Speckle2Void: Deep Unsupervised SAR Despeckling with Blind-Spot Convolutional Neural Networks

Andrea Bordone Molini, Diego Valsesia, Giulia Fracastoro, and Enrico Magli

Politecnico di Torino - Italy

Nov. 26, 2020

Synthetic Aperture Radar (SAR)

- Synthetic Aperture Radar (SAR):
  - **coherent imaging system** able to acquire ground images by emitting radiations and capturing the signals backscattered from the imaged scene.
- SAR images suffer from the presence of speckle, a signal dependent granular noise.
  - **Difficult interpretation** and information extraction.
  - **Reduce the effectiveness of scene analysis algorithms** (e.g., image segmentation, detection and recognition)
- Speckle is usually modelled as multiplicative noise:

$$y = nx$$

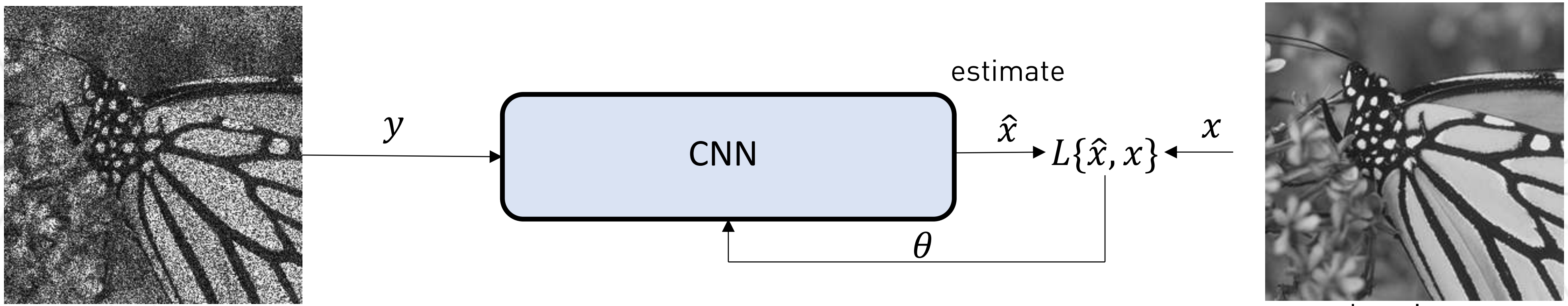
Noisy SAR image (y) — Speckle noise (n) — Unobserved clean image (x)

Deep learning for Despeckling

- Despeckling is the very first step of many scene analysis algorithms.
- Recently, deep learning works have been suggested to exploit the powerful learning capabilities of CNN.
- All the proposed deep learning methods for despeckling exploit supervised learning:
  - Real SAR images do not have the corresponding clean images.
  - Resorting to synthetically speckled optical images or multi-temporal SAR images which are difficult to acquire.

Despeckling with supervised learning

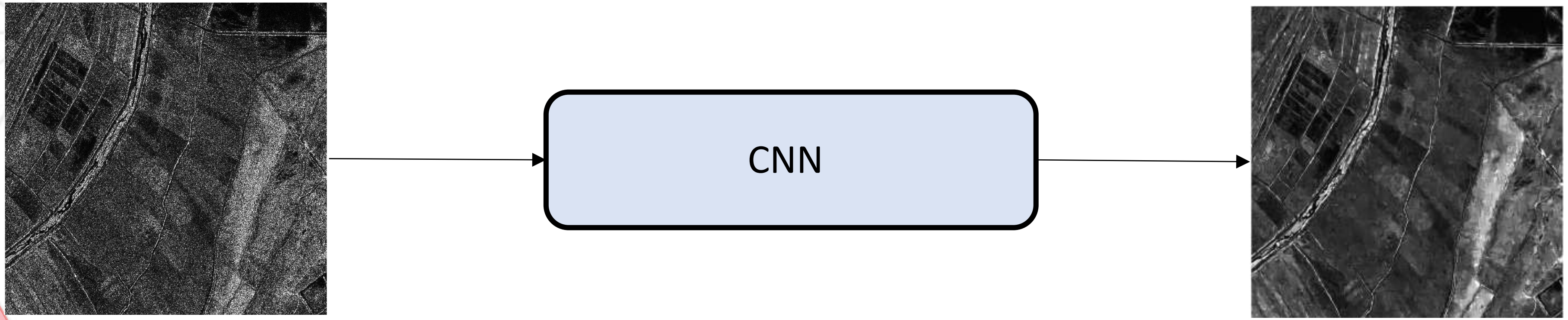
- **Training** → synthetic dataset (optical images are used as ground truth and their artificially speckled version as noisy inputs)



