanisms



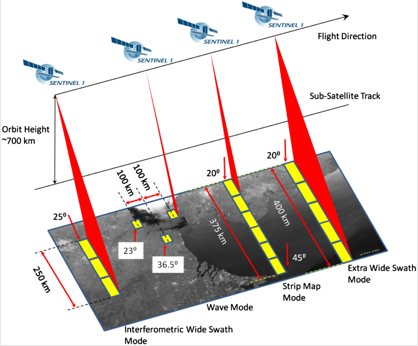
(b) PolSAR image (©RCM-CSA)

Principle, advantages and limits

- Acquisition of many spectral bands
- Provide an information on the spectral signature of the objects on the ground
- Fine characterization of the surface
- Resolution and spectral mixing
- Very high dimension

Spatial resolution

- Wide range of spatial resolutions (from sub-metric to deca-metric resolutions)
- Resolution linked with the spectral or polarimetric diversity
- For a same sensor different modes leading to different resolutions (Example for SAR data: Spotlight, Stripmap, Extra Wide Swath)



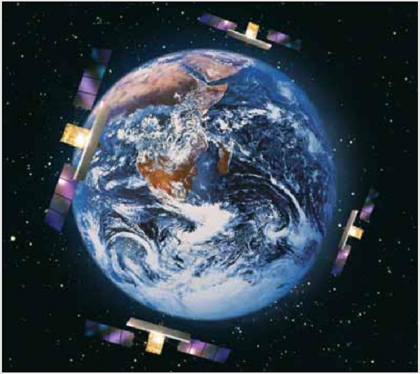
Resolutions - spatial, spectral, temporal

Spectral resolution

- Optic and spectral images (ex: Sentinel-2)
- Hyperspectral images
- Polarimetric SAR images (RadarSat, BIOMASS, ...)

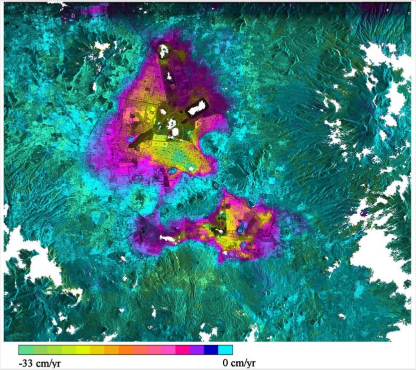
Temporal resolution

- Revisit time of a satellite depends on the orbital parameters (a few days)
- It can be reduced using the agility of the satellite to modify the incidence angle
- It can be reduced thanks to constellations



Multi-temporal series or heterogeneous acquisitions

- SITS (Satellite Image Time Series) ex: Sentinel-1, Sentinel-2
- Applications : DEM (stereo, interferometry), ground movement monitoring (differential interferometry)
- Heterogeneous sensors :
  - registration (image processing methods, exploitation of sensor parameters and acquisition geometry)
  - use of georeferenced images (projection in a common referential, need a DEM) ; ex Sentinel-2 2A product are ortho-rectified tiles

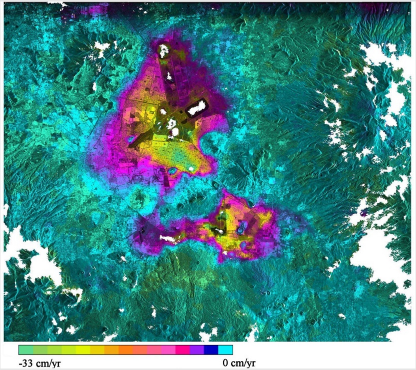


extrait de la thèse de P. Lopes-Quiroz

Combining images

Softwares for processing and visualization

- Different levels of products (geometric and radiometric corrections) provided by the space agencies or provided by specific softwares (SNAP, etc.)
- Commercial softwares doing the processing based on the meta-data associated to the the images or using image processing tools
- Example of Google Earth Engine :
    - allows the manipulation of SITS
    - pre-processing of the data may lose part of the information



extrait de la thèse de P. Lopes-Quiroz

Outline

1. Introduction to satellite imaging
2. Image restoration
   - Remote sensing image restoration
   - Model based methods
   - Deep learning methods
3. Information extraction
   - Issues
   - Probabilistic classification methods
   - Deep learning methods
4. Practical work

Outline

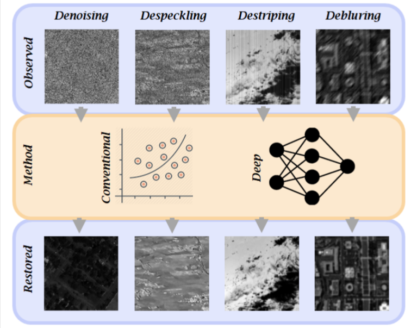
Image degradation

Data are degraded by a variety of noise types and artefacts

Major sources of degradation

- atmospheric perturbation (absorption, scattering, reflection of the solar radiation in the atmosphere -mostly for passive systems)
- imaging systems (speckle for SAR coherent imaging system, striping for multi or hyper-spectral pushbrooming systems)
- instrumental noises (thermal noise, quantization noise, shot noise)

General model of image degradation:
$H = M.X + S$
where M can be a multiplicative noise (SAR data) or the blurring effect of the instrument)

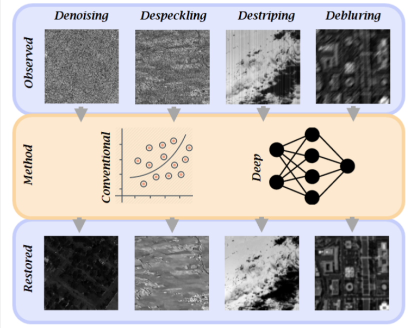


B. Rasti et al., "*Image Restoration for Remote Sensing: Overview and Toolbox," IEEE Geoscience and Remote Sensing Magazine*", 2022.

Image restoration problems

Wide range of tasks

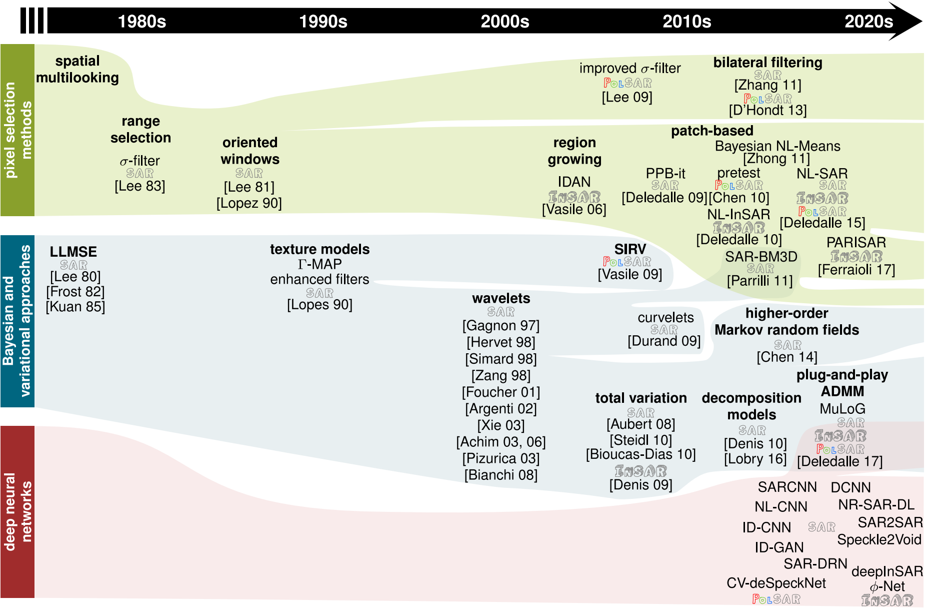
- Denoising: speckle reduction for SAR systems, noise reduction for multi/ hypersctral data
- Deblurring and destriping for multi/ hyper-spectral images
- Pan-sharpening and super-resolution for optical systems



B. Rasti et al., "*Image Restoration for Remote Sensing: Overview and Toolbox," IEEE Geoscience and Remote Sensing Magazine*", 2022.

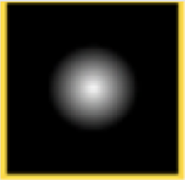
Outline

Example of speckle reduction techniques: an overview of 40+ years of research



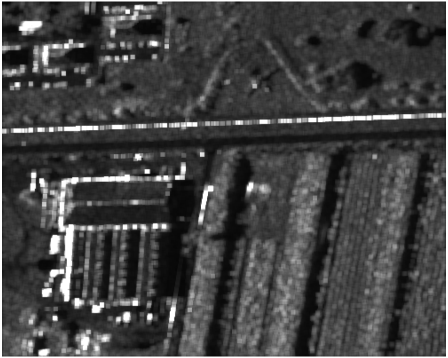
Pixel selection methods

- Local filters: they are based on an image model of local smoothness (neighboring pixels are similar)
  $\Rightarrow$ mean filter (multi-looking)
  $\Rightarrow$ gaussian filter



Example of local filters for speckle reduction

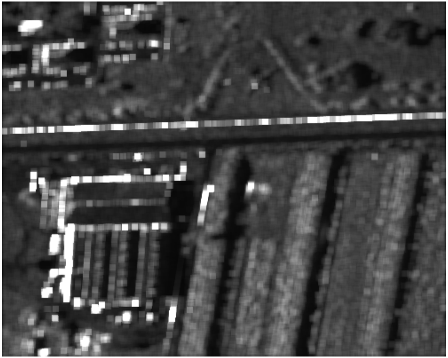
noisy ©ONERA



denoised (5×5 window)

Example of local filters for speckle reduction

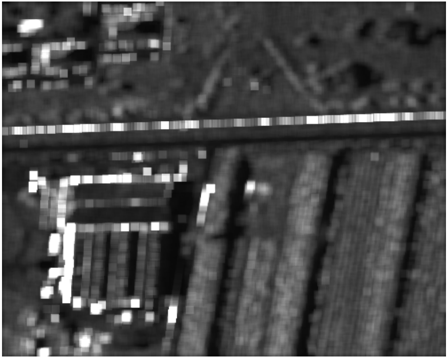
noisy ©ONERA



denoised (7×7 window)

Example of local filters for speckle reduction

noisy ©ONERA



denoised (9×9 window)

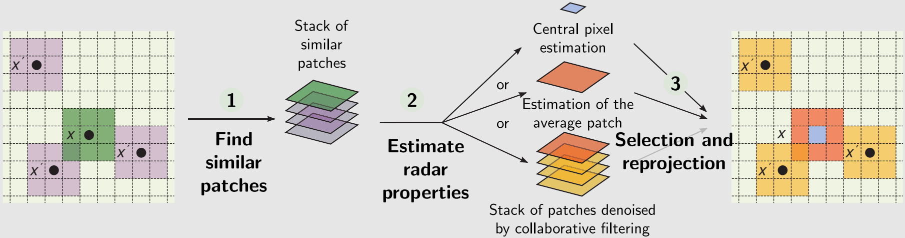
simple and fast

strong resolution loss!

⇝ still in use in several works (e.g. forest/agriculture applications)

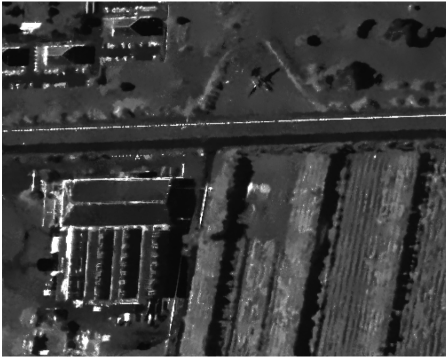
Pixel selection methods

- Non-local filters (patch based filters): they are based on an image model of local redundancy or self-similarity (small patches of the image are similar)



Example of patch-based filter for speckle reduction


noisy ©ONERA


denoised (NL-SAR)

preserves the spatial resolution
also works on polarimetric/interferometric data

some residual speckle may be visible
some low-contrast textures are smoothed

Bayesian and variational approaches

Key idea: minimization of a loss function combining data-fidelity (given by the acquisition system) and regularity (given by a prior model of images)

$$\arg\min_{\mathbf{x}} \underbrace{-\log p_{\mathbf{y}}(\mathbf{y}|\mathbf{x})}_{\text{data-fidelity term}} + \underbrace{\mathscr{R}(\mathbf{x})}_{\text{regularization}}$$

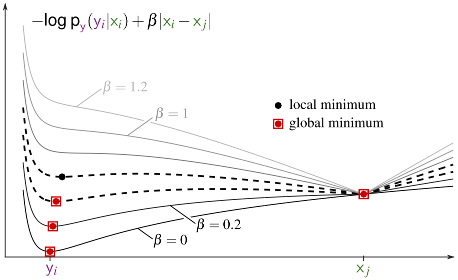
Key idea: minimization of a loss function combining data-fidelity (given by the acquisition system) and regularity (given by a prior model of images)

$$\arg\min_{\mathbf{x}} \underbrace{-\log p_{\mathbf{y}}(\mathbf{y}|\mathbf{x})}_{\text{data-fidelity term}} + \underbrace{\mathscr{R}(\mathbf{x})}_{\text{regularization}}$$

In linear scale: Rayleigh distribution (non-convex)

Total variation:

$$\mathscr{R}(\mathbf{x}) = \beta \sum_i \sqrt{(\mathbf{D}_{\mathsf{h}}\mathbf{x})_i^2 + (\mathbf{D}_{\mathsf{v}}\mathbf{x})_i^2}$$



Bayesian and variational approaches

Key idea: minimization of a loss function combining data-fidelity (given by the acquisition system) and regularity (given by a prior model of images)

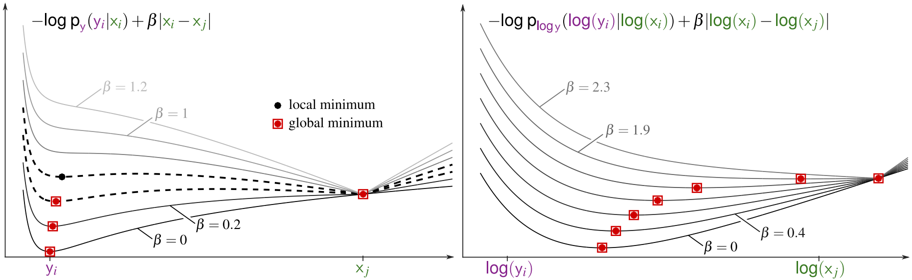
$$\arg\min_{\mathbf{x}} \underbrace{-\log p_{\mathbf{y}}(\mathbf{y}|\mathbf{x})}_{\text{data-fidelity term}} + \underbrace{\mathscr{R}(\mathbf{x})}_{\text{regularization}}$$

