ADMM DIP-TV: combining Total Variation and Deep Image Prior for image restoration

Pasquale Cascarano, *Andrea Sebastiani, Elena Loli Piccolomini,*

University of Bologna

Maria Colomba Comes, Eugenio Martinelli,

University of Rome Tor-Vergata

Warsaw, 15 December 2020

BOS/SOR2020 Conference

Session III: Advances in optimization techniques for machine learning

Outline

1. Imaging Inverse Problems
2. Variational approach
3. Tips on supervised learning
4. Deep Image Prior (DIP)
5. The DIPTV framework: ADMM-DIP TV
6. Results and final remarks

Imaging Inverse Problems

- Imaging inverse problems attempt to retrieve an unknown data from its measurement.
- The model we refer reads:

$$\boxed{Ax + \eta = b} \tag{1}$$

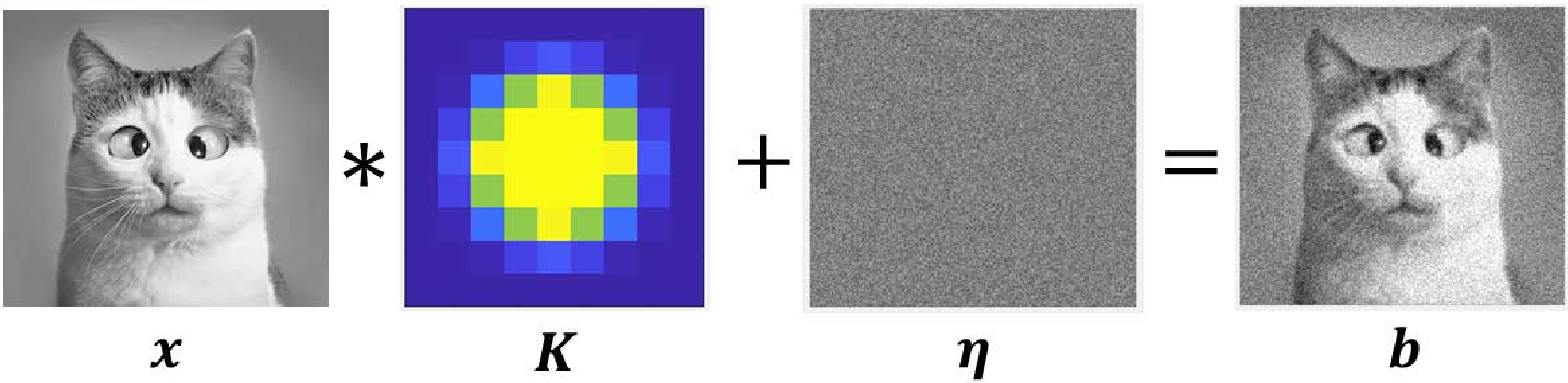
- The linear operator $A \in \mathbb{R}^{l \times n}$ is the forward operator, $x \in \mathbb{R}^n$ is the unknown data, $b \in \mathbb{R}^l$ is the measurement and η is the noisy component.
- Solving (1) means inverting the process:

$$\boxed{b \longrightarrow x}$$

- Different choices for $A \in \mathbb{R}^{l \times n}$ lead to different image restoration/reconstruction tasks.

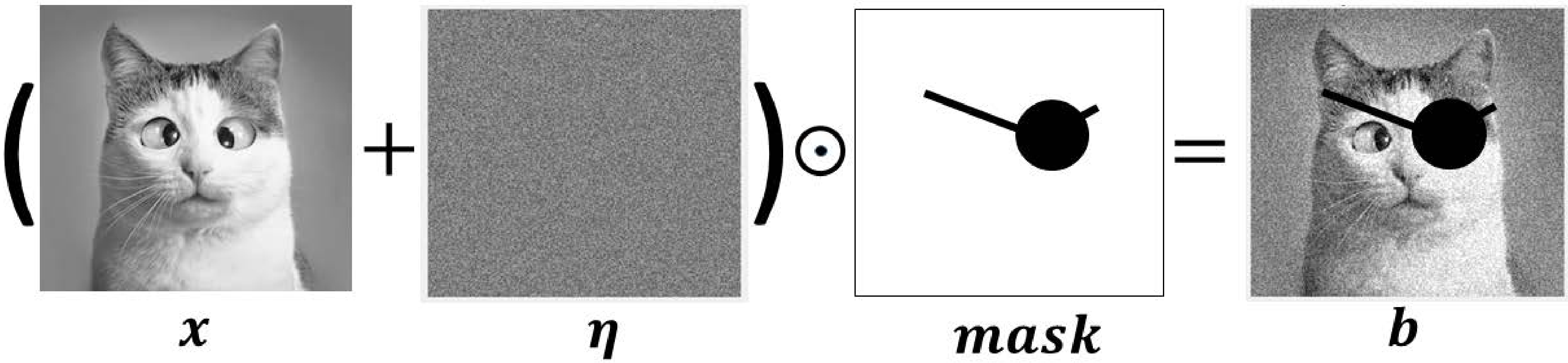
Deblurring

- The matrix A is the discretization of a convolution.



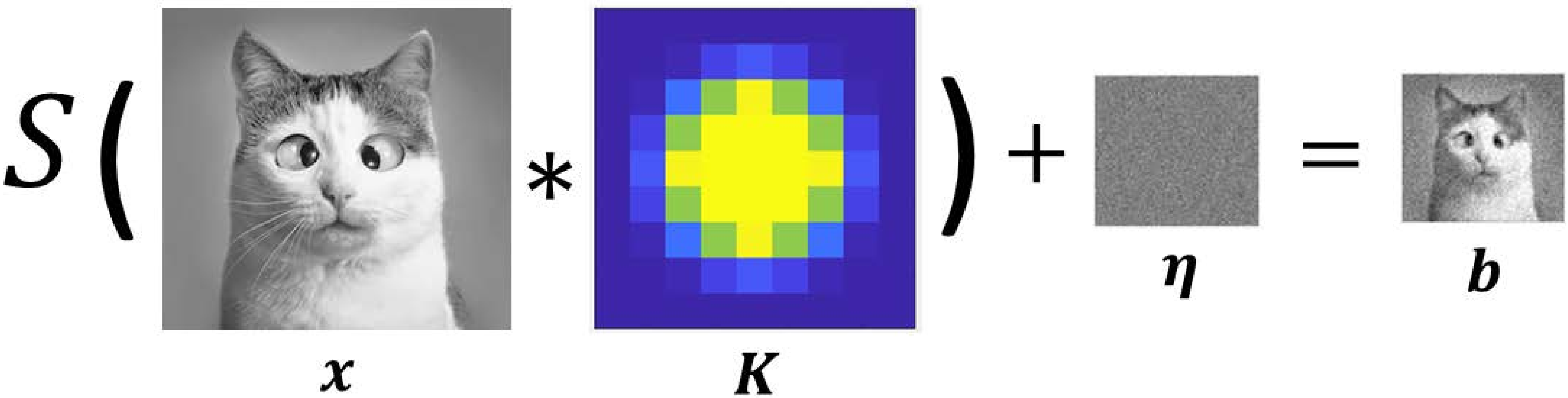
Inpaiting

- The matrix A is the discretization of the Hadamard product.