noisy image

clean image

- Testing → test SAR images are fed to the **pre-trained CNN.**



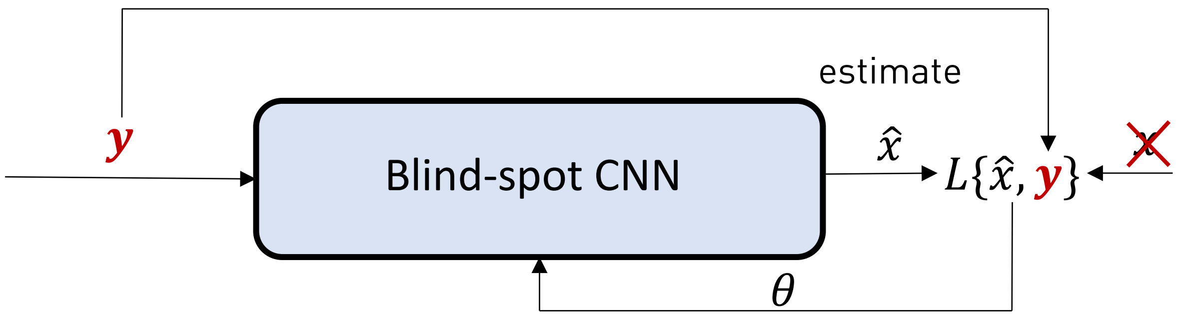
noisy SAR image

SAR estimate

Domain gap

- Transfering knowledge is an issue:
  - Using pre-trained CNN on SAR images for testing:
    - Effective from a noise suppression viewpoint;
    - Presence of artifacts across the image;
    - Poor preservation of the radiometric features (different geometries and distribution properties).

- Solution:
  - Exploiting a particular class of CNN called **deep blind-spot network** to enable direct training on real SAR images (Terra-SarX satellite) and **no ground truth.**



Background

Supervised training

y → CNN → $\hat{x} \longleftrightarrow x$

noisy image

clean estimate

clean image

$$\boldsymbol{max\, p\left(x_i \middle| y_i, \Omega_{y_i}\right)}$$

receptive field of y_i,
excluding y_i itself

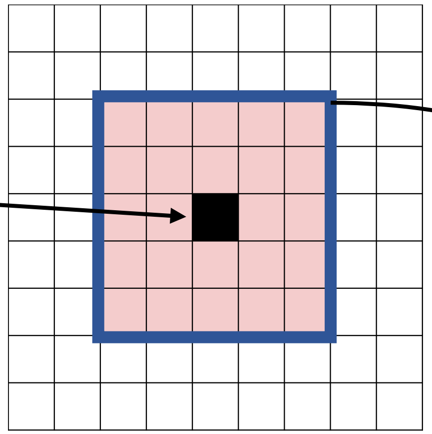
Self-supervised training

y → Blind-spot CNN → $\hat{x} \longleftrightarrow y$

pixelwise i.i.d noise

$$\boldsymbol{max\, p\left(x_i \middle| \Omega_{y_i}\right)}$$

y_i



$$\hat{x}_i = f(\Omega_{y_i})$$

$$\boldsymbol{loss = \left(f(\Omega_{y_i}) - y_i\right)^2}$$

No ground truth

Watch out: the noisy image y is never used to produce the final clean estimation.

Model

- The following Bayesian denoising framework incorporates the noise model allowing to use the noisy pixels to compute the clean estimate.

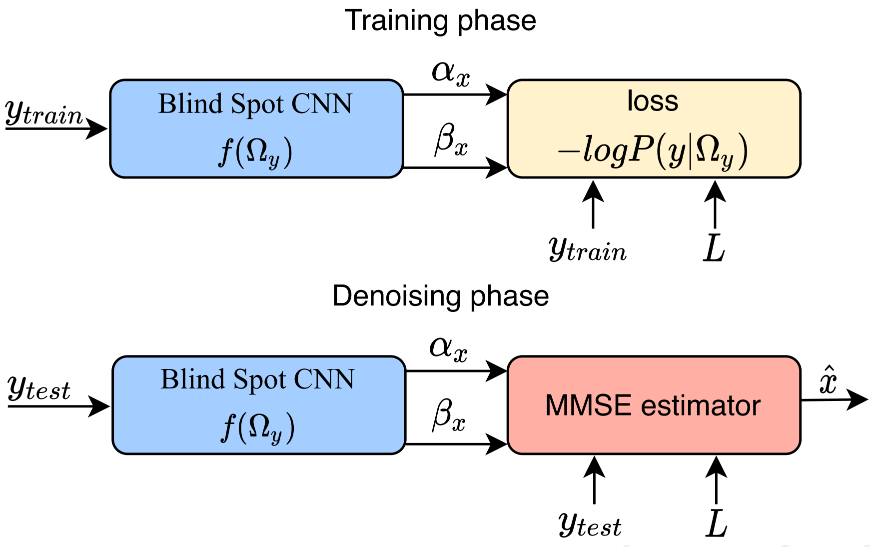
$$p\left(x_i \middle| y_i, \Omega_{y_i}\right) \propto p(y_i|x_i)p\left(x_i \middle| \Omega_{y_i}\right)$$

- $p(y_i|x_i)$ is the noise distribution, modelled as a $\boldsymbol{\Gamma(L, L)}$
- $p\left(x_i \middle| \Omega_{y_i}\right)$ is the conditional image prior, modelled as an $\boldsymbol{inv\Gamma\left(\alpha_{x_i}, \beta_{x_i}\right)}$
- $p\left(x_i \middle| y_i, \Omega_{y_i}\right)$ is the conditional posterior, modelled as an $\boldsymbol{inv\Gamma\left(L + \alpha_{x_i}, \beta_{x_i} + Ly_i\right)}$

Speckle2Void (training/testing)