In linear scale: Rayleigh distribution (non-convex)

In logarithmic scale: Fisher-Tippett distribution (convex)

Total variation:

$$\mathscr{R}(\mathbf{x}) = \beta \sum_i \sqrt{(\mathbf{D}_h\mathbf{x})_i^2 + (\mathbf{D}_v\mathbf{x})_i^2}$$



Bayesian and variational approaches

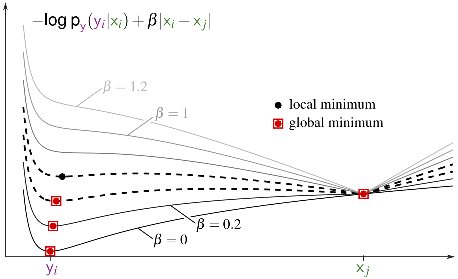
Key idea: minimization of a loss function combining data-fidelity (given by the acquisition system) and regularity (given by a prior model of images)

$$\arg\min_{\mathbf{x}} \underbrace{-\log p_{\mathbf{y}}(\mathbf{y}|\mathbf{x})}_{\text{data-fidelity term}} + \underbrace{\mathscr{R}(\mathbf{x})}_{\text{regularization}}$$

In linear scale: Rayleigh distribution (non-convex)

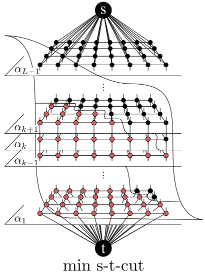
Total variation:

$$\mathscr{R}(\mathbf{x}) = \beta \sum_i \sqrt{(\mathbf{D}_{\mathsf{h}}\mathbf{x})_i^2 + (\mathbf{D}_{\mathsf{v}}\mathbf{x})_i^2}$$



$\rightsquigarrow$

non-convex optimization by graph-cuts:



Bayesian and variational approaches

Key idea: minimization of a loss function combining data-fidelity (given by the acquisition system) and regularity (given by a prior model of images)

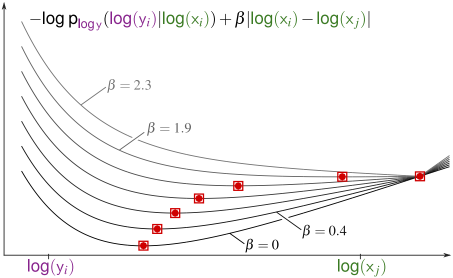
$$\arg\min_{\mathbf{x}} \underbrace{-\log p_{\mathbf{y}}(\mathbf{y}|\mathbf{x})}_{\text{data-fidelity term}} + \underbrace{\mathscr{R}(\mathbf{x})}_{\text{regularization}}$$

In logarithmic scale: Fisher-Tippett distribution (convex)

Total variation:

$$\mathscr{R}(\mathbf{x}) = \beta \sum_i \sqrt{(\mathbf{D}_{\mathsf{h}}\mathbf{x})_i^2 + (\mathbf{D}_{\mathsf{v}}\mathbf{x})_i^2}$$

non-smooth convex optimization: e.g. ADMM $\leadsto$



Variational: total variation minimization [Bioucas-Dias *et al.* IEEE trans. Image Proc. 2010]

Key idea: minimization of a loss function combining data-fidelity and regularity (low total-variation)

noisy ©ONERA

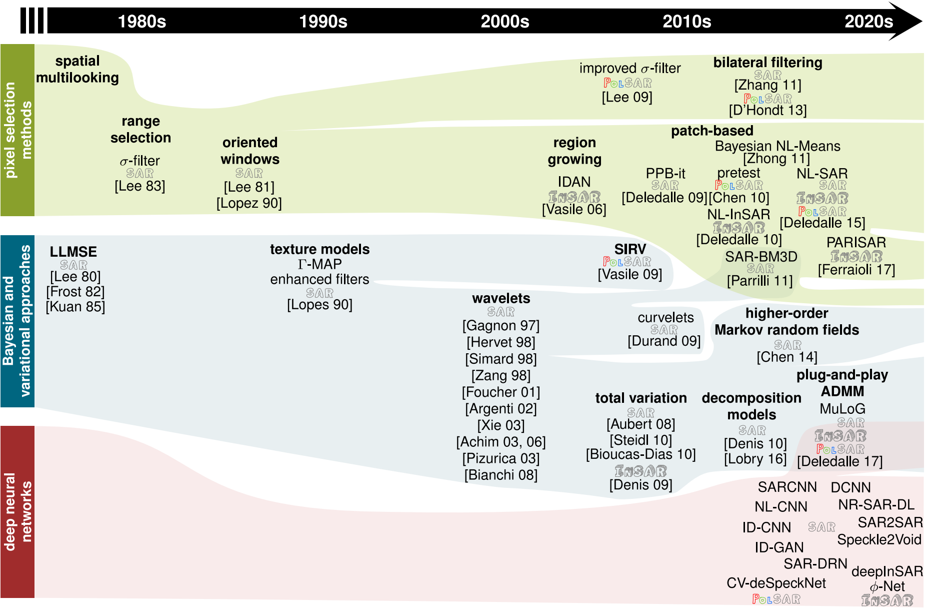
denoised (TV minimization)

sharp edges

restored image is piecewise constant
↳ staircase artifact in slowly varying areas

Outline

Example of speckle reduction techniques: an overview of 40+ years of research



Family of methods

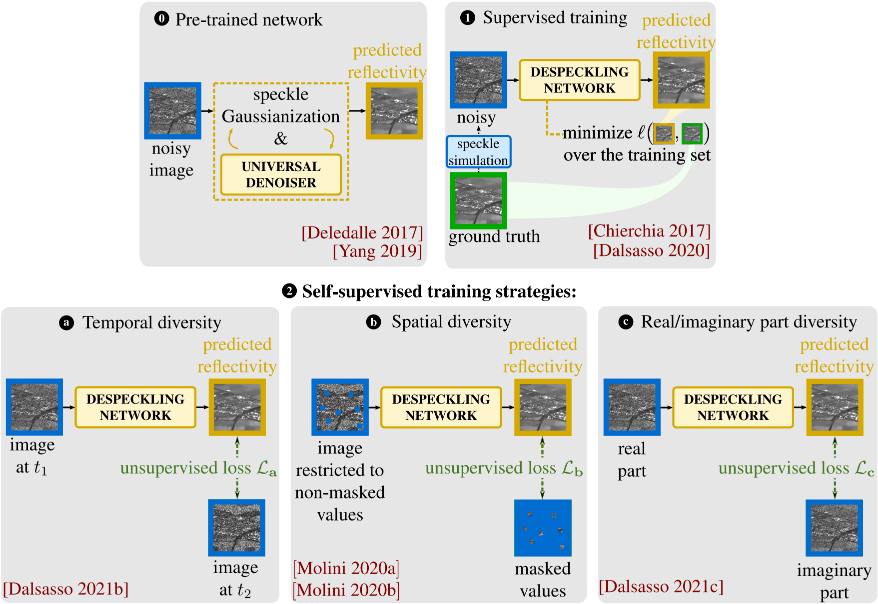
- Plug-and-Play methods: they are based on an optimization framework calling iteratively a gaussian denoiser which can be trained in a deep learning fashion
- Supervised approaches: they are based on the availability of ground truth images (pairs of noisy and ground-truth data)
- Unsupervised approaches: they are based only on noisy data with different strategies: The main idea is to split the data in two parts : one given as input to the network to provide a prediction and the other one used to define the loss to evaluate the prediction.

  The 2 parts can be defined using:
  - temporal diversity (using two -noisy- acquisitions of the same area);
  - spatial diversity (only the neighborhood of a pixel is given to the network to make the perdiction);
  - "data" diversity for instance by splitting real and imaginary parts in case of complex data.

Key ingredients of the network

- The network architecture: the choice of the network architecture has a strong influence on the performances (loss, generalization,...) and the training time
  Many choices are possible : U-net, residual connections, ... (see course of Emanuele)
- The loss definition: it should be adapted to the input and output of the network. Depending of the input / output pair L2 loss or losses inspired from the neg-log-likelihood can be used.

A broad overview of deep-learning strategies (example of speckle reduction)



Pre-trained network in Plug and Play approaches

plug-and-play ADMM: Key idea: use off-the-shelf Gaussian denoiser + apply regularization in log-domain

$$\arg\min_{\mathbf{x}} \underbrace{-\log p_{\mathbf{y}}(\mathbf{y}|\mathbf{x})}_{\text{data-fidelity term}} + \underbrace{\mathscr{R}(\mathbf{x})}_{\text{regularization}}$$

Pre-trained network in Plug and Play approaches

plug-and-play ADMM: Key idea: use off-the-shelf Gaussian denoiser + apply regularization in log-domain

$$\arg\min_{\mathbf{x}} \underbrace{-\log p_{\mathbf{y}}(\mathbf{y}|\mathbf{x})}_{\text{data-fidelity term}} + \underbrace{\mathscr{R}(\mathbf{x})}_{\text{regularization}}$$

$$\Big\downarrow \text{ variable splitting}$$

$$\arg\min_{\mathbf{x},\,\mathbf{z}} \underbrace{-\log p_{\mathbf{y}}(\mathbf{y}|\mathbf{x})}_{\text{data-fidelity term}} + \underbrace{\mathscr{R}(\mathbf{z})}_{\text{regularization}} \qquad \text{subject to} \quad \mathbf{z} = \mathbf{x}$$

Pre-trained network in Plug and Play approaches

plug-and-play ADMM: Key idea: use off-the-shelf Gaussian denoiser + apply regularization in log-domain

$$\arg\min_{\mathbf{x}} \underbrace{-\log p_{\mathbf{y}}(\mathbf{y}|\mathbf{x})}_{\text{data-fidelity term}} + \underbrace{\mathscr{R}(\mathbf{x})}_{\text{regularization}}$$

$$\Big\downarrow \text{ variable splitting}$$

$$\arg\min_{\mathbf{x},\mathbf{z}} \underbrace{-\log p_{\mathbf{y}}(\mathbf{y}|\mathbf{x})}_{\text{data-fidelity term}} + \underbrace{\mathscr{R}(\mathbf{z})}_{\text{regularization}} \quad \text{subject to} \quad \mathbf{z} = \mathbf{x}$$

$$\Big\downarrow \text{ Alternating Direction Method of Multipliers (ADMM)}$$

$$\mathbf{z} \leftarrow \arg\min_{\mathbf{z}} \underbrace{\frac{\beta}{2}\|\mathbf{z}-\mathbf{x}+\mathbf{d}\|_2^2}_{\text{quadratic term}} + \underbrace{\mathscr{R}(\mathbf{z})}_{\text{regularization}}$$

$$\mathbf{d} \leftarrow \mathbf{d} + \mathbf{z} - \mathbf{x}$$

$$\mathbf{x} \leftarrow \arg\min_{\mathbf{x}} \underbrace{\frac{\beta}{2}\|\mathbf{z}-\mathbf{x}+\mathbf{d}\|_2^2}_{\text{quadratic term}} - \underbrace{\log p_{\mathbf{y}}(\mathbf{y}|\mathbf{x})}_{\text{data-fidelity term}}$$

Pre-trained network in Plug and Play approaches

plug-and-play ADMM: Key idea: use off-the-shelf Gaussian denoiser + apply regularization in log-domain

$$\arg\min_{\mathbf{x}} \underbrace{-\log p_{\mathbf{y}}(\mathbf{y}|\mathbf{x})}_{\text{data-fidelity term}} + \underbrace{\mathscr{R}(\mathbf{x})}_{\text{regularization}}$$

$\downarrow$ variable splitting

$$\arg\min_{\mathbf{x},\mathbf{z}} \underbrace{-\log p_{\mathbf{y}}(\mathbf{y}|\mathbf{x})}_{\text{data-fidelity term}} + \underbrace{\mathscr{R}(\mathbf{z})}_{\text{regularization}} \quad \text{subject to} \quad \mathbf{z}=\mathbf{x}$$

$\downarrow$ Alternating Direction Method of Multipliers (ADMM)

plug-in ADMM

$$\mathbf{z} \leftarrow \arg\min_{\mathbf{z}} \underbrace{\frac{\beta}{2}\|\mathbf{z}-\mathbf{x}+\mathbf{d}\|_2^2}_{\text{quadratic term}} + \underbrace{\mathscr{R}(\mathbf{z})}_{\text{regularization}}$$

$\rightarrow$ state-of-the-art Gaussian denoiser

$$\mathbf{d} \leftarrow \mathbf{d}+\mathbf{z}-\mathbf{x}$$

$$\mathbf{x} \leftarrow \arg\min_{\mathbf{x}} \underbrace{\frac{\beta}{2}\|\mathbf{z}-\mathbf{x}+\mathbf{d}\|_2^2}_{\text{quadratic term}} - \underbrace{\log p_{\mathbf{y}}(\mathbf{y}|\mathbf{x})}_{\text{data-fidelity term}}$$

Pre-trained network in Plug and Play approaches

plug-and-play ADMM: Key idea: use off-the-shelf Gaussian denoiser + apply regularization in log-domain

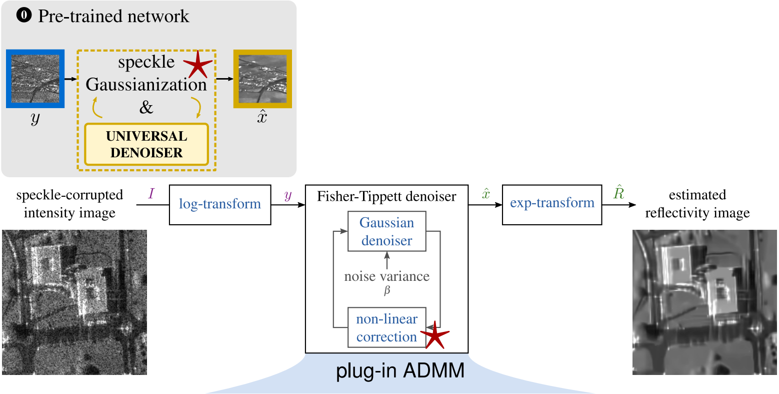
noisy ©ONERA

denoised (MuLoG)

high-quality results
meta-algorithm: can be used with many different denoising algorithms
↳ results can be compared to rule out algorithm-specific artifacts

sensitive to spatial correlations in the speckle fluctuations.