Super Resolution

- The matrix A is the discretization of the composition between a downsampling and a convolution operator.



Variational approach

- Imaging inverse problems are usually ill-posed problems which means that the properties of existence, uniqueness and stability of the solution are not all verified

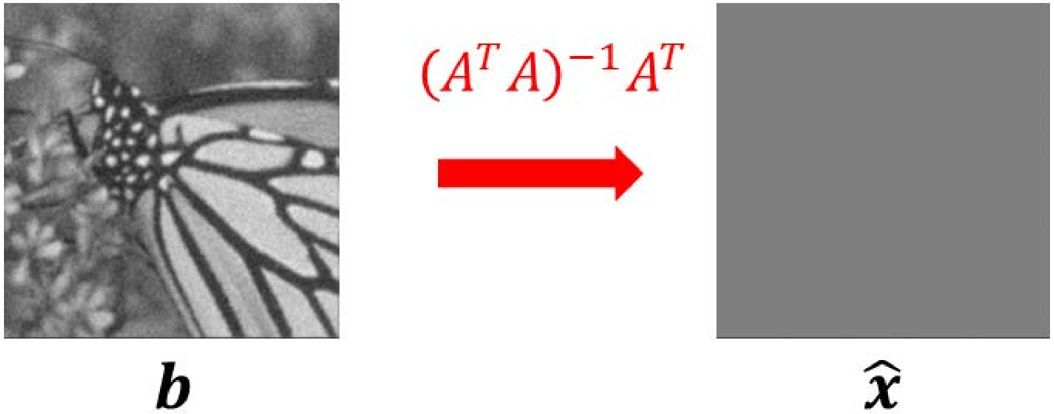
- They are re-casted as an optimization problem of the following form:

$$\boxed{\hat{x} = \arg\min_{x \in \mathbb{R}^n} \{f(x) + \lambda \phi(x)\}} \qquad (2)$$

where f encodes the information on the noise, ϕ reflects prior information on the desired solution and λ is a trade-off parameter.

Under Gaussian Noise assumptions and if no prior is defined, (2) reads:

$$\arg\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - b\|_2^2$$



Total Variation

- The prior is usually defined by assuming some geometric and smoothness properties.

- One of the most famous prior is the Total Variation defined as follows:

$$TV(x) = \sum_{i=1}^{n} \sqrt{(D_h x)_i^2 + (D_v x)_i^2} \longrightarrow \text{images are piecewise constant}$$

$$\boxed{\hat{\mathbf{x}}_{\mathbf{TV}} = \arg\min_{\mathbf{x}\in\mathbb{R}^{\mathbf{n}}} \frac{1}{2}\|Ax - b\|_2^2 + \lambda TV(x)}$$

$\boldsymbol{b}$

$\hat{x}_{TV}$

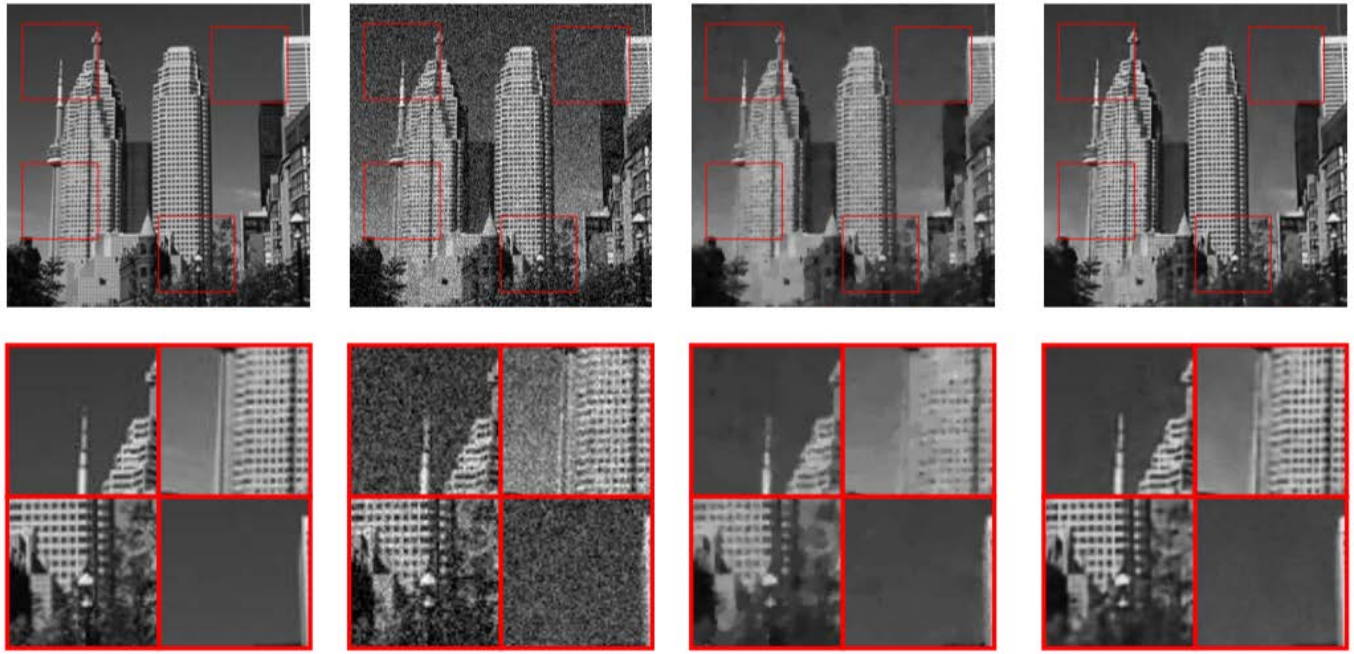
TV Prior

PROS

- Closed form.
- Strong mathematical foundations.
- Convergence guaranteed.

CONS

- Lack of flexibility.
- Complexity of image statistics is only partially reflected.



GT noise+blur TV Learning based

Tips on supervised learning

Ingredients:

- A set of example pairs (*training set)*:

$$(\mathbf{b_1}, \mathbf{x_1}), \ldots, (\mathbf{b_N}, \mathbf{x_N}) \in \mathbf{B} \times \mathbf{X}$$

- An unknown target function which interpolates the training set:

$$\mathbf{h} \in \mathcal{H} := \{\mathbf{h} | \mathbf{h} : \mathbf{B} \to \mathbf{X}\} \quad \text{s.t.} \quad \mathbf{h}(\mathbf{b_i}) = \mathbf{x_i}$$

- A fixed parametric space (*hypothesis space*):

$$\mathcal{F}_\theta \subset \mathcal{H}$$

- A loss function deduced from a distance defined on $\mathcal{H}$

Tips on supervised learning

Goal:

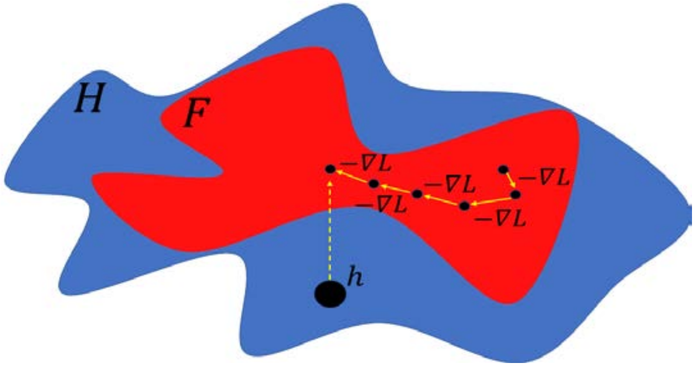
- Approximate the target function by

$$\mathbf{NET}_{(\theta^*,\mathbf{N})} \in \mathcal{F}_\theta, \quad \mathbf{NET}_{(\theta^*,\mathbf{N})} : \mathbf{B} \to \mathbf{X}$$

that is

$$\mathbf{b} \in \mathbf{B} \quad \mathbf{NET}_{(\theta^*,\mathbf{N})}(\mathbf{b}) \approx \mathbf{h}(\mathbf{b})$$

Approximation by gradient flow…

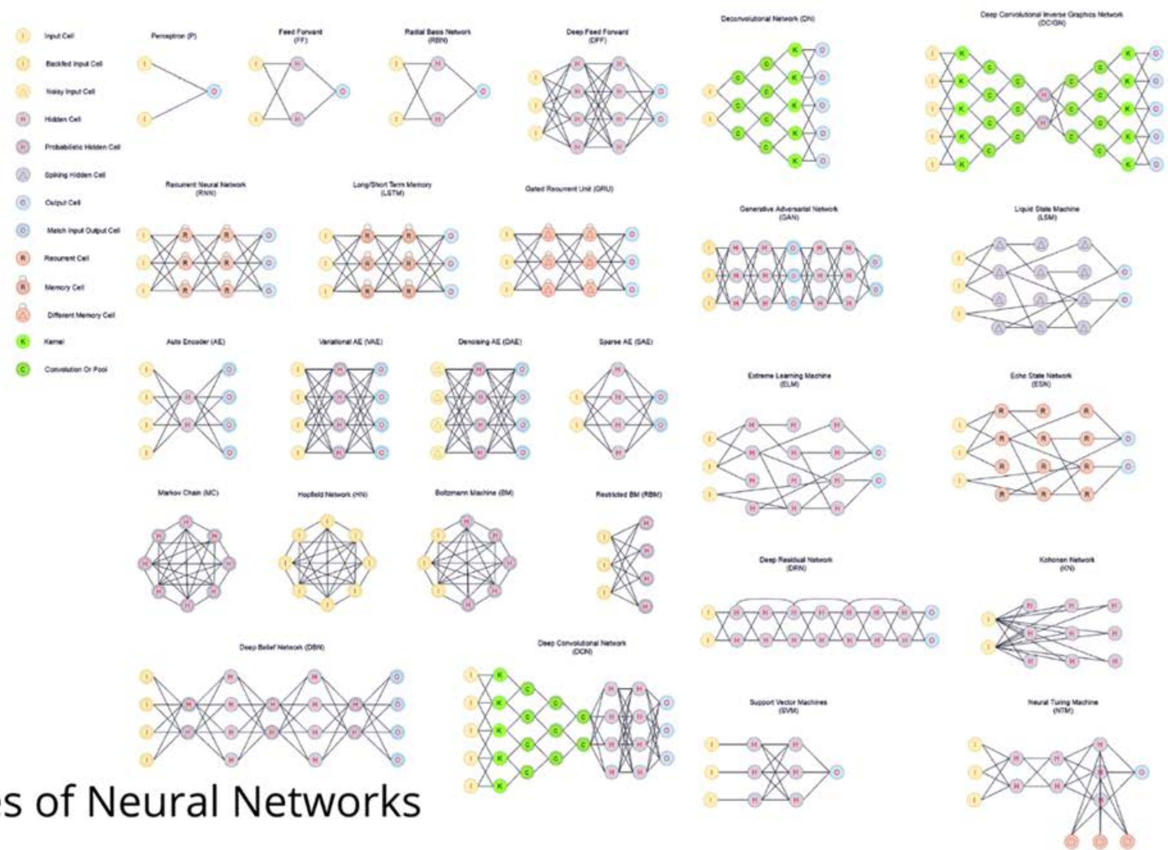


For example: Assuming the standard L_2-norm distance

$$\boxed{\mathbf{NET}_{\theta^*} \in \underset{\mathbf{NET}_\theta \in \mathcal{F}_\theta}{\arg\min} \frac{\mathbf{1}}{\mathbf{N}} \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{N}} \|\mathbf{NET}_\theta(\mathbf{b_i}) - \mathbf{h}(\mathbf{b_i})\|_{\mathbf{2}}^{\mathbf{2}}}$$

A plenty of NETs

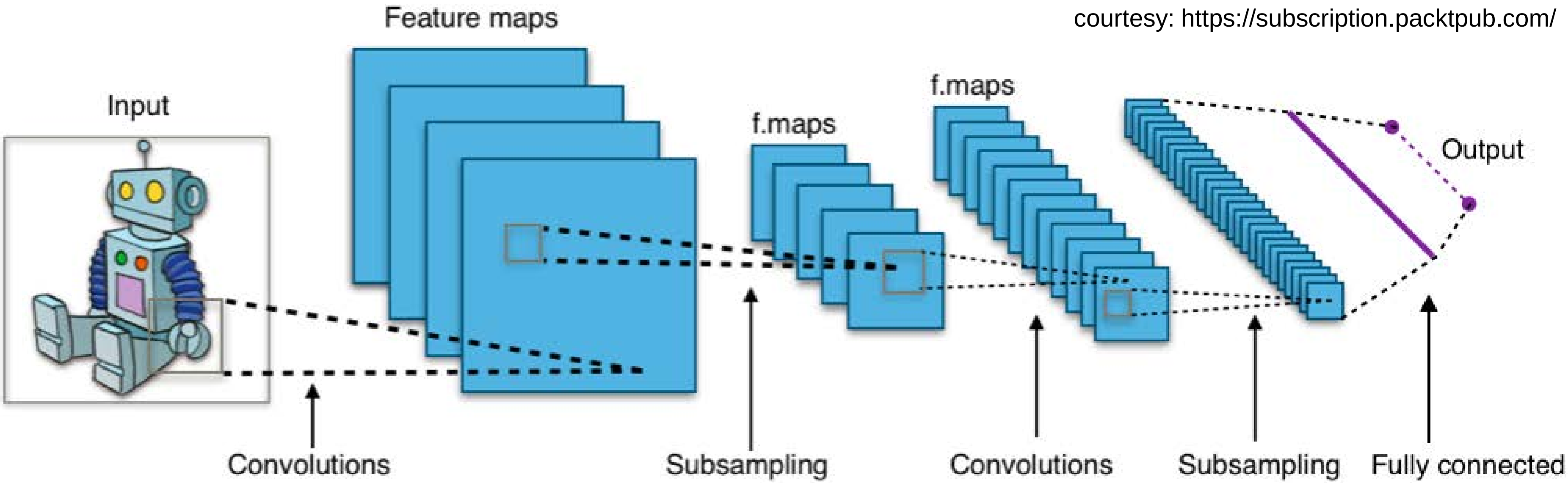
- $\mathbf{NET}_\theta = \mathbf{f}^{\mathbf{1}}_{\theta_\mathbf{1}} \circ \mathbf{f}^{\mathbf{2}}_{\theta_\mathbf{2}} \circ \cdots \circ \mathbf{f}^{\mathbf{t}}_{\theta_\mathbf{t}}$, $\mathbf{f}^{\mathbf{i}}_{\theta_\mathbf{i}}$ is called *layer*