- Training:
  - The **blind-spot CNN is trained** to minimize the negative log likelihood $p(y_i|\Omega_{y_i})$ to produce $\alpha_{x_i}, \beta_{x_i}$ of $p(x_i|\Omega_{y_i})$ that best fit the noisy observations y_i.

$$p(y_i|\Omega_{y_i}) = \frac{L^L y_i^{L-1}}{\beta_{x_i}^{-\alpha_{x_i}} Beta(L, \alpha_{x_i})(\beta_{x_i} + L y_i)^{L+\alpha_{x_i}}}$$

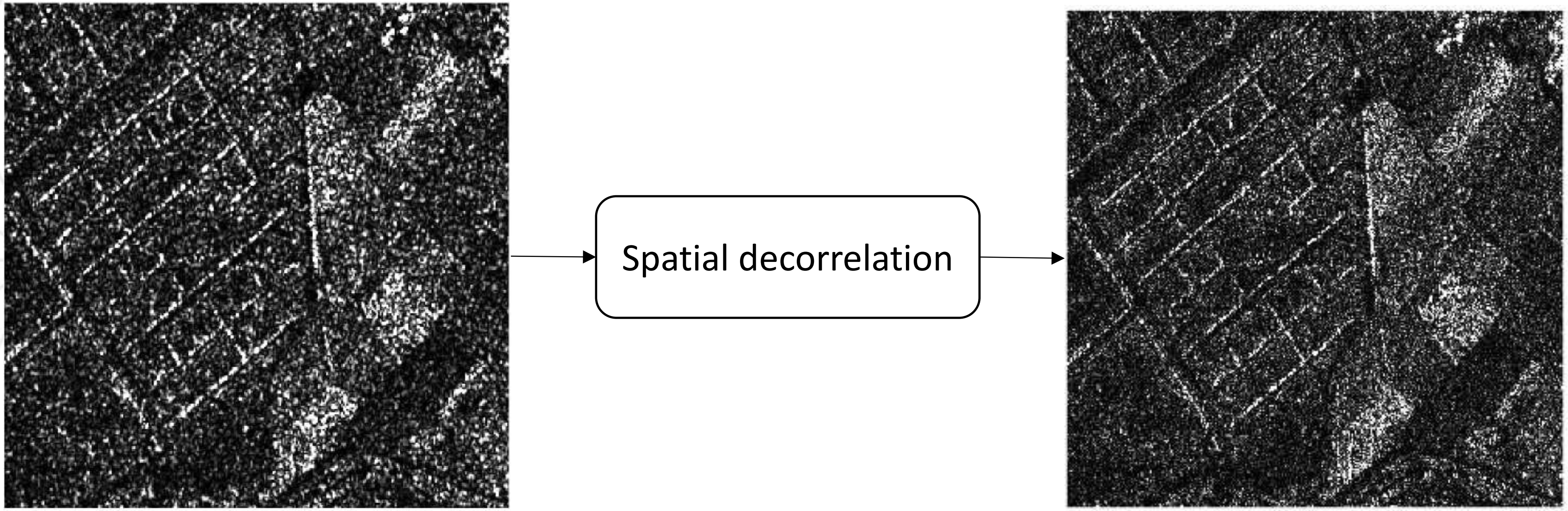


- Testing:
  - **Step 1**: the blind-spot CNN produces α_{x_i} and β_{x_i} for each pixel.
  - **Step 2**: use α_{x_i} and β_{x_i} to compute the expected value of the posterior distribution:

$$\hat{x}_i = E[x_i|y_i, \Omega_{y_i}] = \frac{\beta_{x_i} + L y_i}{L + \alpha_{x_i} - 1}$$

Correlated noise (I)

- Training a blind-spot network requires the noise to be spatially uncorrelated:
  - In SAR images, noise is spatially correlated.
- To solve this, we leverage two aspects:
  - whitening procedure [1] as a pre-processing step, to decorrelate the speckle.
  - a regularized training procedure with a variable blind-spot shape (in order to account for the autocorrelation of the speckle process);