Applying a pre-trained network (universal Gaussian denoiser)



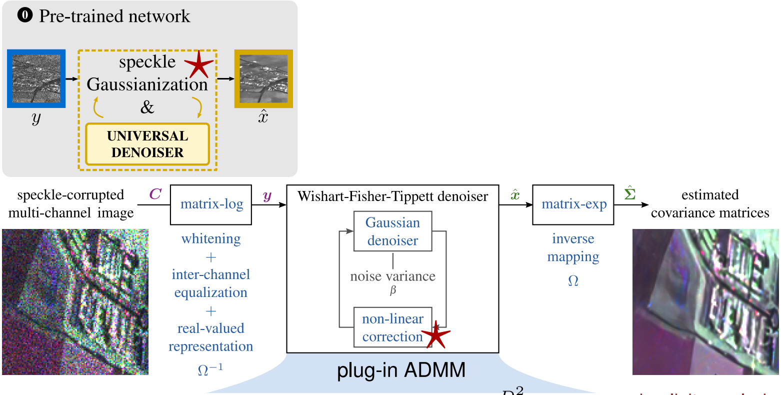
$$\hat{\boldsymbol{z}} \leftarrow \underset{\boldsymbol{z}}{\operatorname{argmin}} \; \frac{\beta}{2}\|\boldsymbol{z} - \hat{\boldsymbol{x}} + \hat{\boldsymbol{d}}\|^2 + \quad \mathcal{R}(z^c) \; ,$$

implicit regularization: CNN trained on natural images

$$\hat{\boldsymbol{d}} \leftarrow \hat{\boldsymbol{d}} + \hat{\boldsymbol{z}} - \hat{\boldsymbol{x}} \; ,$$

$$\hat{\boldsymbol{x}} \leftarrow \underset{\boldsymbol{x}}{\operatorname{argmin}} \; \frac{\beta}{2}\|\boldsymbol{x} - \hat{\boldsymbol{z}} - \hat{\boldsymbol{d}}\|^2 - \log p_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{x}) \; .$$

Applying a pre-trained network (universal Gaussian denoiser)



$$\hat{\boldsymbol{z}} \leftarrow \underset{\boldsymbol{z}}{\operatorname{argmin}} \frac{\beta}{2}\|\boldsymbol{z} - \hat{\boldsymbol{x}} + \hat{\boldsymbol{d}}\|^2 + \sum_{c=1}^{D^2} \mathcal{R}(z^c)\,,$$

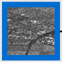
implicit regularization: CNN trained on natural images

$$\hat{\boldsymbol{d}} \leftarrow \hat{\boldsymbol{d}} + \hat{\boldsymbol{z}} - \hat{\boldsymbol{x}}\,,$$

$$\hat{\boldsymbol{x}} \leftarrow \underset{\boldsymbol{x}}{\operatorname{argmin}} \frac{\beta}{2}\|\boldsymbol{x} - \hat{\boldsymbol{z}} - \hat{\boldsymbol{d}}\|^2 - \log p_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{x})\,.$$

Applying a pre-trained network (universal Gaussian denoiser)

0 Pre-trained network

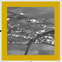


y

speckle Gaussianization

&

UNIVERSAL DENOISER



$\hat{x}$

requires no training!
generalizes to multi-channel images

network not refined for SAR scenes.

speckle-corrupted multi-channel image

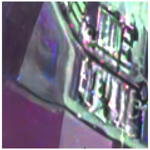
$\boldsymbol{C}$ → matrix-log → $\boldsymbol{y}$ → Wishart-Fisher-Tippett denoiser → $\hat{\boldsymbol{x}}$ → matrix-exp → $\hat{\boldsymbol{\Sigma}}$

matrix-log: whitening + inter-channel equalization + real-valued representation Ω^{-1}

Wishart-Fisher-Tippett denoiser: Gaussian denoiser — noise variance β — non-linear correction

matrix-exp: inverse mapping Ω

estimated covariance matrices



plug-in ADMM

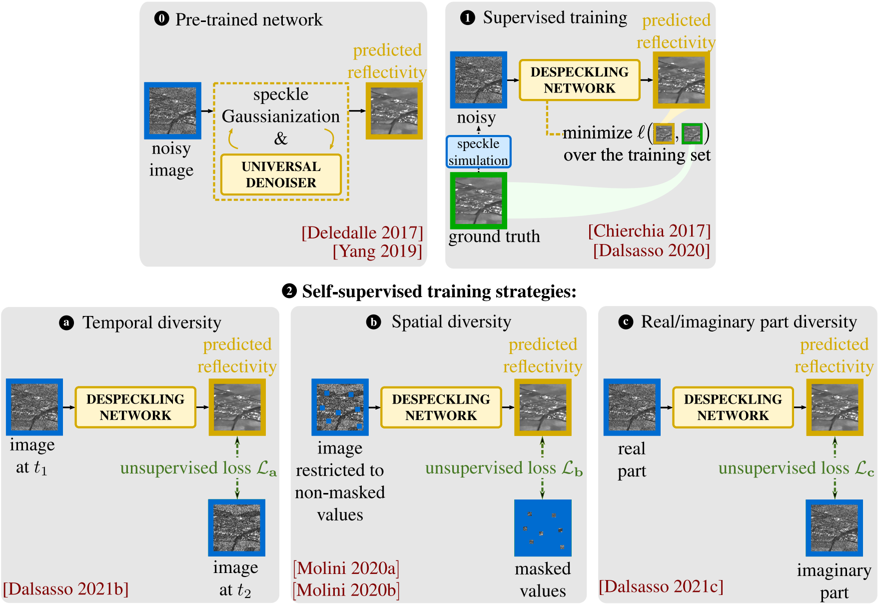
$$\hat{\boldsymbol{z}} \leftarrow \operatorname*{argmin}_{\boldsymbol{z}} \frac{\beta}{2}\|\boldsymbol{z} - \hat{\boldsymbol{x}} + \hat{\boldsymbol{d}}\|^2 + \sum_{c=1}^{D^2} \mathcal{R}(\boldsymbol{z}^c)\,,$$

$$\hat{\boldsymbol{d}} \leftarrow \hat{\boldsymbol{d}} + \hat{\boldsymbol{z}} - \hat{\boldsymbol{x}}\,,$$

$$\hat{\boldsymbol{x}} \leftarrow \operatorname*{argmin}_{\boldsymbol{x}} \frac{\beta}{2}\|\boldsymbol{x} - \hat{\boldsymbol{z}} - \hat{\boldsymbol{d}}\|^2 - \log p_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{x})\,.$$

implicit regularization: CNN trained on natural images

A broad overview of deep-learning strategies (example of speckle reduction)



Supervised training for denoising

Principle

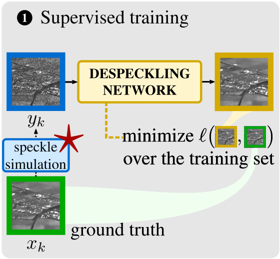
- It supposes that some training pairs are available $\{(\boldsymbol{y}_k, \boldsymbol{x}_k^\star)\}_{k=1..K}$ forming the training set ($\boldsymbol{y}_k$ noisy image, $\boldsymbol{x}_k^\star$ the ground truth image)
- The parameters of the network are optimized to minimize :

$$\arg\min_{\boldsymbol{\theta}} \sum_{k=1}^{K} \mathcal{L}\left[f_{\boldsymbol{\theta}}(\boldsymbol{y}_k), \boldsymbol{x}_k^\star\right], \quad (1)$$

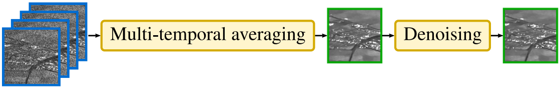
  where L can be a quadratic L2 loss or L1 loss (more robust to outliers)

- The training pairs can be defined using different strategies:
  - using natural images (photos) and a degradation model (synthetic noise, blurring kernel)
  - using "improved" images (higher resolution, longer exposure time, multi-temporal filtering...) and a degradation model

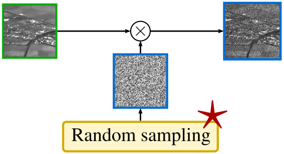
Supervised training of a despeckling network



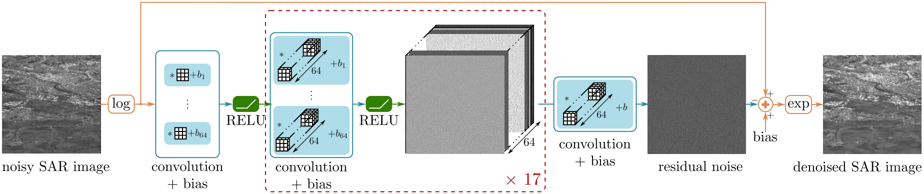
Ground-truth generation:



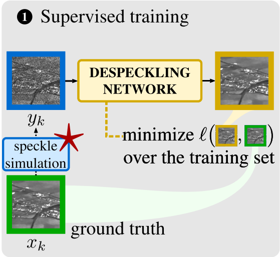
Speckle simulation:



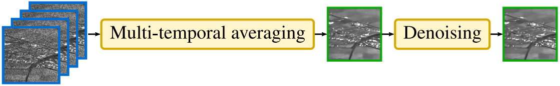
Residual CNN:



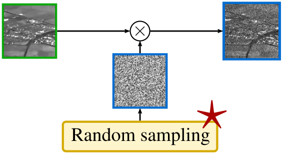
Supervised training of a despeckling network



Ground-truth generation:



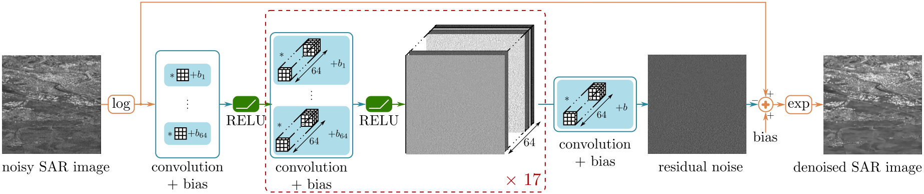
Speckle simulation:



time series provide high-quality ground truth.

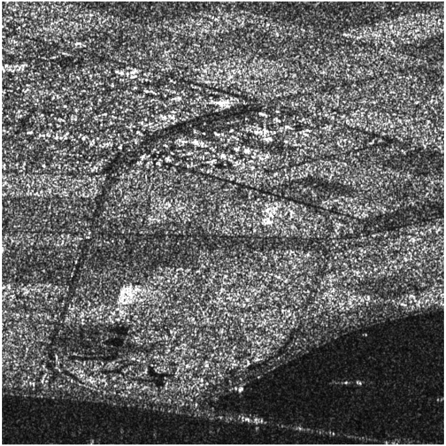
if speckle correlations are ignored in the simulator, requires some downsampling ⇒ resolution loss

Residual CNN:

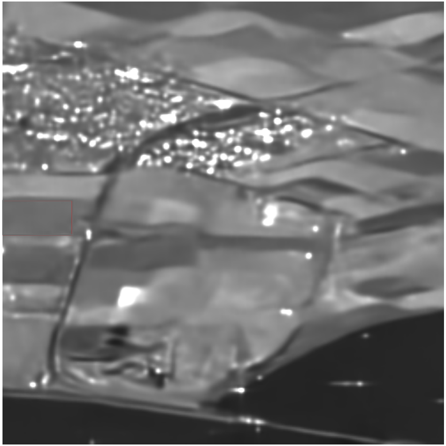


Supervised training of a despeckling network

Restoration results with SARCNN: Sentinel-1 SLC IW image (©ESA, images have been pre-processed to reduce sidelobes and limit speckle correlation)

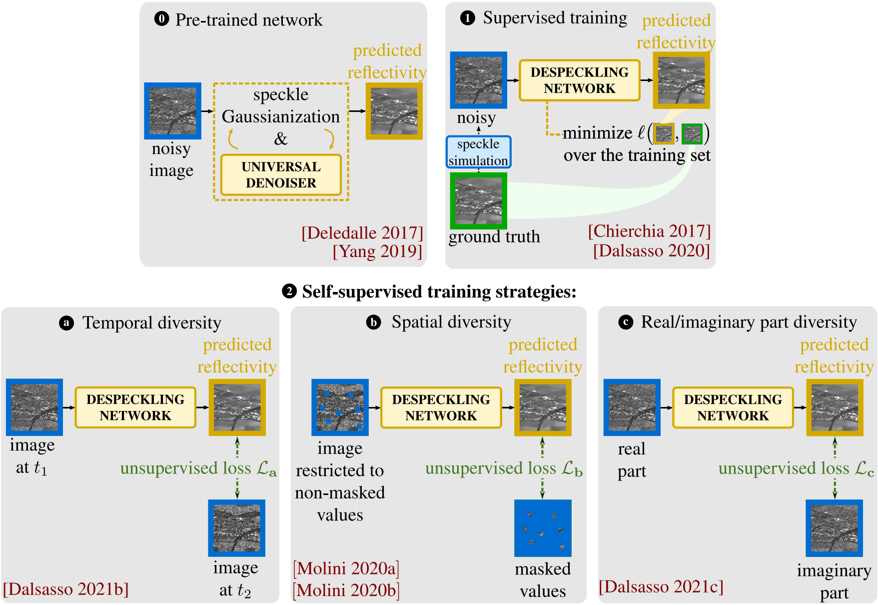


Single-look Sentinel-1 image



Restored image

A broad overview of deep-learning strategies for despeckling



Semi-supervised training with noisy pairs

Noise2noise principle

- Noise2noise: the main idea is that instead of using a pair of noisy (input) and ground-truth (loss definition) images, the network output is the same when trained with noisy / noisy pairs $((\boldsymbol{y}_k^{\text{in}}), \boldsymbol{y}_k^{\text{val}})$ under certain conditions:

$$\arg\min_{\boldsymbol{\theta}} \sum_{k=1}^{K} \mathcal{L}\left[f_{\boldsymbol{\theta}}(\boldsymbol{y}_k), \boldsymbol{x}_k^\star\right] = \arg\min_{\boldsymbol{\theta}} \sum_{k=1}^{K} \mathcal{L}\left[f_{\boldsymbol{\theta}}(\boldsymbol{y}_k^{\text{in}}), \boldsymbol{y}_k^{\text{val}}\right]. \tag{2}$$