Main Types of Neural Networks

Convolutional Neural Networks

courtesy: https://subscription.packtpub.com/



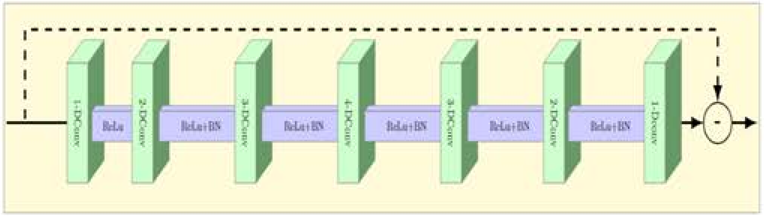
<u>In Convolutional Neural Networks the layers are convolutions</u>

An example: image denosing

- The target is the function which denoises a noisy input.
- The training set is a set of noisy-cleaned image pairs.
- The parametric space is fixed as the set of Convolutional Neural Networks with 7 convolutions, ReLu and Batch Normalization.
- The target is approximated by solving:

$$\mathbf{NET}_{\theta^*} \in \underset{\mathbf{NET}_\theta \in \mathbf{CNN}_\theta}{\arg\min} \frac{\mathbf{1}}{\mathbf{N}} \sum_{\mathbf{i=1}}^{\mathbf{N}} \|\mathbf{NET}_\theta(\mathbf{b_i}) - \mathbf{x_i}\|_{\mathbf{2}}^{\mathbf{2}}$$

Noisy image $\boldsymbol{b}$



Neural Network Denoiser
NET_{θ^*}

Denoised image $\widehat{\boldsymbol{x}} = \mathrm{NET}_{\theta^*}\ (\boldsymbol{b})$

Learning Pros and Cons

PROS

- The learning process manages a huge number of training data and allows to capture a prior which reflects the complexity of image statistics.
- After training, the computation if really fast.
- Currently the state-of-the-art for several inverse problems related to images even with heavily degraded acquisition.

CONS

- Different image restoration/reconstruction tasks require different deep architectures.
- Unfortunately not all the real applications can count on a large number of training examples

Question

Is there a way to avoid the training set?

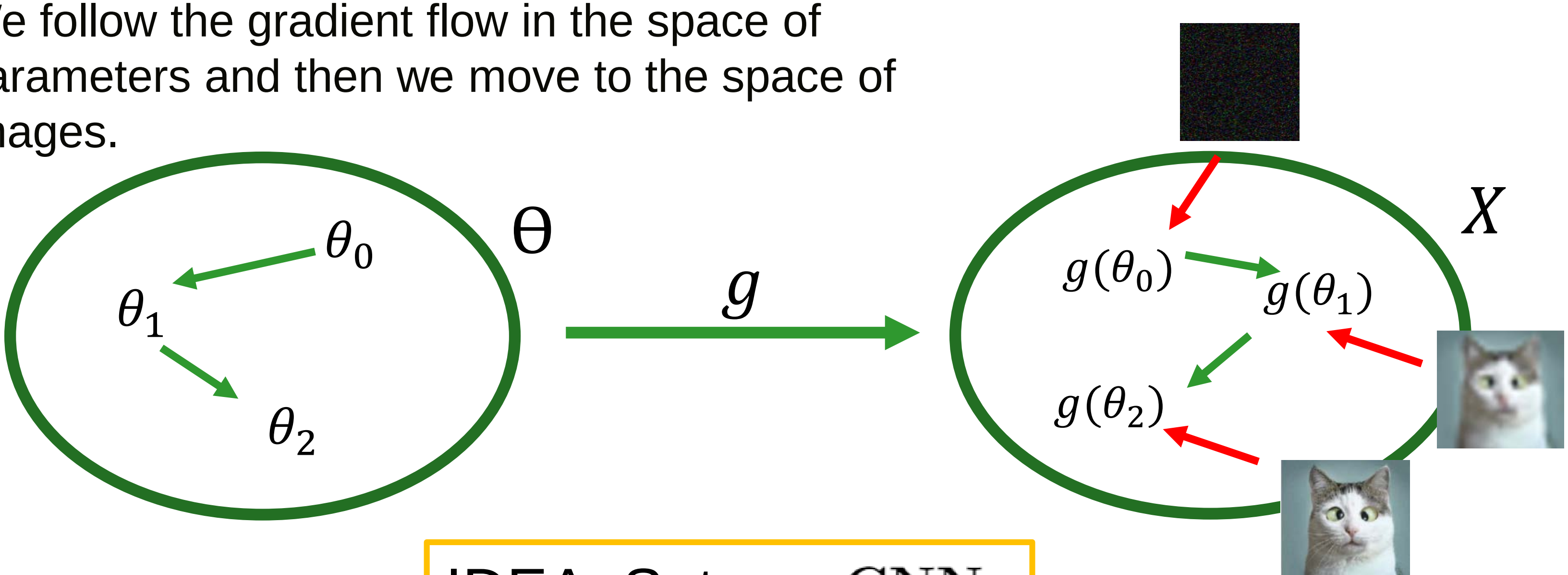
Parametrizing the space of images

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathbb{R}^\mathbf{n}} \frac{1}{2}\|Ax-b\|_2^2 + \lambda R(x) \longrightarrow \theta^* = \arg\min_{\theta\in\mathbb{R}^\mathbf{s}} \frac{1}{2}\|Ag(\theta)-b\|_2^2 + \lambda R(g(\theta))$$

$$\text{s.t. } x^* = g(\theta^*)$$

- The prior is usually defined on a transformed domain, such as wavelet, gradient or overcomplete dictionary.

- We follow the gradient flow in the space of parameters and then we move to the space of images.