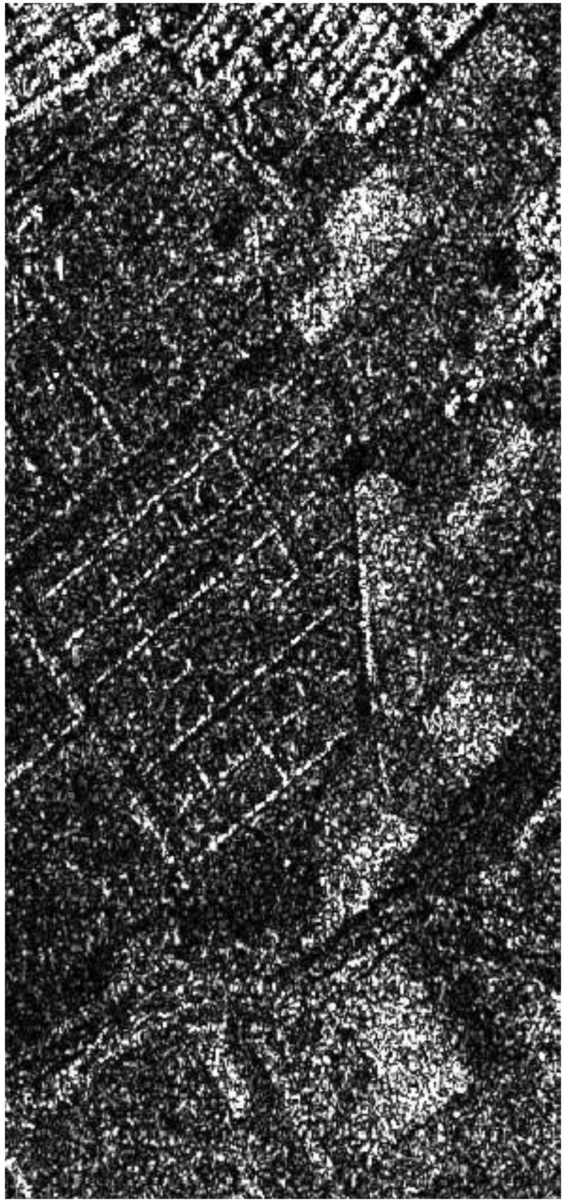


[1] A. Lapini, T. Bianchi, F. Argenti, and L. Alparone, "Blind speckle decorrelation for SAR image despeckling,"IEEE Transactions on Geoscienceand Remote Sensing, vol. 52, no. 2, pp. 1044–1058, Feb 2014.

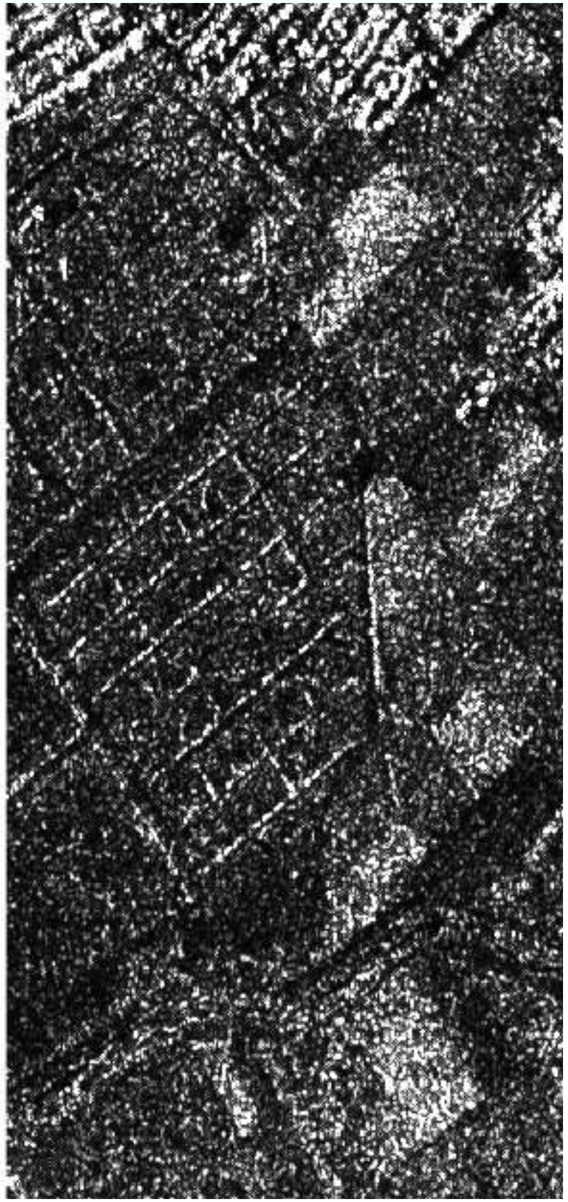
Whitening procedure (I)

- Applying a decorrelation process before training our blind-spot network is **compelling** to obtain a decent output clean SAR image