- Conditions and choice of the loss:
  For the training to work, images $\boldsymbol{y}_k^{\text{in}}$ and $\boldsymbol{y}_k^{\text{val}}$ must correspond to a common (yet unknown) ground truth image $\boldsymbol{x}_k^\star$ and also be statistically independent. Under these conditions, a natural choice for the loss function $\mathcal{L}$ is the neg-log-likelihood:

$$\mathcal{L}\left[f_{\boldsymbol{\theta}}(\boldsymbol{y}_k^{\text{in}}), \boldsymbol{y}_k^{\text{val}}\right] = -\log \mathsf{p}_Y\left(\boldsymbol{y}_k^{\text{val}} | \boldsymbol{x}_k^\star = f_{\boldsymbol{\theta}}(\boldsymbol{y}_k^{\text{in}})\right). \tag{3}$$

J. Lehtinen et al., "*Noise2Noise: Learning Image Restoration without Clean Data*", 2018.

Self-supervised SAR despeckling: loss definition

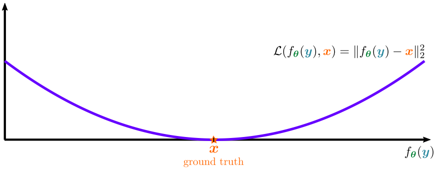
Supervised loss for AWGN

$$\hat{\boldsymbol{\theta}}_{\text{supervised}} \in \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \ \mathbb{E}_X \Big\{ \mathbb{E}_{Y|X} \big[\mathcal{L}\big(f_{\boldsymbol{\theta}}(\boldsymbol{y}_i), \boldsymbol{x}_i\big)\big]\Big\}$$

network parameters

noisy input

ground truth



J. Lehtinen et al., "*Noise2Noise: Learning Image Restoration without Clean Data*", 2018.

Self-supervised SAR despeckling: loss definition

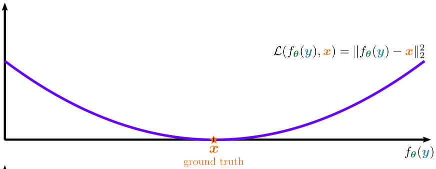
Supervised loss for AWGN

$$\hat{\boldsymbol{\theta}}_{\text{supervised}} \in \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \; \mathbb{E}_X \Big\{ \mathbb{E}_{Y|X} \big[\mathcal{L}\big(f_{\boldsymbol{\theta}}(\boldsymbol{y}_i), \boldsymbol{x}_i \big) \big] \Big\}$$

network parameters

noisy input

ground truth

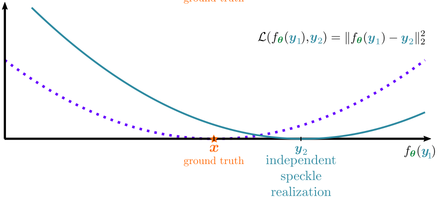


Unsupervised loss for AWGN

$$\hat{\boldsymbol{\theta}}^{(\mathcal{L}_2)}_{\text{self-supervised}} \in \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \; \mathbb{E}_X \Big\{ \mathbb{E}_{Y|X} \big[\| f_{\boldsymbol{\theta}}(\boldsymbol{y}_1) - \boldsymbol{y}_2 \|^2 \big] \Big\}$$

noisy input

independent noise realization

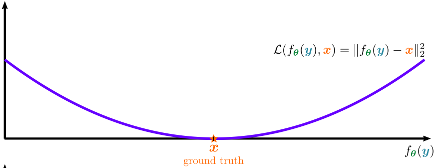


J. Lehtinen et al., "*Noise2Noise: Learning Image Restoration without Clean Data*", 2018.

Supervised loss for AWGN

$$\hat{\boldsymbol{\theta}}_{\text{supervised}} \in \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \ \mathbb{E}_X \left\{ \mathbb{E}_{Y|X} \left[\mathcal{L}\left(f_{\boldsymbol{\theta}}(\boldsymbol{y}_i), \boldsymbol{x}_i \right) \right] \right\}$$

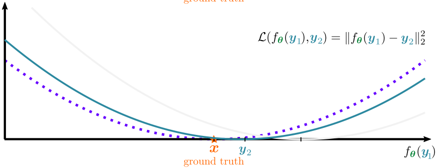
network parameters — noisy input — ground truth



Unsupervised loss for AWGN

$$\hat{\boldsymbol{\theta}}^{(\mathcal{L}_2)}_{\text{self-supervised}} \in \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \ \mathbb{E}_X \left\{ \mathbb{E}_{Y|X} \left[\| f_{\boldsymbol{\theta}}(\boldsymbol{y}_1) - \boldsymbol{y}_2 \|^2 \right] \right\}$$

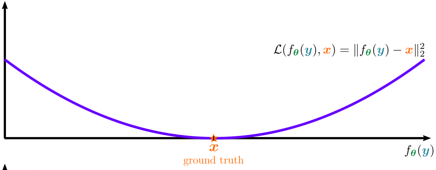
noisy input — independent noise realization



J. Lehtinen et al., "*Noise2Noise: Learning Image Restoration without Clean Data*", 2018.

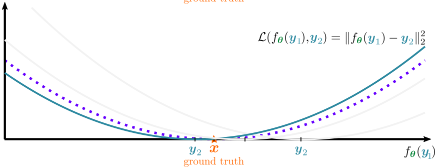
Supervised loss for AWGN

$$\underbrace{\hat{\boldsymbol{\theta}}_{\text{supervised}}}_{\text{network parameters}} \in \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \; \mathbb{E}_X\Big\{ \mathbb{E}_{Y|X}\big[\mathcal{L}\big(f_{\boldsymbol{\theta}}(\underbrace{\boldsymbol{y}_i}_{\text{noisy input}}), \underbrace{\boldsymbol{x}_i}_{\text{ground truth}}\big)\big]\Big\}$$



Unsupervised loss for AWGN

$$\hat{\boldsymbol{\theta}}^{(\mathcal{L}_2)}_{\text{self-supervised}} \in \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \; \mathbb{E}_X\Big\{\mathbb{E}_{Y|X}\big[\|f_{\boldsymbol{\theta}}(\underbrace{\boldsymbol{y}_1}_{\text{noisy input}}) - \underbrace{\boldsymbol{y}_2}_{\text{independent noise realization}}\|^2\big]\Big\}$$

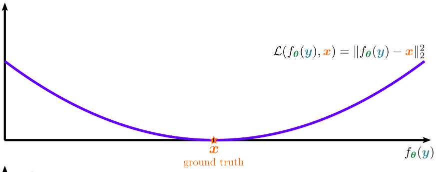


J. Lehtinen et al., "*Noise2Noise: Learning Image Restoration without Clean Data*", 2018.

Self-supervised SAR despeckling: loss definition

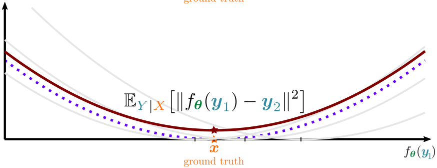
Supervised loss for AWGN

$$\underbrace{\hat{\boldsymbol{\theta}}_{\text{supervised}}}_{\text{network parameters}} \in \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \ \mathbb{E}_X \Big\{ \mathbb{E}_{Y|X} \big[\mathcal{L}\big(f_{\boldsymbol{\theta}}(\underbrace{\boldsymbol{y}_i}_{\text{noisy input}}), \underbrace{\boldsymbol{x}_i}_{\text{ground truth}} \big) \big] \Big\}$$



Unsupervised loss for AWGN

$$\hat{\boldsymbol{\theta}}^{(\mathcal{L}_2)}_{\text{self-supervised}} \in \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \ \mathbb{E}_X \Big\{ \mathbb{E}_{Y|X} \big[\| f_{\boldsymbol{\theta}}(\underbrace{\boldsymbol{y}_1}_{\text{noisy input}}) - \underbrace{\boldsymbol{y}_2}_{\text{independent noise realization}} \|^2 \big] \Big\}$$

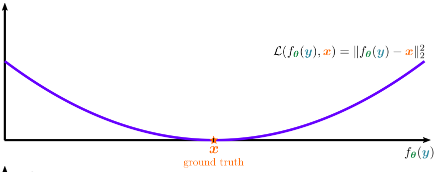


J. Lehtinen et al., "*Noise2Noise: Learning Image Restoration without Clean Data*", 2018.

Supervised loss for AWGN

$$\hat{\boldsymbol{\theta}}_{\text{supervised}} \in \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\ \mathbb{E}_X \Big\{ \mathbb{E}_{Y|X} \left[\mathcal{L}\big(f_{\boldsymbol{\theta}}(\boldsymbol{y}_i), \boldsymbol{x}_i\big) \right] \Big\}$$

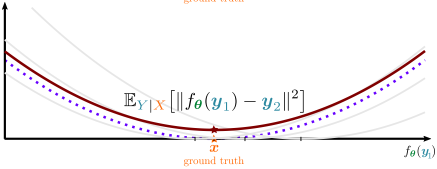
network parameters — noisy input — ground truth



Unsupervised loss for AWGN

$$\hat{\boldsymbol{\theta}}^{(\mathcal{L}_2)}_{\text{self-supervised}} \in \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\ \mathbb{E}_X \Big\{ \mathbb{E}_{Y|X} \left[\|f_{\boldsymbol{\theta}}(\boldsymbol{y}_1) - \boldsymbol{y}_2\|^2 \right] \Big\}$$

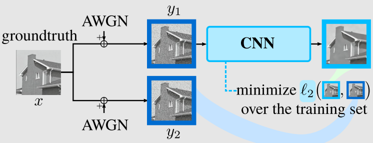
noisy input — independent noise realization



The noise2noise framework: training strategy

❶ Generate noisy pairs (y_1, y_2) from a groundtruth image

❷ Update network

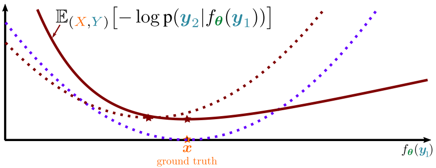


J. Lehtinen et al., "*Noise2Noise: Learning Image Restoration without Clean Data*", 2018.

Unsupervised loss for speckle

$$\hat{\boldsymbol{\theta}}^{(\text{lik})}_{\text{self-supervised}} \in \operatorname{argmin}_{\boldsymbol{\theta}} \; \mathbb{E}_{(X,Y)}\Big[\underbrace{-\log \mathrm{p}(\boldsymbol{y}_2 | f_{\boldsymbol{\theta}}(\boldsymbol{y}_1))}_{\text{Goodman's model of speckle}}\Big]$$

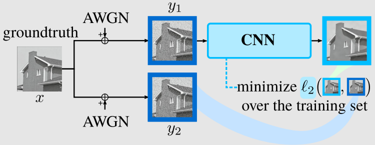
$$\mathrm{p}(\boldsymbol{y}|\boldsymbol{x}) = e^{y-x} \exp\left(-e^{y-x}\right)$$

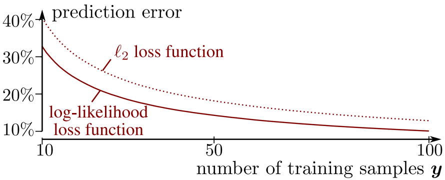


The noise2noise framework: training strategy

❶ Generate noisy pairs (y_1, y_2) from a groundtruth image

❷ Update network



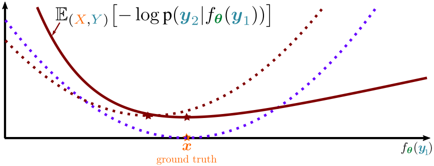


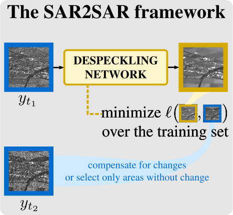
J. Lehtinen et al., "*Noise2Noise: Learning Image Restoration without Clean Data*", 2018.
E. Dalsasso et al., "*SAR2SAR: a semi-supervised despeckling algorithm for SAR images.*", 2021.

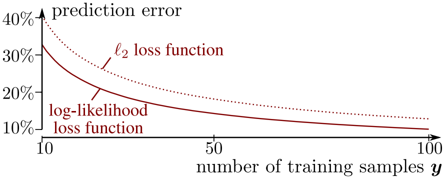
Unsupervised loss for speckle

$$\hat{\boldsymbol{\theta}}^{(\text{lik})}_{\text{self-supervised}} \in \operatorname{argmin}_{\boldsymbol{\theta}} \; \mathbb{E}_{(X,Y)}\Big[\underbrace{-\log \mathrm{p}(\boldsymbol{y}_2 | f_{\boldsymbol{\theta}}(\boldsymbol{y}_1))}_{\text{Goodman's model of speckle}}\Big]$$

$$\mathrm{p}(y|x) = e^{y-x} \exp\left(-e^{y-x}\right)$$

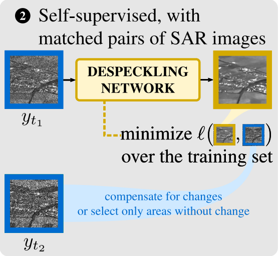




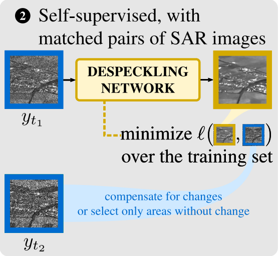


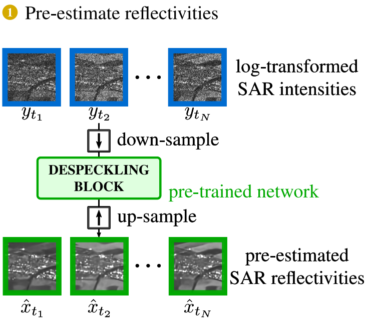
J. Lehtinen et al., "*Noise2Noise: Learning Image Restoration without Clean Data*", 2018.
E. Dalsasso et al., "*SAR2SAR: a semi-supervised despeckling algorithm for SAR images.*", 2021.