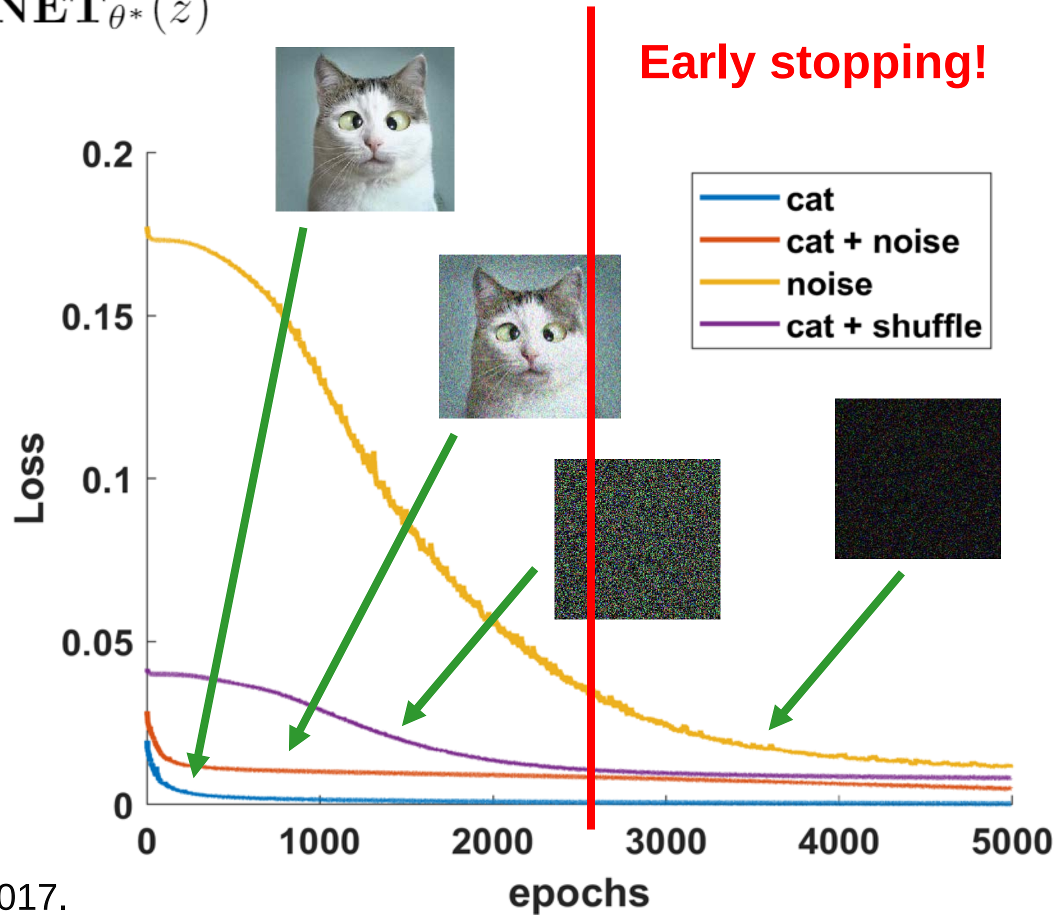
IDEA: Set $\mathbf{g} = \mathrm{CNN}_\theta$

Deep Image Prior (DIP)

$$\mathbf{NET}_{\theta^*} = \underset{\mathbf{NET}_\theta \in \mathbf{CNN}_\theta}{\arg\min} \frac{1}{2} \| A \ \mathbf{NET}_\theta(z) - b \|_2^2$$

$$\text{s.t. } x^* = \mathbf{NET}_{\theta^*}(z)$$

High noise
impedance property



Deep Image Prior, Dimitri Ulyanov et al., 2017.

High noise impedance property

LR noisy baby

GT baby

Stopping criterion needed!

ADMM-DIP TV

- Solving the constrained formulation of DIP TV functional using ADMM.

From an unconstrained formulation…

$$\mathbf{NET}_{\theta^*} = \underset{\mathbf{NET}_\theta \in \mathbf{CNN}_\theta}{\arg\min} \frac{1}{2}\|A\ \mathbf{NET}_\theta(z) - b\|_2^2 + \lambda \sum_{i=1}^{N} \sqrt{(D_h \mathbf{NET}_\theta(z))_i^2 + (D_v \mathbf{NET}_\theta(z))_i^2}$$

To a constrained formulation…

$$\mathbf{NET}_{\theta^*} = \underset{\mathbf{NET}_\theta \in \mathbf{CNN}_\theta}{\arg\min} \frac{1}{2}\|A\ \mathbf{NET}_\theta(z) - b\|_2^2 + \sum_{i=1}^{N} \|t_i\|_2$$

$$\text{s.t. } D(\mathbf{NET}_\theta(z)) = t$$

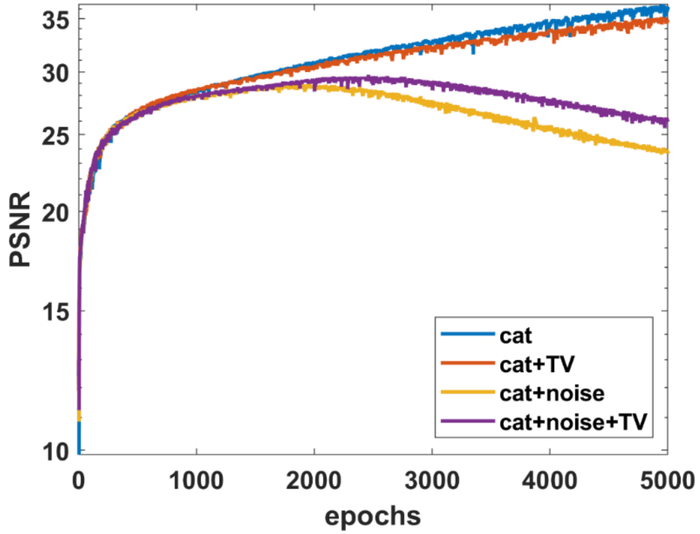
ADMM-DIP TV

$$\begin{cases} \theta^{k+1} \in \arg\min_{\theta} \frac{1}{2}\|A\ \mathbf{NET}_\theta(z) - b\|_2^2 + \frac{\beta_t}{2}\|D(\mathbf{NET}_\theta(z)) - t^k + \frac{\lambda_t^k}{\beta_t}\|_2^2 & (1) \\ t^{k+1} = \arg\min_t \lambda \sum_{i=1}^{N} \|t_i\|_2 + \frac{\beta_t}{2}\|t - (D(\mathbf{NET}_{\theta_{k+1}}(z)) + \frac{\lambda_t^k}{\beta_t})\|_2^2 & (2) \\ \lambda_t^{k+1} = \lambda_t^k + \beta_t(D(\mathbf{NET}_{\theta_{k+1}}(z)) - t^{k+1}) & (3) \end{cases}$$