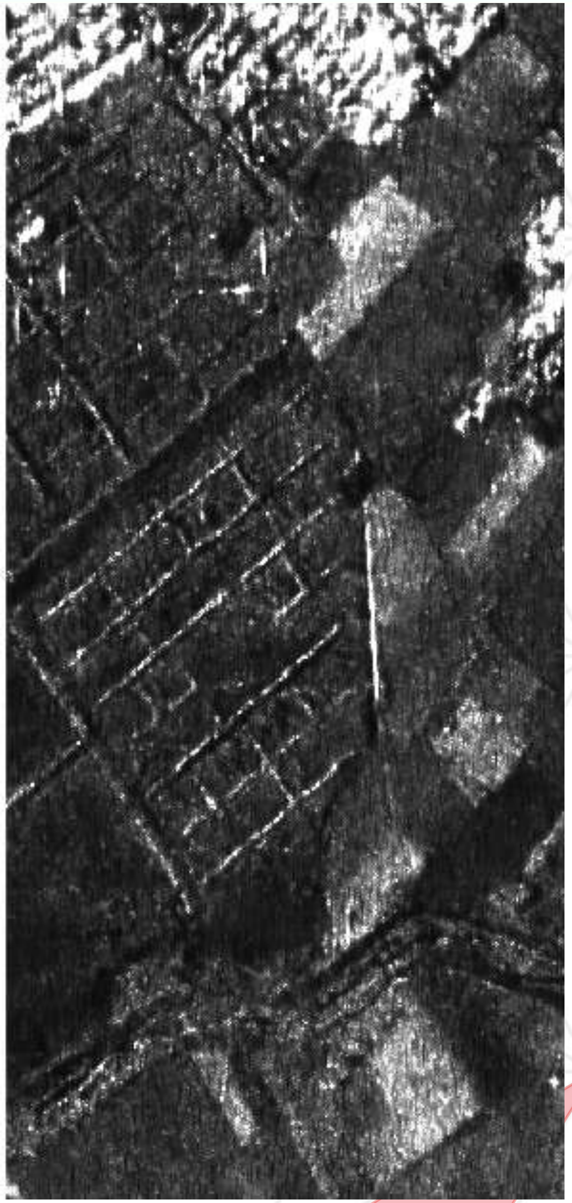
Original noisy



1x1 spot

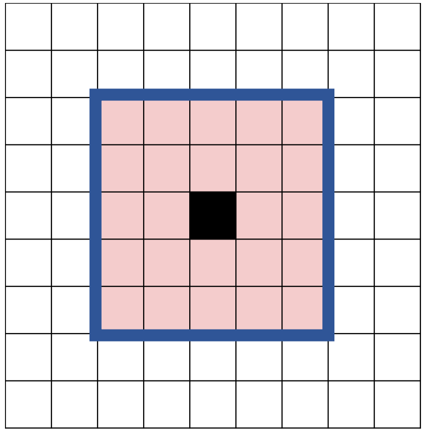


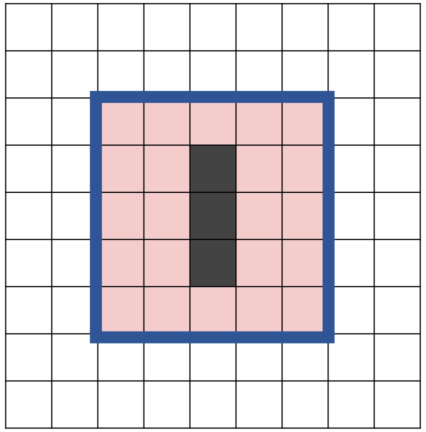
White, 1x1 spot



Regularized training procedure

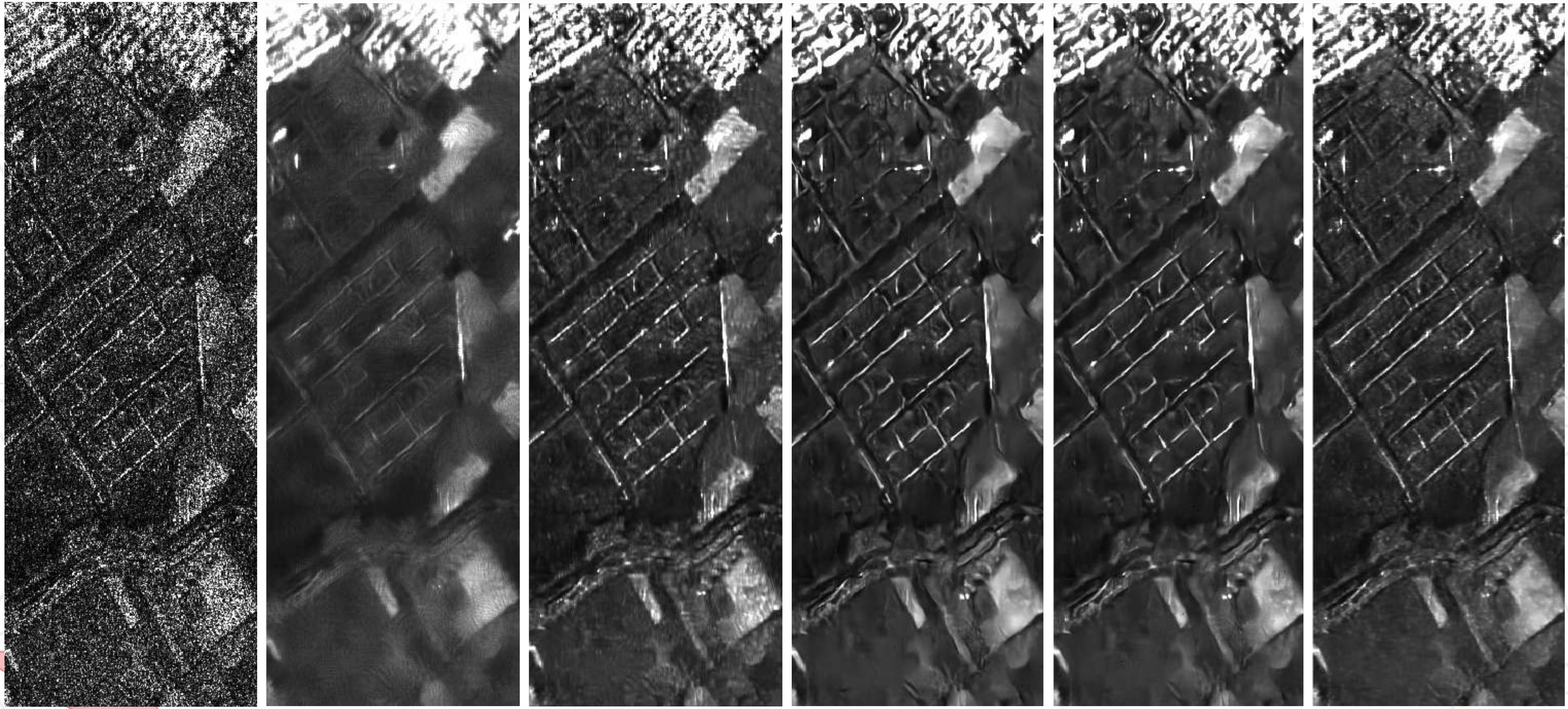
- During training, we alternate the following blind-spot shapes with predefined probability:





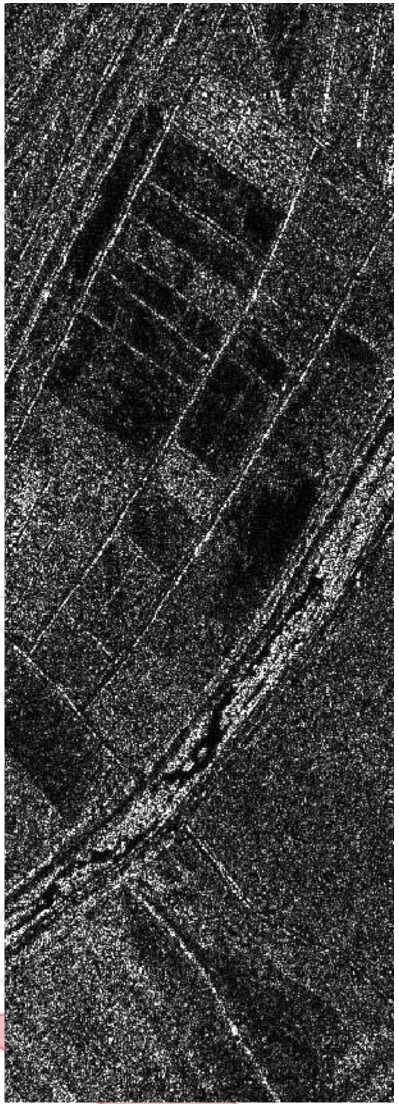
- It allows to tune the degree of reliance of the CNN on the immediate neighbours
  - improvement of the high frequency details in the denoised image;
  - suppression most of the noise correlation.

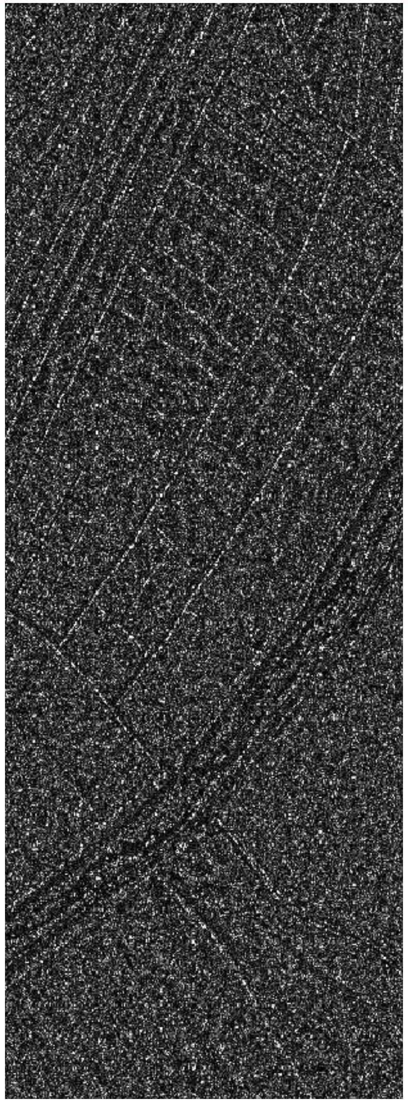
Qualitative performance (I)