Semi-supervised method with matched pairs of SAR images

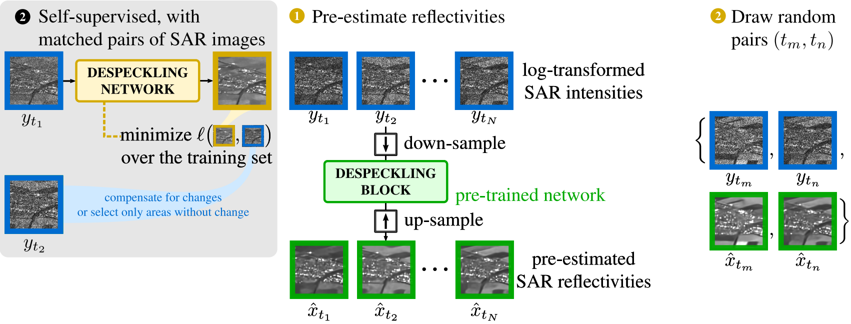


Semi-supervised method with matched pairs of SAR images

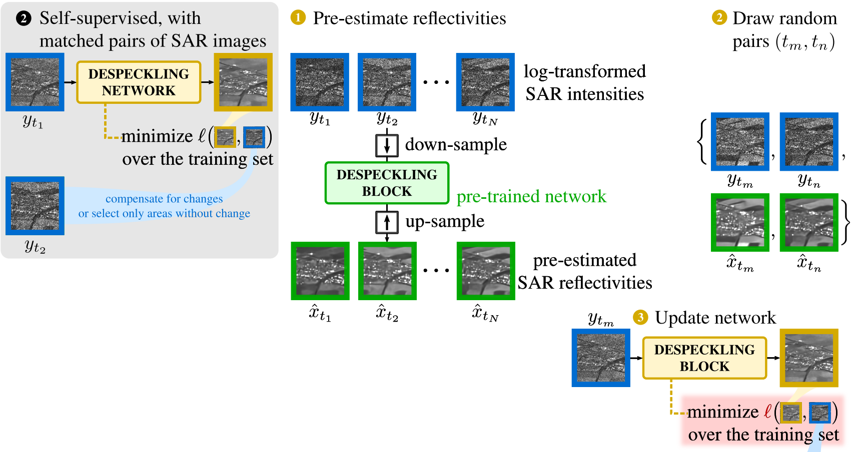




Semi-supervised method with matched pairs of SAR images

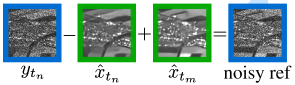


Semi-supervised method with matched pairs of SAR images

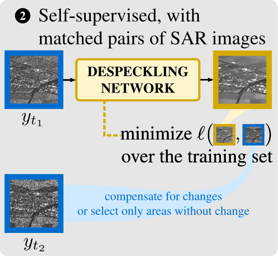


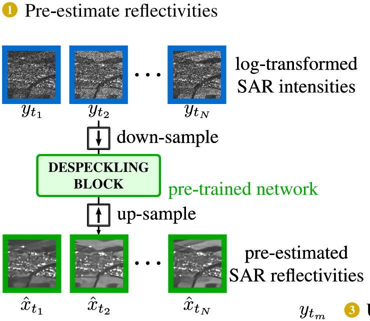
$$\hat{\theta}^{(\mathrm{lik})}_{\text{self-supervised}} \in \underset{\theta}{\arg\min}\ \mathbb{E}_X\left[\underbrace{-\log p(\boldsymbol{y}_2|f_{\boldsymbol{\theta}}(\boldsymbol{y}_1))}_{\ell}\right]$$

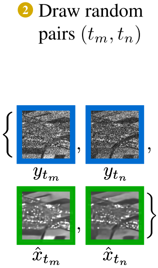
Goodman's speckle model: $\sum_k f_{\boldsymbol{\theta}}([\boldsymbol{y}_1]_k) - [\boldsymbol{y}_2]_k + \exp\left([\boldsymbol{y}_2]_k - f_{\boldsymbol{\theta}}([\boldsymbol{y}_1]_k)\right)$



Semi-supervised method with matched pairs of SAR images

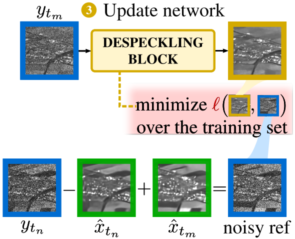






very good restoration performance

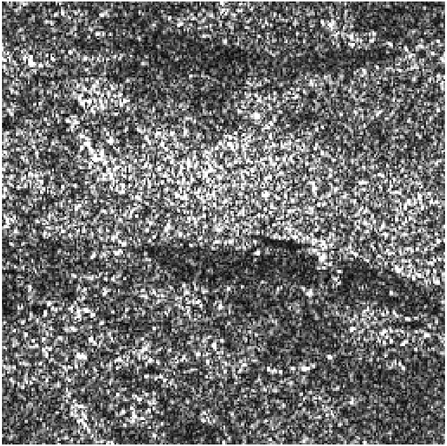
the training procedure is a little heavy



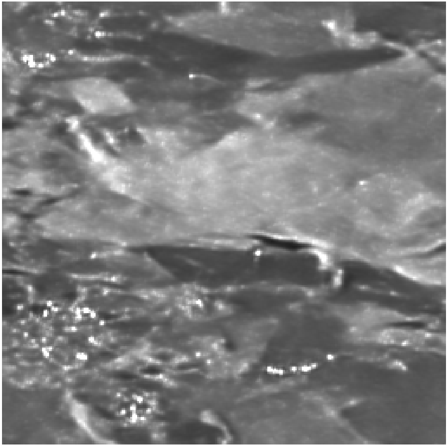
$$\hat{\theta}^{(\text{lik})}_{\text{self-supervised}} \in \underset{\theta}{\operatorname{argmin}}\ \mathbb{E}_X\left[-\log p(\boldsymbol{y}_2 | f_\theta(\boldsymbol{y}_1))\right] \quad \ell$$

Goodman's speckle model: $\displaystyle\sum_k f_\theta([\boldsymbol{y}_1]_k) - [\boldsymbol{y}_2]_k + \exp\left([\boldsymbol{y}_2]_k - f_\theta([\boldsymbol{y}_1]_k)\right)$

Semi-supervised method with matched pairs of SAR images



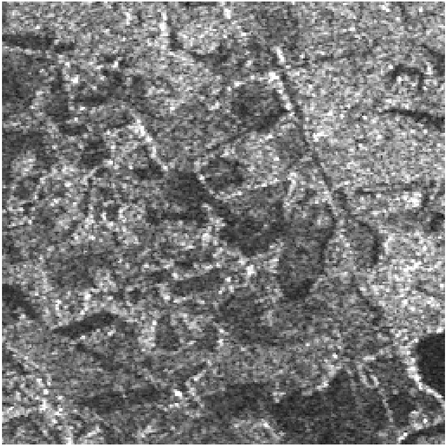
noisy ©ESA (Sentinel-1)



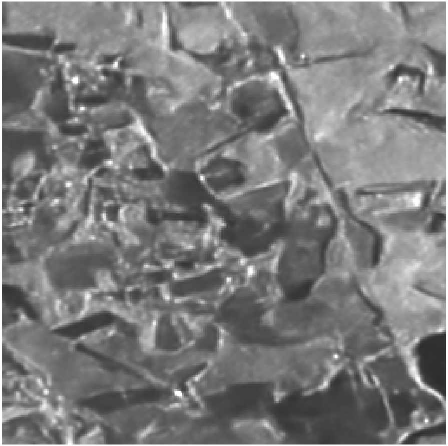
denoised (SAR2SAR)

E. Dalsasso et al. "*SAR2SAR: a semi-supervised despeckling algorithm for SAR images*", IEEE JSTARS, 2021.

Semi-supervised method with matched pairs of SAR images



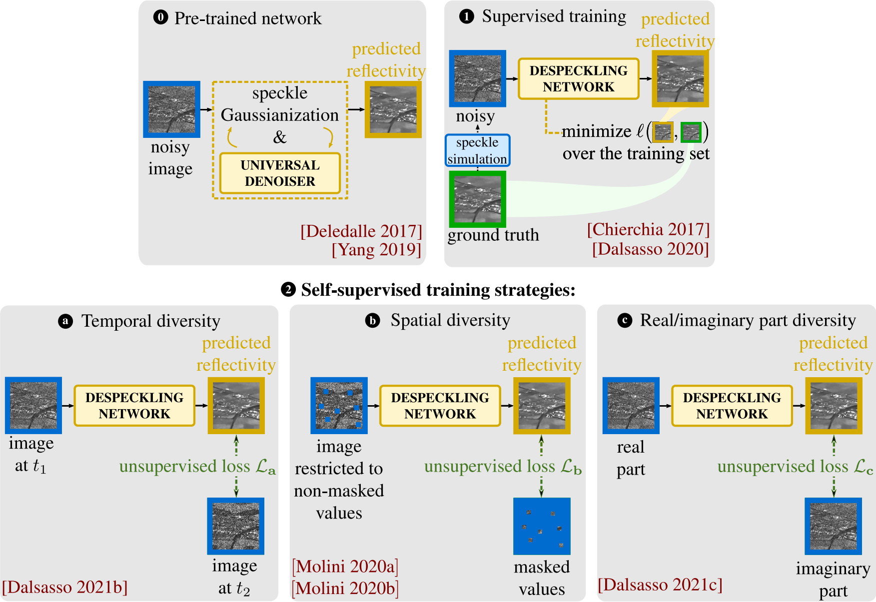
noisy ©ESA (Sentinel-1 GRD)



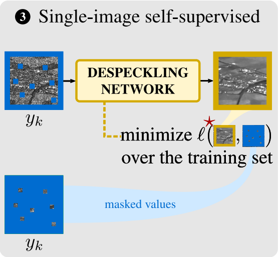
denoised (SAR2SAR-GRD)

N. Gasnier et al. "*Despeckling Sentinel-1 GRD images by deep learning and application to narrow river segmentation*", IGARSS, 2021.

A broad overview of deep-learning strategies for despeckling



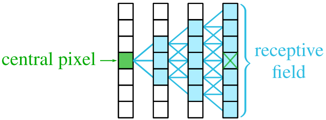
Self-supervised with a single image



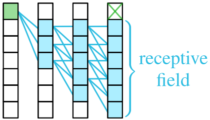
Main idea: train by cross-validation

Improvement 1: alternately mask out each pixel ⇝ dense validation

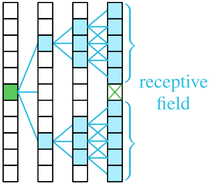
build network architecture to exclude the central pixel
from the receptive field



with conventional convolutions
the central pixel is at the center of
the receptive field



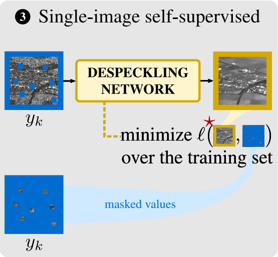
by shifting the convolution kernels
the central pixel is next to the
receptive field



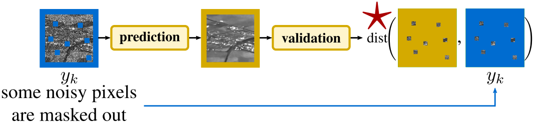
combining dilated convolutions
and conventional convolutions
can also exclude the central pixel

Laine et al. “High-Quality Self-Supervised Deep Image Denoising”, NeurIPS 2019.
Molini et al. “Speckle2Void: Deep self-supervised SAR despeckling with blind-spot convolutional neural networks”, IEEE TGRS 2021

Self-supervised with a single image



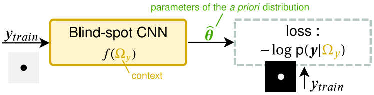
Main idea: train by cross-validation



Improvement 2: include the noisy measurement at the central pixel in the final estimation:

$\rightsquigarrow$ Bayesian framework

network training:

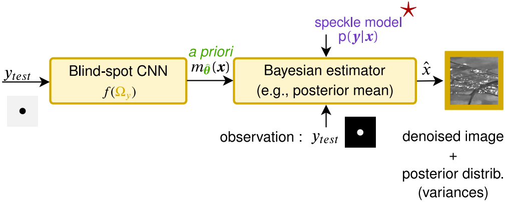


training : maximize the conditional likelihood

$$p(\boldsymbol{y}|\Omega_{\boldsymbol{y}}) = \int \underbrace{p(\boldsymbol{y}|\Omega_{\boldsymbol{y}},\boldsymbol{x})}_{\text{likelihood (speckle model)}} \underbrace{p(\boldsymbol{x}|\Omega_{\boldsymbol{y}})}_{m_{\boldsymbol{\theta}}(\boldsymbol{x})} d\boldsymbol{x}$$

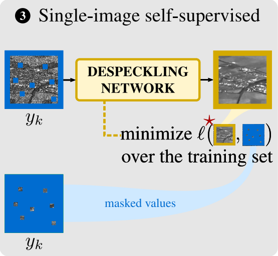
a priori distribution
(conjugate prior)

applying the network to denoise an image:

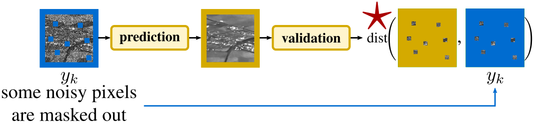


Laine et al. “High-Quality Self-Supervised Deep Image Denoising”, NeurIPS 2019.
Molini et al. “Speckle2Void: Deep self-supervised SAR despeckling with blind-spot convolutional neural networks”, IEEE TGRS 2021

Self-supervised with a single image



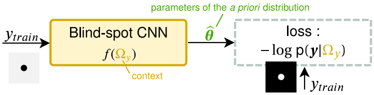
Main idea: train by cross-validation



Improvement 2: include the noisy measurement at the central pixel in the final estimation:

$\rightsquigarrow$ Bayesian framework

network training:

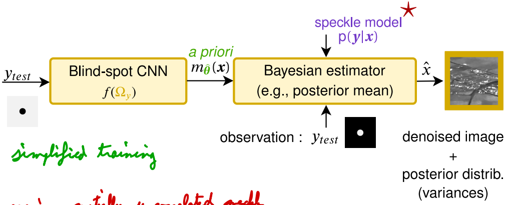


training : maximize the conditional likelihood

$$p(\boldsymbol{y}|\Omega_y) = \int \underbrace{p(\boldsymbol{y}|\Omega_y, \boldsymbol{x})}_{\substack{\text{likelihood} \\ \text{(speckle model)}}} \underbrace{p(\boldsymbol{x}|\Omega_y)}_{m_{\boldsymbol{\theta}}(\boldsymbol{x})} \, d\boldsymbol{x}$$

a priori distribution
(conjugate prior)

applying the network to denoise an image:



simplified training

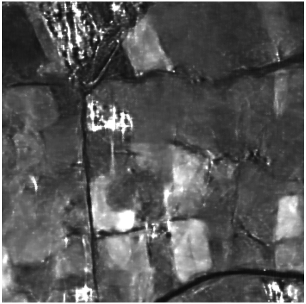
requires spatially uncorrelated speckle

Laine et al. "High-Quality Self-Supervised Deep Image Denoising", NeurIPS 2019.
Molini et al. "Speckle2Void: Deep self-supervised SAR despeckling with blind-spot convolutional neural networks", IEEE TGRS 2021

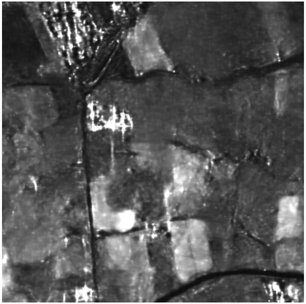
Self-supervised with a single image

Restoration results with Speckle2Void: TerraSAR-X image (©DLR, image pre-processed)

Single-look TerraSAR-X image



Speckle2Void

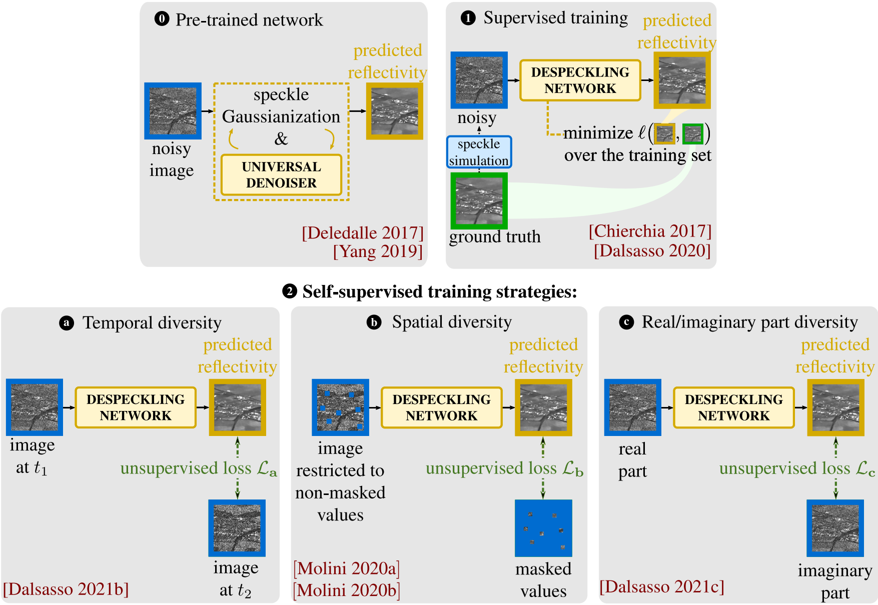


Speckle2Void-NL

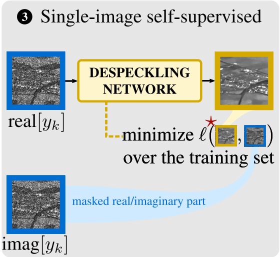
(source: results provided by the Authors at `https://diegovalsesia.github.io/speckle2void`)

Molini et al. "Speckle2Void: Deep self-supervised SAR despeckling with blind-spot convolutional neural networks", IEEE TGRS 2021

A broad overview of deep-learning strategies for despeckling

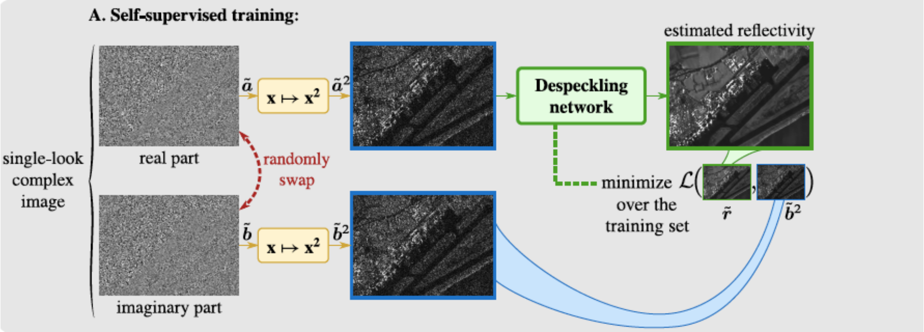


Self-supervised with a single image: real-/imaginary-part decomposition



Main idea: speckle in the real and imaginary parts is independent

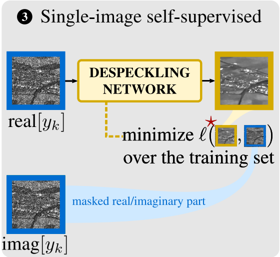
Self-supervised with a single image: real-/imaginary-part decomposition



$$
\begin{aligned}
\mathrm{p}_Z(z) = \mathrm{p}_Z(a+jb) &= \frac{1}{\pi r}\exp(-(a^2+b^2)/r) \\
&= \underbrace{\frac{1}{\sqrt{2\pi}\sqrt{r/2}}\exp(-a^2/r)}_{\mathcal{N}(0,r/2)}\underbrace{\frac{1}{\sqrt{2\pi}\sqrt{r/2}}\exp(-b^2/r)}_{\mathcal{N}(0,r/2)},
\end{aligned} \tag{4}
$$

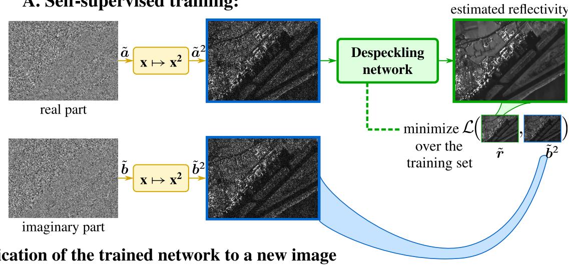
$$
\mathcal{L}(\tilde{\boldsymbol{r}}, \tilde{\boldsymbol{b}}) = \sum_k \frac{1}{2}\log(\tilde{r}_k) + \frac{\tilde{b}_k^2}{\tilde{r}_k}, \tag{5}
$$

Self-supervised with a single image: real-/imaginary-part decomposition

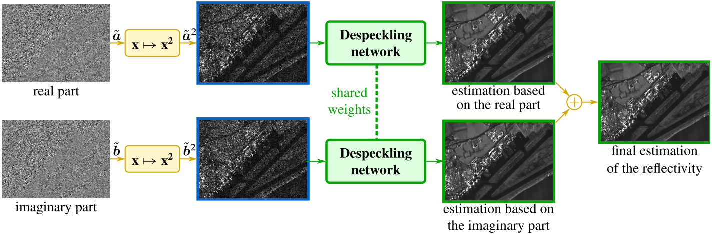


Main idea: speckle in the real and imaginary parts is independent

A. Self-supervised training:

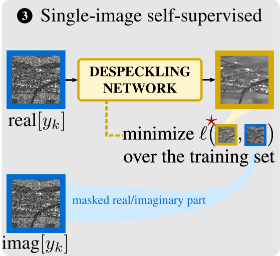


B. Testing: application of the trained network to a new image



Dalsasso et al. “As if by magic: self-supervised training of deep despeckling networks with MERLIN”, TGRS 2022

Self-supervised with a single image: real-/imaginary-part decomposition

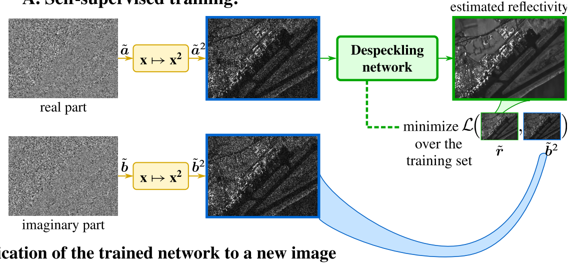


- building the training set is inexpensive
- handles images with spatially correlated speckle
- very good performance

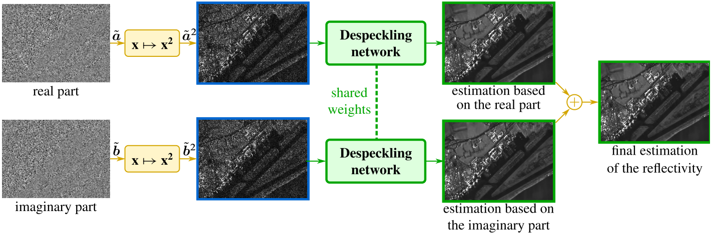
- the real/imaginary parts have a lower SNR than the original image
- requires special care (pre-processing step) for some imaging modalities (e.g., SPOTLIGHT, TOPS)

Main idea: speckle in the real and imaginary parts is independent

A. Self-supervised training:

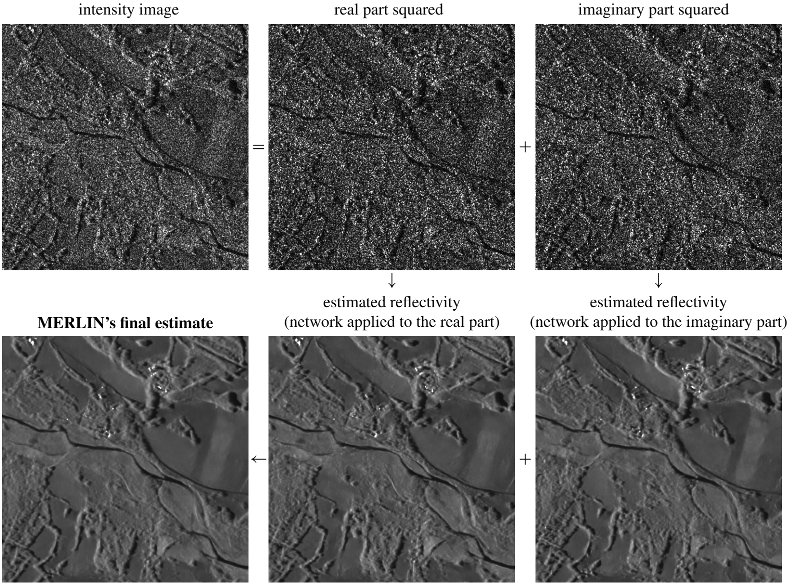


B. Testing: application of the trained network to a new image



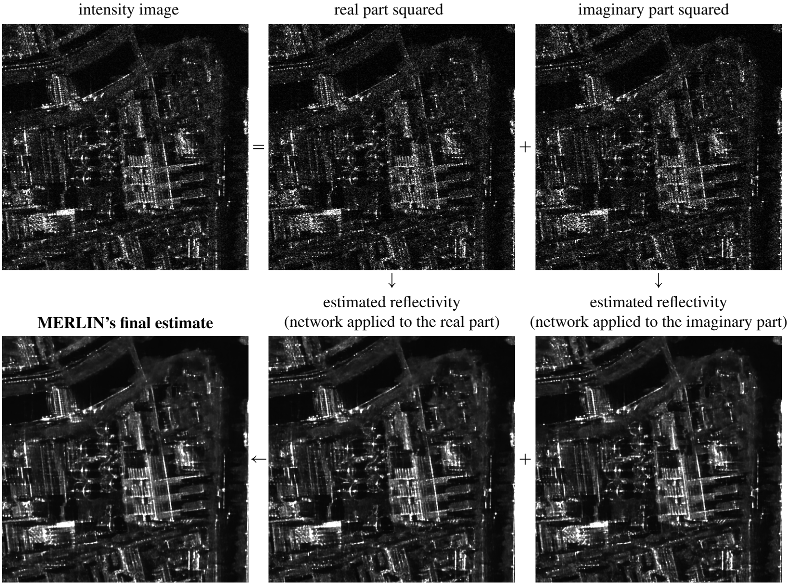
Dalsasso et al. “As if by magic: self-supervised training of deep despeckling networks with MERLIN”, TGRS 2022

Self-supervised with a single image: MERLIN (TerraSAR-X image ©DLR)



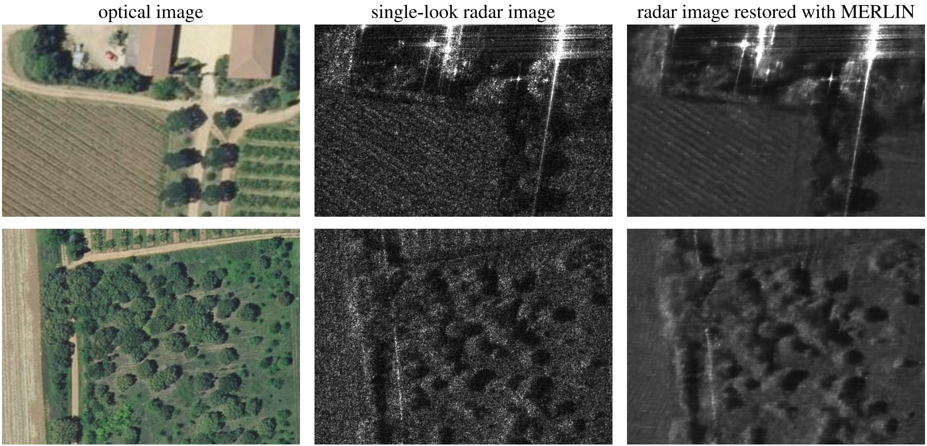
Dalsasso et al. "As if by magic: self-supervised training of deep despeckling networks with MERLIN", TGRS 2022

Self-supervised with a single image: MERLIN (TerraSAR-X image ©DLR)



Dalsasso et al. "As if by magic: self-supervised training of deep despeckling networks with MERLIN", TGRS 2022

Self-supervised with a single image: MERLIN (SETHI image ©ONERA)



Dalsasso et al. “As if by magic: self-supervised training of deep despeckling networks with MERLIN”, TGRS 2022

Outline

Outline

Information extraction

Main tasks for single image processing

- detection of specific objects / classes for a specific application (airports, roads, ships, buildings, hedges, fires,...)
- "semantic" segmentation (global understanding of the scene)

Vienna　　Vienna – Reference

https://project.inria.fr/aerialimagelabeling/

Information extraction

Main tasks for multi-temporal image processing

- detection of specific objects / classes using their temporal signature (type of crops,...) [single classification]
- change detection between two dates ; example of deforestation, of building damage detection)
  semantic change detection (labeling of the type of change)
- change monitoring : monitoring of a specific class in time (evolution of urban areas, forested areas, water areas,...)