(1) One step of ADAM

(2) L2 proximity operator

(3) Simple update



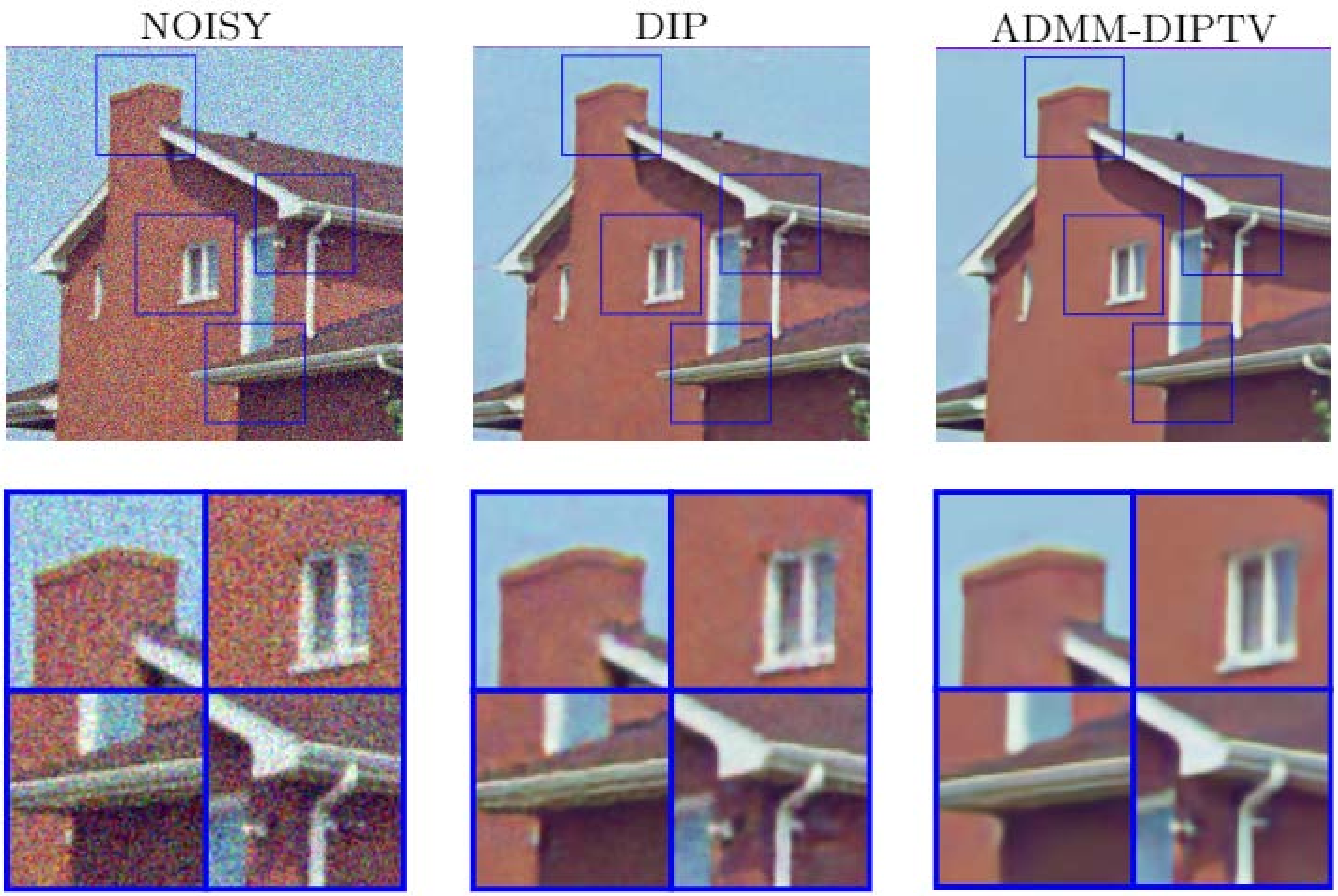
Remark: Changing A in (1) allows to solve different image restoration tasks.

Boosting the standard DIP

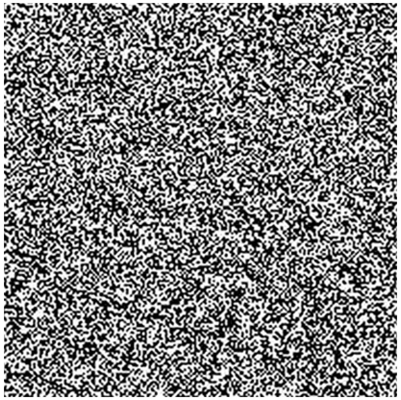
ADMM-DIP TV

DIP

ADMM-DIP TV vs DIP



Inpaiting with ADMM-DIP TV

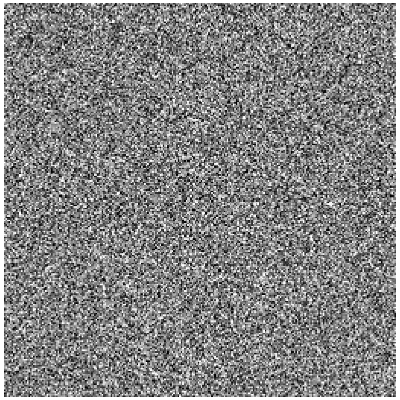



gt

Mask (50%)

gt+mask

ADMM-DIP TV

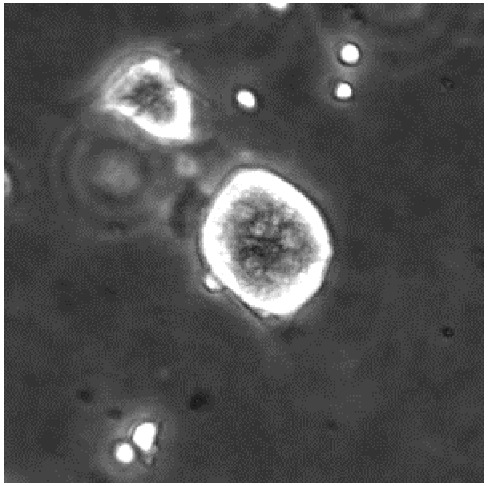
Time-Lapse Microscopy Video

- Time-Lapse Microscopy Videos are sequences of frame with biological contents whose temporal frame rate goes from seconds to minutes, acquired by a TLM microscope.

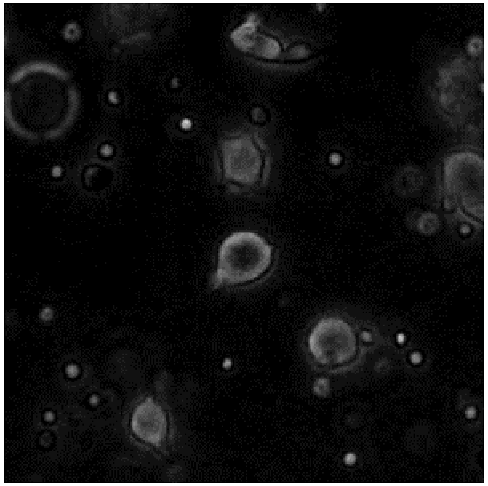
- They are used combined with automated software to track the cells.

- The way in which the cells move, indeed, has been discovered meaningful to understand wound healing, morphogenesis, cancer growth and spread of metastasis.