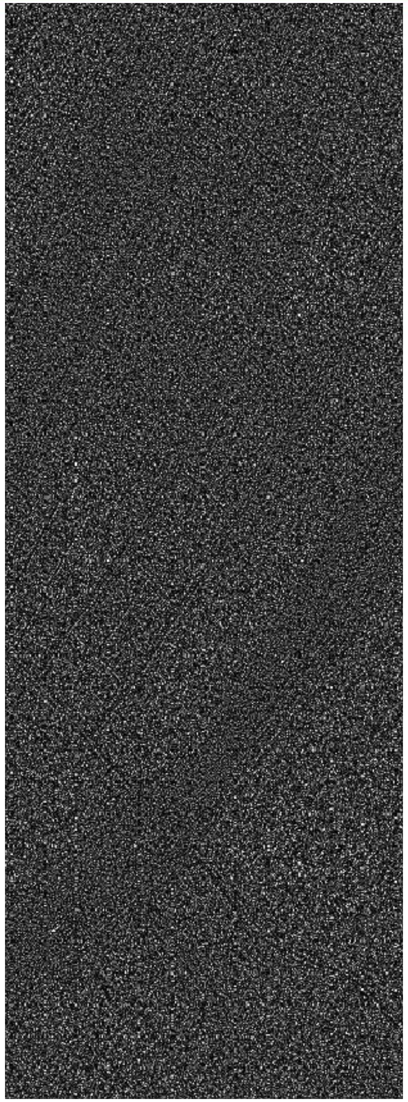


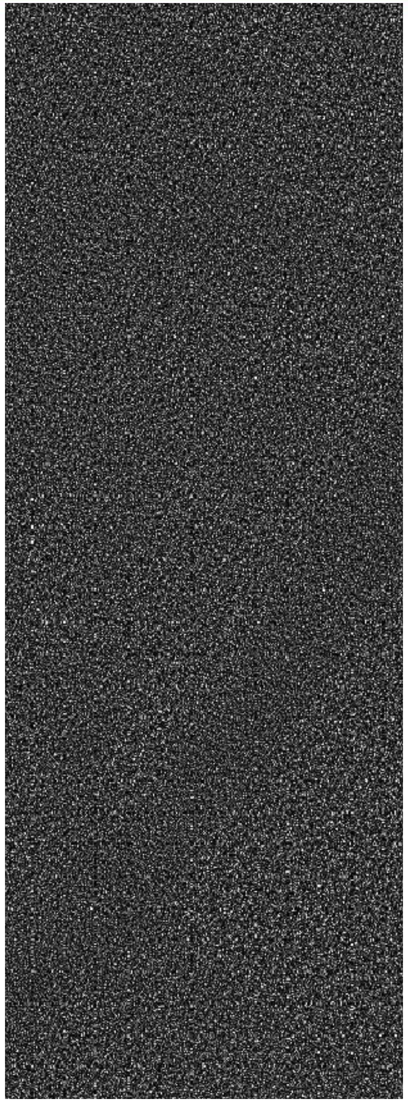
Noisy(white) | PPB | SAR-BM3D | CNN baseline | ID-CNN | Speckle2Void

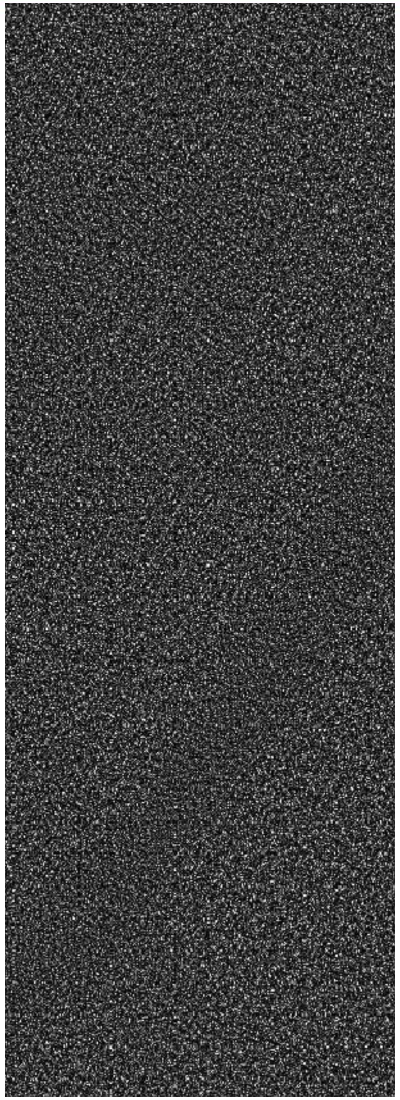
Qualitative performance (II)

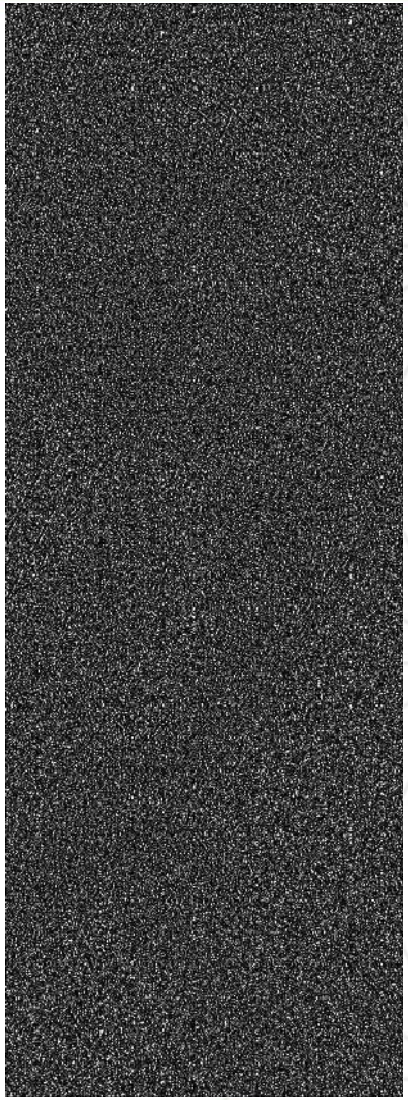

Noisy(white)