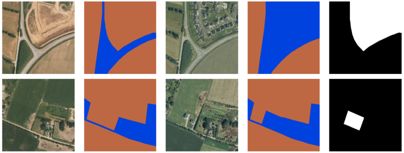


Figure 4.1: Examples of image pairs, land cover maps (LCMs) and associated pixel-wise change maps from the HRSCD dataset. In the depicted LCMs, blue represents the "artificial surfaces" class, and orange represents the "agricultural areas" class.

https://rcdaudt.github.io/hrscd/

Families of methods

- Probabilistic based classification (supervised)
  - use of a statistical model for the distribution of the data (single or multi-channel) conditionally to the classes
  - modeling of the structural information (object shape) using a prior term
  - energy minimization using an adapted optimization framework
- Deep learning (supervised or semi-supervised)
  - training of a network using input / labeled pairs (supervised)
  - building data representation using unlabeled data for a down-stream task (self-supervised learning)

Outline

Modeling

- Random fields
  - the input data are modeled as the realizations of some random variables associated to each pixel of the image defining the observation field $\boldsymbol{D}$
  - the labels are also modeled as a random field $\boldsymbol{L}$ (with values in a fixed set)
  - Using a Maximum a posteriori criterion, the objective is to recover $\hat{\boldsymbol{L}}$ minimizing the following energy :

$$\arg\min_{\boldsymbol{L}} \; \mathcal{L}(\boldsymbol{D},\boldsymbol{L}) + \mathcal{U}(\boldsymbol{L}). \tag{6}$$

- Energy terms
  - (neg-log)Likelihood term $\mathcal{L}(\boldsymbol{D},\boldsymbol{L}) = -\log P(\boldsymbol{D}|\boldsymbol{L})$ (data attachment term)
    This term is usually defined in a supervised way using the sensor physical knowledge and some samples of each class
  - prior term $\mathcal{U}(\boldsymbol{L})$
    This term usually introduces some prior knowledge on the shape of the objects (compact areas, linear structures, etc.)

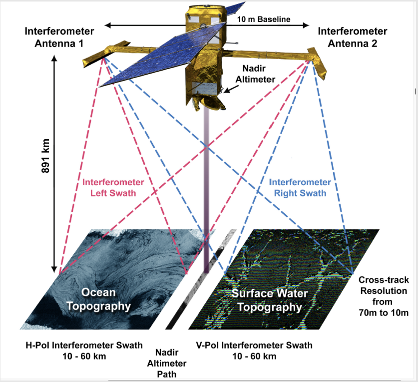
- Minimization algorithms
  simulated annealing, graph-cut based, ...

SWOT mission

NASA / CNES satellite launched dec. 2022

Principal instrument of SWOT: KaRIn (Ka band Radar Interferometer)

- Ka-band: f = 35.75GHz, λ = 8.6mm
- near-nadir incidence angle: 0.6° to 3.9°
- resolution $5 \times 70m$ (near range) to $5 \times 10m$ (far range)



https.swot.cnes.fr

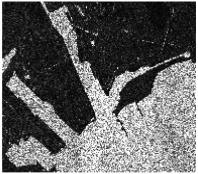
SWOT mission

NASA / CNES satellite launched dec. 2022

Objectives of SWOT mission

- ocean monitoring (ocean mode)
- water surface monitoring (lakes, rivers)
  - automatic extraction
  - elevation extraction using interferometric phase

Landsat 8 (optic) image



SWOT (SAR)

Example of probabilitic method

NASA / CNES satellite launched dec. 2022

Example of SWOT water surface extraction

- Random fields

$$\arg\min_{\boldsymbol{L}} \ \mathcal{L}(\boldsymbol{D}, \boldsymbol{L}) + \mathcal{U}(\boldsymbol{L}).$$

  For each pixel s, $\boldsymbol{l}_s \in \{0, 1\}$ (land, water)

- Energy terms
  - (neg-log)Likelihood term $\mathcal{L}(\boldsymbol{D}, \boldsymbol{L}) = -\log P(\boldsymbol{D}|\boldsymbol{L})$ (data attachment term) Goodman model of SAR amplitude distribution

$$\mathcal{L}(\boldsymbol{D}, \boldsymbol{L}) = \sum_s 2\log(\boldsymbol{\mu}_{\boldsymbol{l}_s}) + \frac{\boldsymbol{d}_s^2}{\boldsymbol{\mu}_{\boldsymbol{l}_s}^2}$$

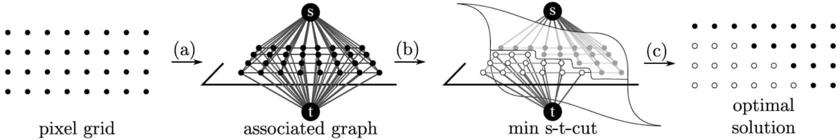
  - prior term $\mathcal{U}(\boldsymbol{L})$

$$\mathcal{U}(\boldsymbol{L}) = \sum_{(s,t)} \psi(\boldsymbol{l}_s, \boldsymbol{l}_t)$$

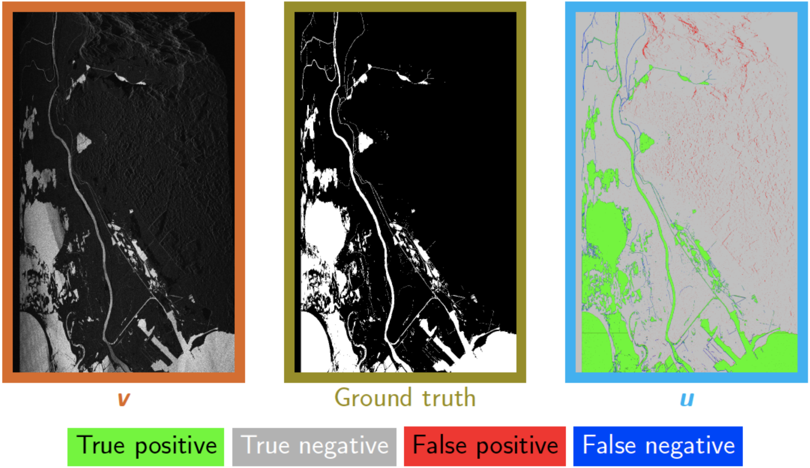
- Minimization algorithms efficient graph-cut optimization for binary cases

Efficient optimization with graph-cut

- definition of a graph with pixels as nodes plus 2 terminal nodes
- weighting of the edges between pixel-node and terminal-node using the data attachment terms
- min-cut computation and deduction of the associated labeling



Example of probabilitic method



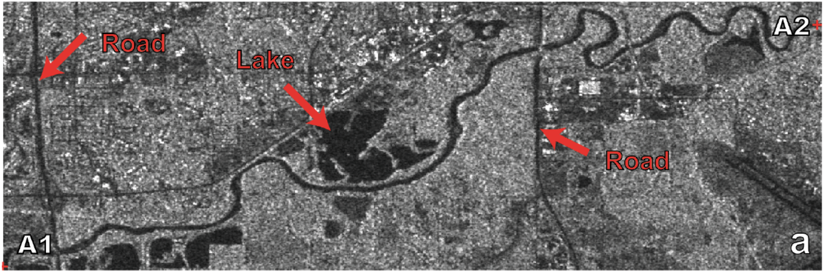
Extrait de la thèse de S. Lobry `https://www.sylvainlobry.com/`

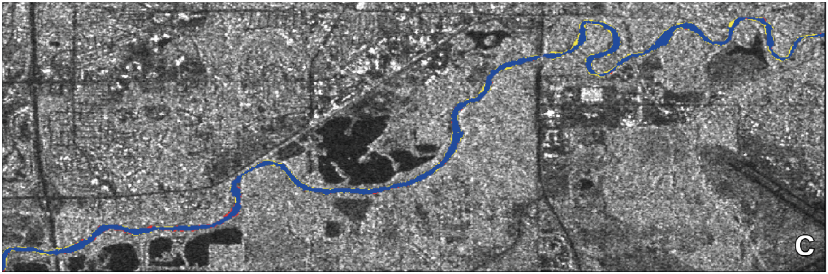
Example of probabilitic method

Advantages of the supervised framework

- Very flexible framework: in the case of SWOT data it is possible to improve the models in multiple ways :
  - variable parameters for land and water classes to take into account the position in the swath beacuse of the antenna pattern
  - prior field to favor linear strucures using an external field
  - extension of the framework to multi-temporal series by introducing temporal links in the classification (temporal neighbors)
- Very controlable framework (small set of parameters to modify to adapt to new data)

Example of probabilitic method - SWOT river segmentation





Extrait de la thèse de N. Gasnier `https://www.theses.fr/2022IPPAT002`

Specificities of multi-temporal series

Talk of M. Fauvel (wednesday 7th)

Classification of muli-temporal series

- pixel-wise classification (no spatial term)
- supervised framework
- modeling with mixture of Gaussian processes
- application to multi-temporal satellite data (possibly incomplete - cloudy images)

Outline

Supervised learning

- Training set with labeled data
- Architecture and loss : ResNet backbone, CE (cross-entropy loss) (see course of Emanuele)
- Limits
  - limited size of the available training set
  - noise in the labels of the training sets
  - lack of generalization (loss of performances when different acquisition conditions or sensors)
  - computational cost and need of memory size
- Fine tuning, transfer learning, self-supervision

https://captain-whu.github.io/GID15/

Deep learning methods: from supervision to self-supervision

Self-supervised learning

- Context:
  High availability of raw data
  Low availablity of labeled data

- Objective: exploiting the huge set of unlabeled remote sensing data to improve methods

- Principle: the model (network) learns a good representation of the data using supervision signals generated from the data themselves

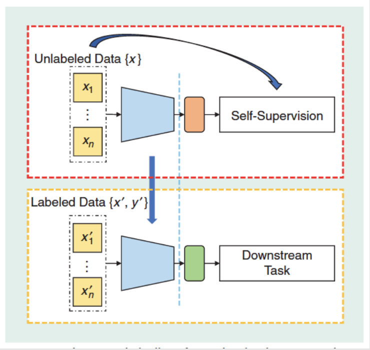


Figure from Y. Wang et al.

Y. Wang et al., "*Self-supervised learning in remote sensing - A review," IEEE Geoscience and Remote Sensing Magazine*", 2022.

Example of a semi-supervised learning framework (1)

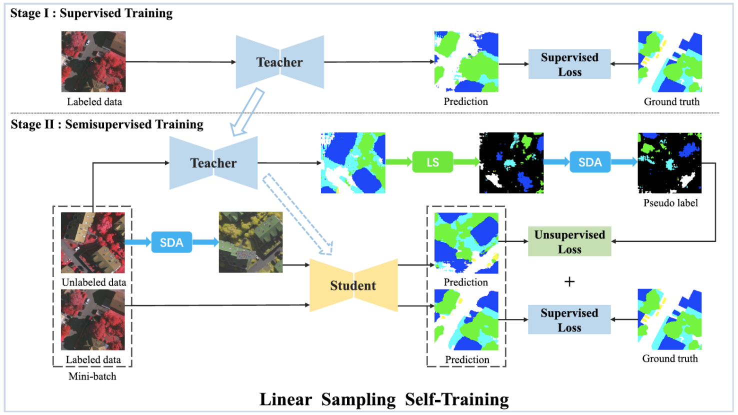
Semi-supervised semantic segmentation

- Context: a reduced set of labeled data but a large database of unlabeled data
- Principle:
  - use knowledge distillation (teacher-student network) [the teacher learns on the labeled data and is then used to generate labels for the student network training]
  - strong data augmentation :
    color transformations (color modifications and blur)
    geometric transformations: rotations
    cutout operation (masking with noise)
  - reliable pseudo-label selection (only the most reliable labels are selected and used in the loss definition)

X. Lu et al., "*Simple and efficient: A semi-supervised Learning Framework for Remote Sensing Image Semantic Segmentation", IEEE Trans. on Geoscience and Remote Sensing*", 2022.

Example of a semi-supervised learning framework (1)

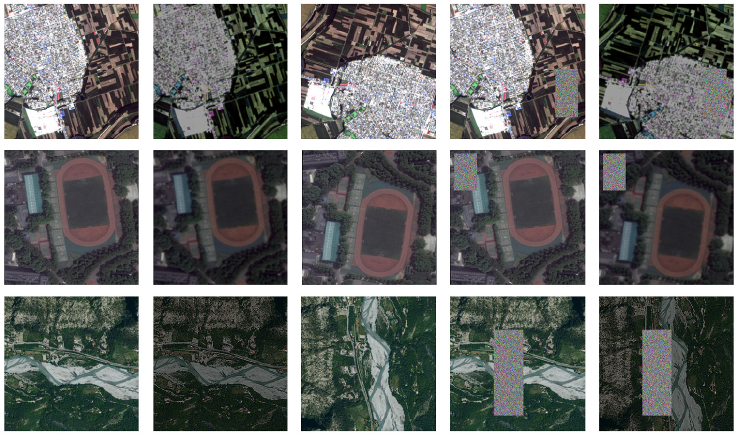
Semi-supervised semantic segmentation



X. Lu et al., "*Simple and efficient: A semi-supervised Learning Framework for Remote Sensing Image Semantic Segmentation", IEEE Trans. on Geoscience and Remote Sensing*", 2022.