- For example, some motility descriptors are extracted to uncover and quantitatively evaluate the response to target therapeutic agents

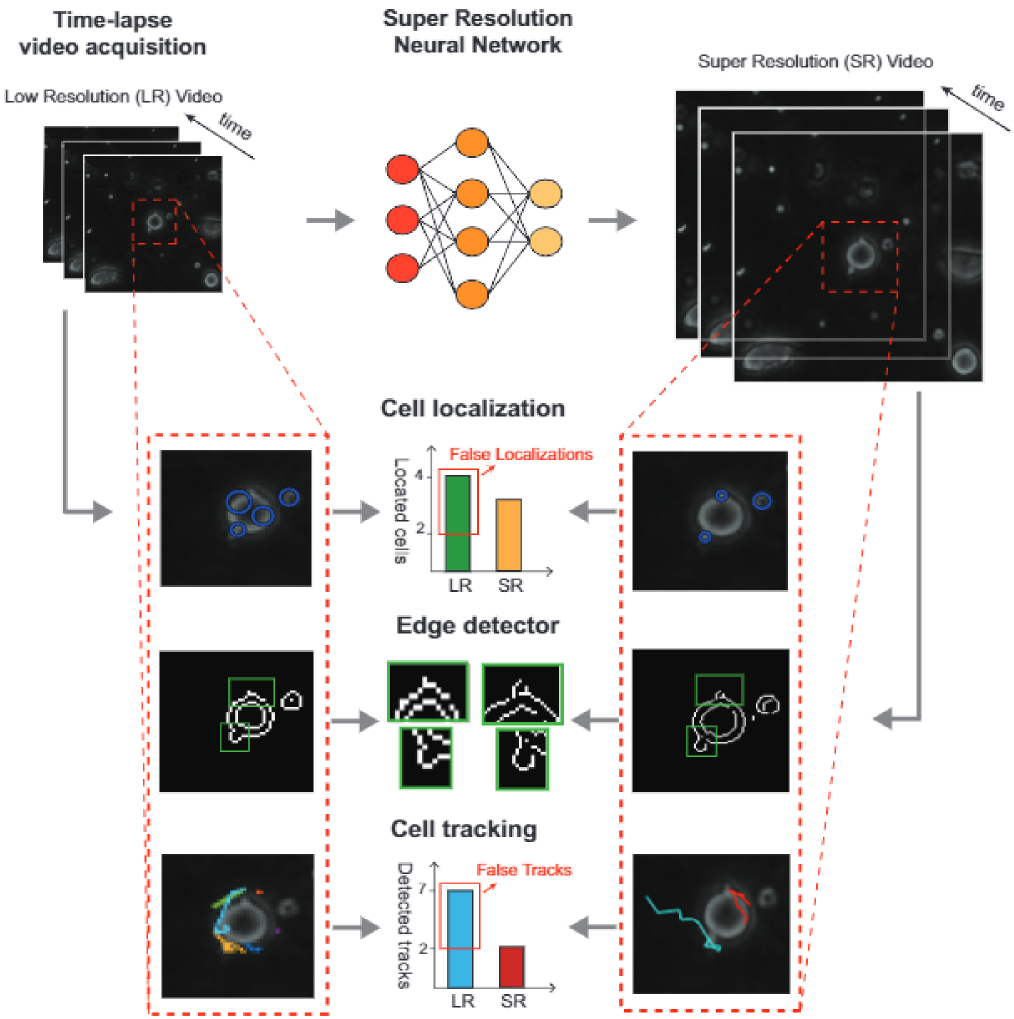


Cell Apoptosis



Breast cancer and immune cells interaction

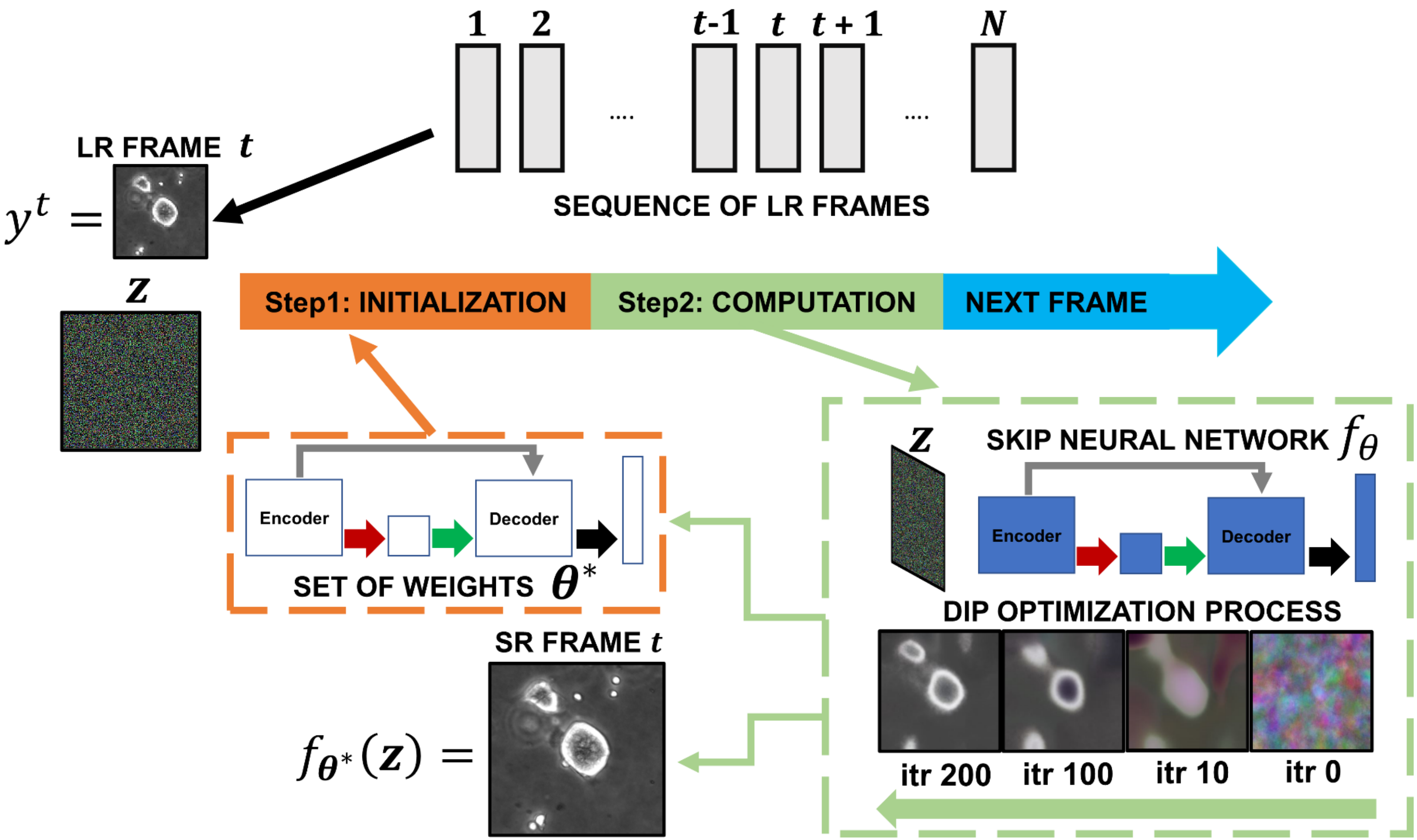
Time-Lapse Microscopy Video



- The resolution of a TLM microscope is limited.
- An high spatial resolution positively affects the trustworthiness of the cell tracking.
- For this application it is not easy to acquire a representative dataset.