PPB ratio


SAR-BM3D ratio


CNN baseline ratio


ID-CNN ratio


Speckle2Void ratio

Quantitative performance

Metric	Image	PPB [31]	SAR-BM3D [7]	CNN baseline	ID-CNN [12]	Speckle2Void	Speckle2Void NL
ENL ↑	1	82	46.2	52.9	76.5	**88.5**	86.5
	2	78.6	49.1	48.7	69.9	**89.9**	81.8
	3	76.9	58.1	52.5	73.1	84.0	**86.0**
	4	54.2	40.4	37.6	46.2	**54.7**	53.1
	5	**22.9**	16.2	14.6	16.6	18.9	17.5
μ_r ↑	1	0.887	0.919	0.963	0.943	0.966	**0.970**
	2	0.925	0.938	**0.969**	0.964	0.966	0.967
	3	0.926	0.941	**0.974**	0.969	0.968	0.968
	4	0.933	0.942	0.974	0.976	0.962	**0.977**
	5	0.853	0.894	**0.950**	0.918	0.947	0.946
σ_r ↑	1	**0.847**	0.627	0.726	0.745	0.803	0.800
	2	**0.886**	0.674	0.740	0.803	0.829	0.817
	3	**0.874**	0.684	0.756	0.817	0.816	0.814
	4	**0.876**	0.688	0.755	0.846	0.823	0.837
	5	**0.891**	0.549	0.683	0.664	0.748	0.736
$\mathcal{M}$ [44] ↓	1	24.4	16.5	11.9	14.6	7.72	**6.71**
	2	10.1	11.6	11.6	9.12	9.11	**8.04**
	3	9.82	11.3	11.3	6.93	6.24	**5.44**
	4	10.6	10.5	12.3	9.7	8.07	**7.74**
	5	14.4	14.3	9.76	10.4	8.91	**7.9**
RIS [45] ↓	1	0.402	0.186	0.145	0.242	0.0929	**0.0817**
	2	0.114	0.0765	0.0925	0.112	0.0918	**0.075**
	3	0.114	0.0782	0.113	0.0643	0.0396	**0.0257**
	4	0.0962	**0.0392**	0.127	0.106	0.0873	0.0804
	5	0.159	0.114	0.0566	0.130	0.0708	**0.0547**

Future work

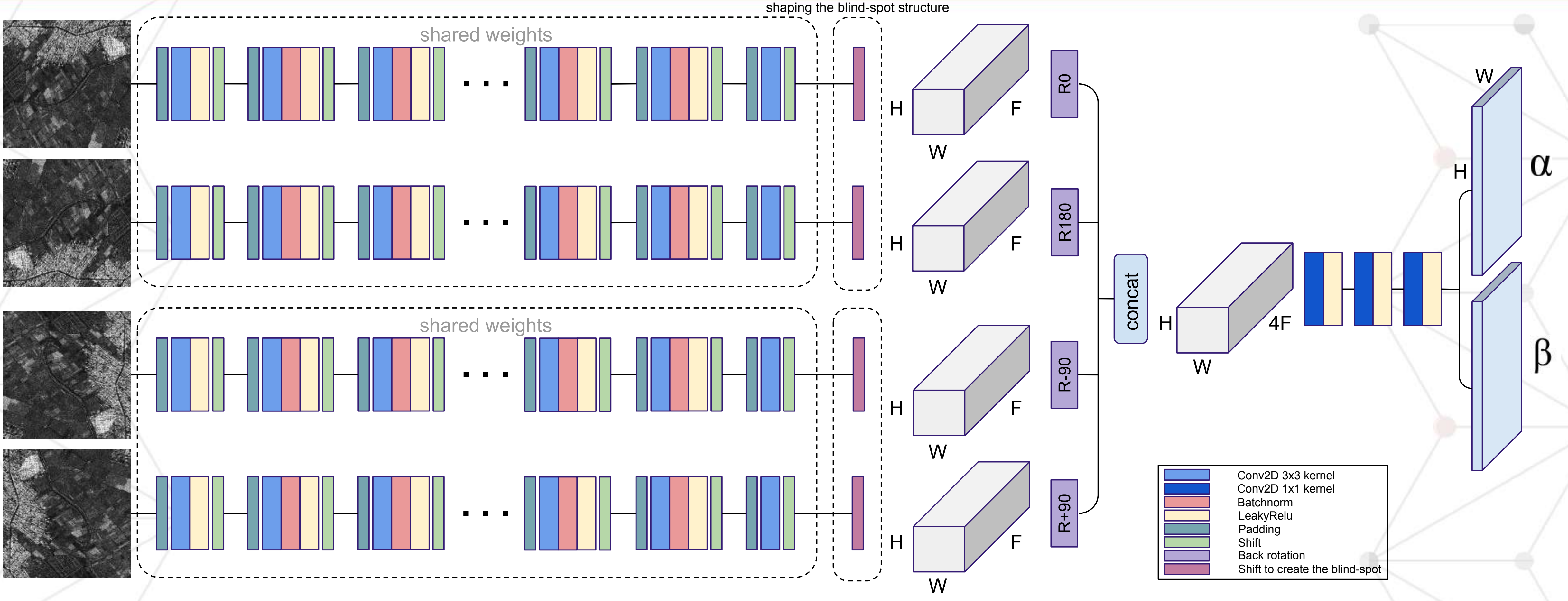
- Devising a method to get rid of the noise correlation:
  - fully incorporated in the blind-spot network;
  - End-to-end trainable by adding a contrastive component in the current loss.

www.ipl.polito.it

Follow us on Twitter: @IPLPolito

Thank you!

Speckle2Void Architecture



Speckle2Void Architecture