Example of a semi-supervised learning framework (1)

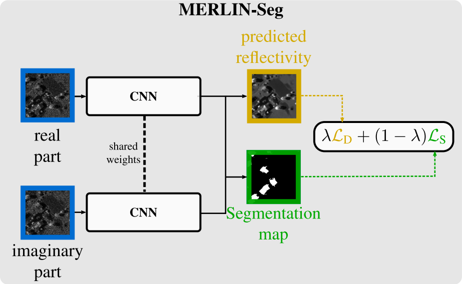
Importance of the data augmentation step



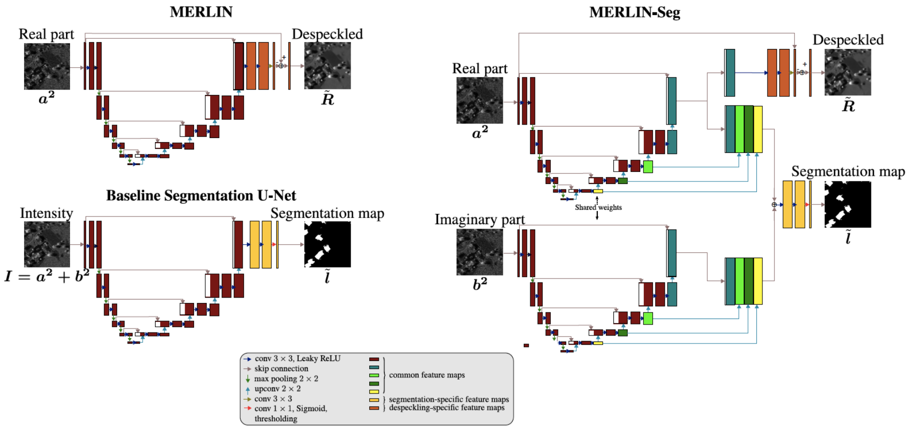
X. Lu et al., “*Simple and efficient: A semi-supervised Learning Framework for Remote Sensing Image Semantic Segmentation", IEEE Trans. on Geoscience and Remote Sensing*”, 2022.

Example of a semi-supervised learning framework (2)

- Context: low availability of labels for semantic segmentation (namely footprint extraction)
- Learn a structured latent space: while learning to suppress speckle, the network co-learns to segment images
- Label-efficient strategy relying on self-supervised despeckling



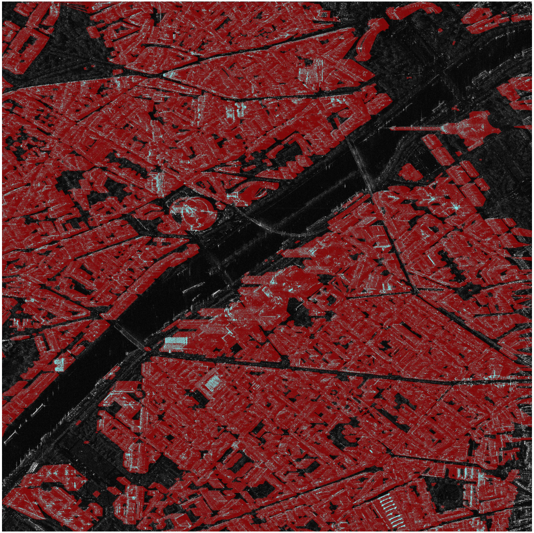
Example of a semi-supervised learning framework (2)



Example of a semi-supervised learning framework (2): EMPRISE dataset

- Simulated image of 12600×10800 pixels
- The slant-range resolution is 0.71 m and azimuth resolution is 1 m
- Patches of size 256×256
- 11161 image patches for training
- 5 image patches for validation
- 70 image patches for testing
- The U-Net backbone is composed of 5 downsampling layers 523 (compressing the 256×256 image patch down to an 8×8 latent representation) and 5 upscaling layers

Example of a semi-supervised learning framework (2): TerraSAR-X dataset



- SpotLight image of 6000×6000 pixels
- The slant-range resolution is 0.45 m and azimuth resolution is 0.87 m
- Patches of size 512×512
- 1189 image patches for training
- 3 image patches for validation
- 65 image patches for testing
- The U-Net backbone is composed of 6 downsampling layers 523 (compressing the 256×256 image patch down to an 8×8 latent representation) and 6 upscaling layers

MERLIN-Seg: quantitative results

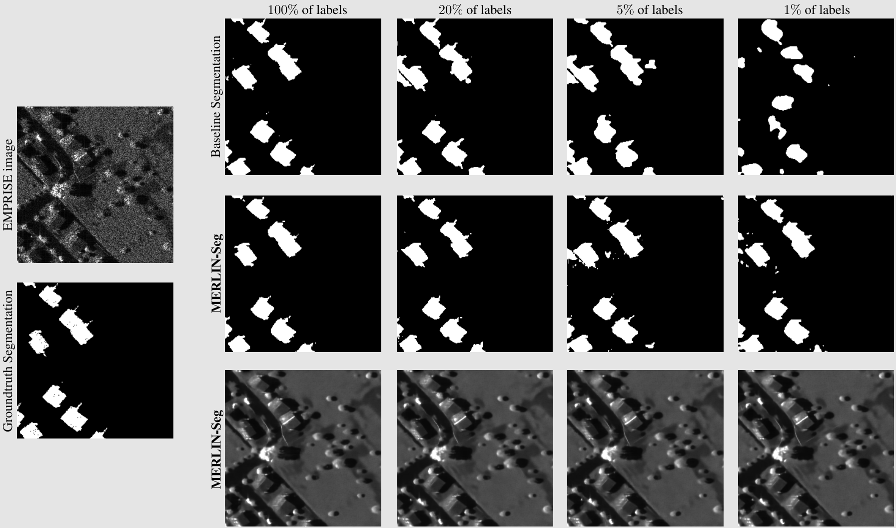
Results with $\lambda = 0.5$ on EMPRISE simulated dataset

	100%			**20%**			**5%**			**1%**		
	mIoU	F1	Acc	mIoU	F1	Acc	mIoU	F1	Acc	mIoU	F1	Acc
Baseline	0.710	**0.803**	0.974	0.674	0.758	0.969	0.551	0.669	0.952	0.195	0.299	0.897
MERLIN-Seg	**0.711**	0.799	**0.976**	**0.707**	**0.794**	**0.976**	**0.680**	**0.781**	**0.971**	**0.634**	**0.745**	**0.965**

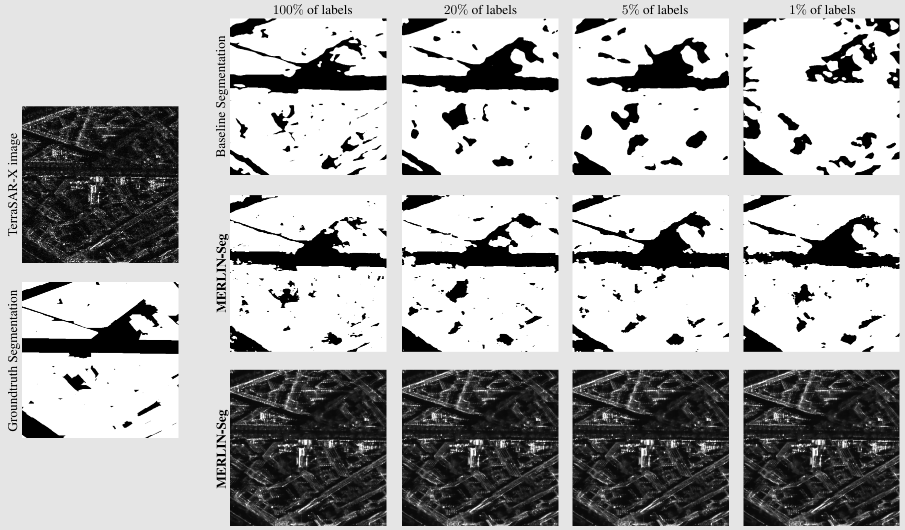
Results with $\lambda = 0.5$ on TerraSAR-X dataset

	100%			**20%**			**5%**			**1%**		
	mIoU	F1	Acc	mIoU	F1	Acc	mIoU	F1	Acc	mIoU	F1	Acc
Baseline	0.748	0.832	0.874	0.746	0.831	0.873	0.714	0.806	0.847	0.633	0.750	0.790
MERLIN-Seg	**0.765**	**0.844**	**0.886**	**0.760**	**0.840**	**0.882**	**0.760**	**0.840**	**0.883**	**0.756**	**0.838**	**0.882**

MERLIN-Seg: results on EMPRISE dataset



MERLIN-Seg: results on TSX dataset



Overview of self-supervised learning methods

Principle of self-supervised learning methods

Main idea : learning a representation of the data generic enough to be used later for an other task (downstream task)

3 families of methods :

- Generative methods
- Predictive methods
- Contrastive methods

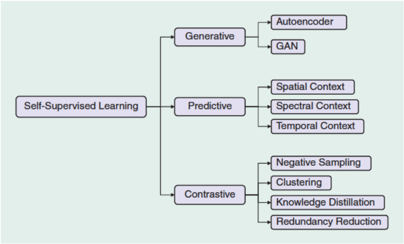


FIGURE 3. A taxonomy of SSL.

Figure from Y. Wang et al. on data augmentation

Y. Wang et al., "*Self-supervised learning in remote sensing - A review," IEEE Geoscience and Remote Sensing Magazine*", 2022.

Overview of self-supervised learning methods

Principle of self-supervised learning methods

Main idea : learning a representation of the data generic enough to be used later for an other task (downstream task)

3 families of methods :

- Generative methods
- Predictive methods
- Contrastive methods

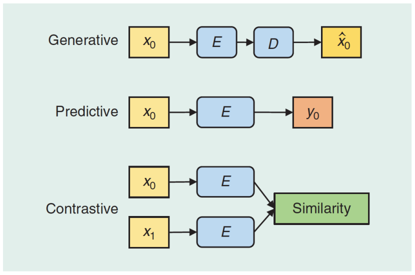


Figure from Y. Wang et al. on data augmentation

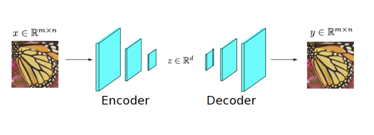
Y. Wang et al., "*Self-supervised learning in remote sensing - A review," IEEE Geoscience and Remote Sensing Magazine*", 2022.

Self-supervised learning methods with generative methods

Principle of generative methods

Learning of representation by reconstruction or generation of the input data.

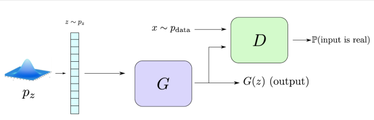
- Auto-encoders:
  - encoder E maps the input x to a latent vector $z = E(x)$ (usually of reduced dimension)
  - decoder D reconstructs $\hat{x}$ from z
  - self-supervised loss $||x - D(E(x))||$
- Variational auto-encoders:
  - instead of encoding a deterministic latent representation z, the encoders generates the parameters $\boldsymbol{\mu}$ and σ of a latent Gaussian distribution
  - the output is reconstructed from the sampling of this distribution
  - loss : reconstruction loss + control of the latent distribution (forced to be close to $\mathcal{N}(0, I)$)

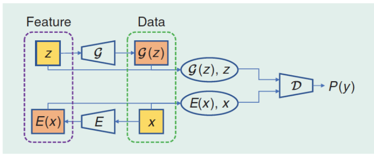


Use of GAN (Generative Adversarial Networks)

Learning a generative model in a two-player game between a generator $\mathcal{G}$ and a discriminator $\mathcal{D}$

- Adversarial Auto-encoders:
  - encoder E maps the input x to a latent vector $E(x)$ (usually of reduced dimension)
  - generator $\mathcal{G}$ maps a z code to a generated data $\mathcal{G}(z)$
  - the discriminator is trained to discriminate between the pairs $(E(x), x)$ and $(z, \mathcal{G}(z))$
  - E, $\mathcal{G}$, $\mathcal{D}$ are trained simultaneously





Use of self generated labels

The network is trained using a pretext task based on self-generated labels. Different context information from the input data can be used.

- Spatial context and pretext tasks:
    - predict the relative position between crops in the image
    - solve complex spatial puzzle
    - find the rotation between patches
    - inpainting
- Spectral context and pretext tasks:
    - (predict the a color channel using other color channels of the image)
    - reducing spectral dimension for multi-spectral image
- Temporal context and pretext tasks:
    - (frame order prediction)
    - (missing frame prediction)
    - association of similar patches

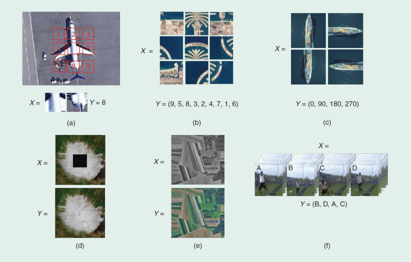


FIGURE 6. Examples of different predictive pretext tasks: (a) relative position, (b) jigsaw puzzle, (c) rotation, and (d) inpainting are based on spatial contexts; (e) colorization is based on spectral context; (f) frame order illustrates temporal context.

Figure from Y. Wang et al.

Extracting a semantic information

The network is trained to put close the representations of samples semantically similar (the loss is computed in the representation space).

- Negative sampling:
  - both positive and negative pairs of samples; positive samples come from data augmentation and negative samples are other datapoints in the dataset.
  - losses: triplet loss, contrastive predictive coding, ...
- Clustering:
  - pseudo-labels using k-means clustering
  - different views of the same image should be assigned the same label
- Knowledge distillation:
  - teacher-student framework
  - no negative samples
- Redundancy reduction:
  - optimization of the cross-correlation matrix (to be similar to the identity matrix) between the representations of distorted samples

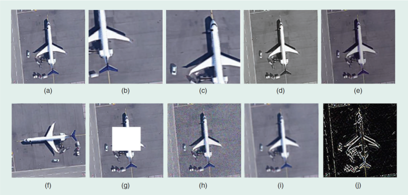


FIGURE 12. Data augmentation operators studied in SimCLR [76]. The selected best group of data augmentations (random cropping, color jittering, grayscaling, Gaussian blurring, and horizontal flipping) are widely referenced in the following self-supervised studies. (a) Original. (b) Crop and resize. (c) Crop, resize, and flip. (d) Color drop. (e) Color jitter. (f) Rotate. (g) Cutout. (h) Gaussian noise. (i) Gaussian blur. (j) Sobel filter.

Figure from Y. Wang et al.

Outline

Practical work

Change detection with optical images

- deforestation detection using spectral features
- change detection using features extracted by CNN layers

Change detection and multi-temporal analysis with SAR data

- change detection
- effect of speckle reduction
- multi-temporal visualization (color composition, REACTIV method)

Conclusion

Impact of very large data

- evolution of image processing tools to model-based to data-based methods
- taking into account physical knowledge on the acquisition system or data properties helps to "improve" the networks (faster to train, lighter networks, higher interpretability, etc.)

Active fiels

- representation learning
- explainability of the networks
- uncertainty computation