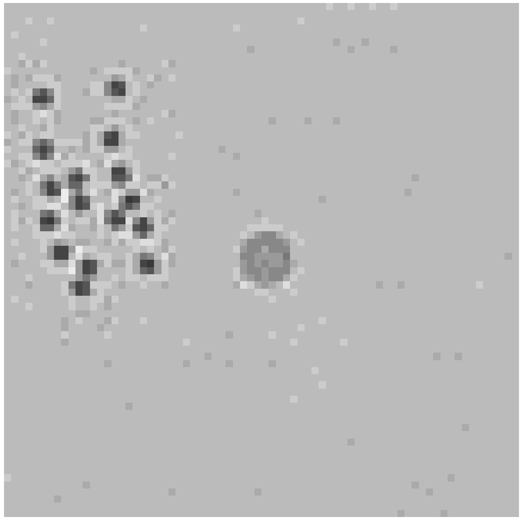
IDEA: Let’s use DIPTV

An extension of DIPTV to Video Super-Resolution

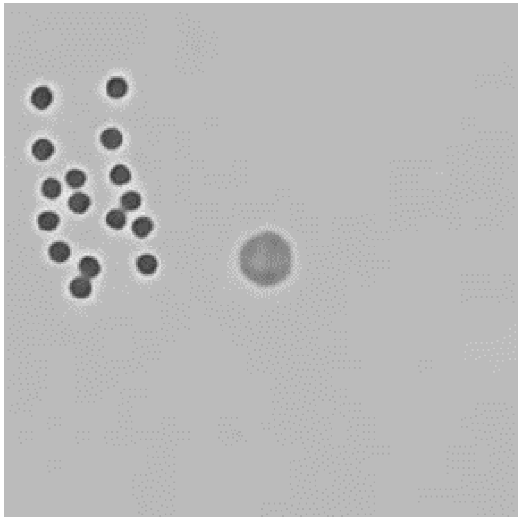


RDPV TV scheme

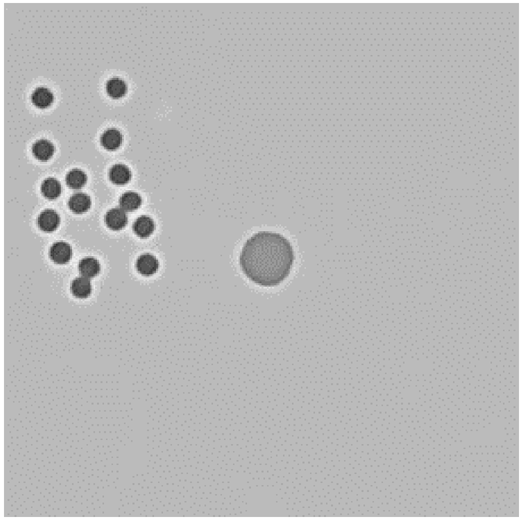
Results on synthetic Cell Video



LR TLM video

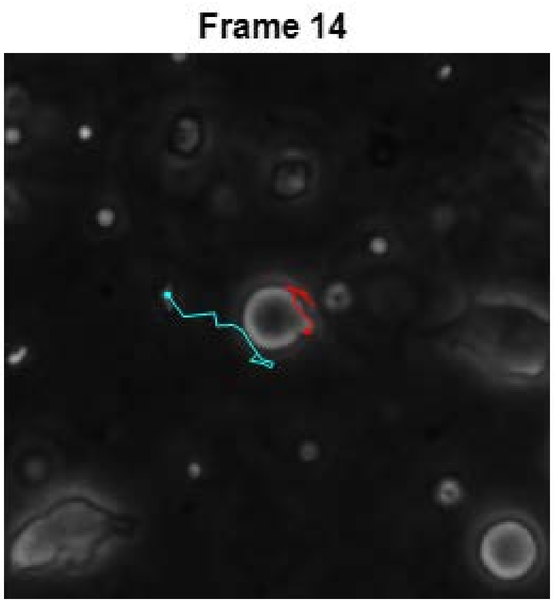


HR TLM video by DIP

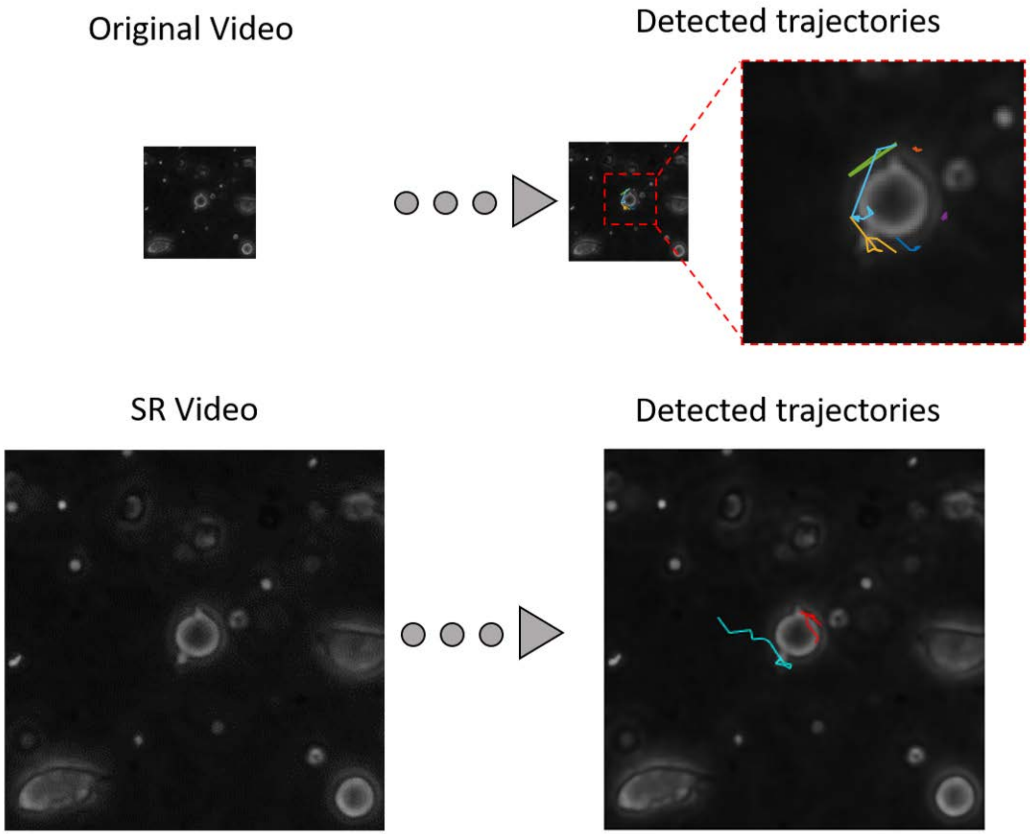


HR TLM video by RDPV-TV

Real Cell Video



HR video trajectories



Computed trajectories on LR and HR videos

Final remarks and future works

- The most striking property of DIP and DIPTV framework is that they do not require any training phase.
- DIP and DIPTV are useful for the solution of generic imaging inverse problems
- ADMM DIPTV leads to better results with respect to the DIP.
- DIPTV framework adapted to Time Lapse Video Microscopy Super Resolution.
- Future works will address the problem of solving the semi-convergence.

References:

1. ADMM-DIPTV: combining Total Variation and Deep Image Prior for image restoration, P.C, A. Sebastiani, M.C. Comes, arXiv.

2. Recursive Deep Prior Video: a Super Resolution algorithm for Time-Lapse Microscopy of organ-on-chip experiments, P.C, M.C. Comes et al., arXiv.

Thank for your attention