Deep-learning for restoration and super-resolution of satellite panchromatic images

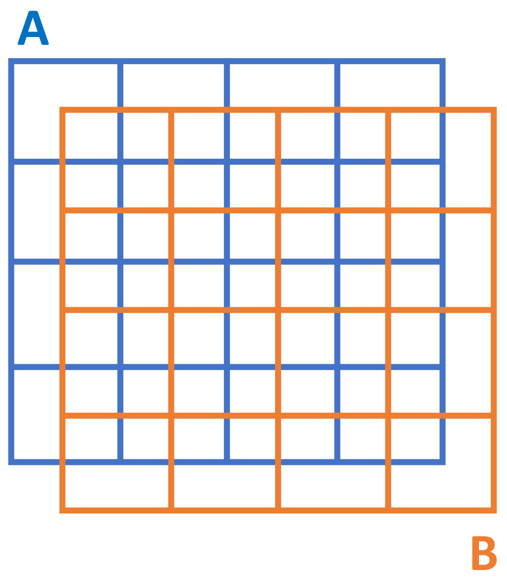
M. Lo Schiavo, E. Magli, D. Valsesia, N. Baccichet, F. Bocchini, L. Giunti, M. Baroni, C. Simoncelli

Politecnico di Torino, Ad Maiorem Italia, s.r.l., Teoresi S.p.A., Leonardo S.p.A.

SUREDOS 2024 - May 30, 2024

Context

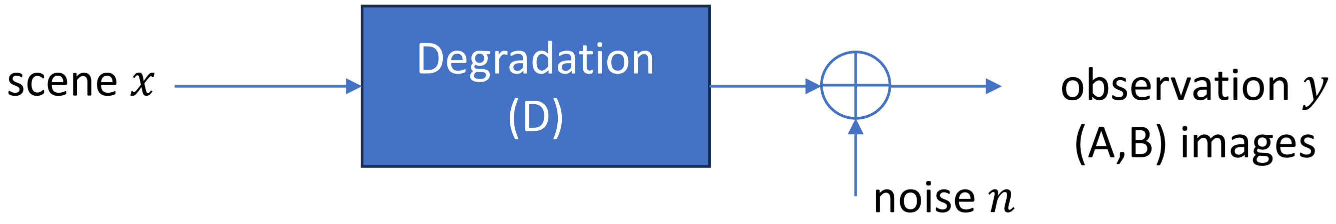
- Activity carried out in collaboration with Leonardo S.p.A.
- Considers a panchromatic imaging sensor at very high resolution, based on PLATiNO-3 VHR mission (ASI)
- The sensor is composed of two individual TDI detectors staggered by 0.5 pixels horizontally and vertically
- How to optimally combine the two images (A and B) at the ground segment (L1 product) to generate a single **2x high-resolution** image?



A and B images are observed through **system PSF**

Problem statement

- Super-resolution problem involves:
  - "interpolation" to increase resolution
  - **deconvolving the PSF** to improve MTF (includes **denoising**)
- Linear degradation model with <u>known</u> degradation operator $\boldsymbol{D}$



- This problem can be solved:
  - via a model-based regularizer, $\arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}\| + \lambda \boldsymbol{R}(\boldsymbol{x})$
  - via a **deep neural network**
    - Datasets to train it?
    - Accuracy and **complexity**? (Output images have **32k x 32k pixels**!)

Approach 1: model-based

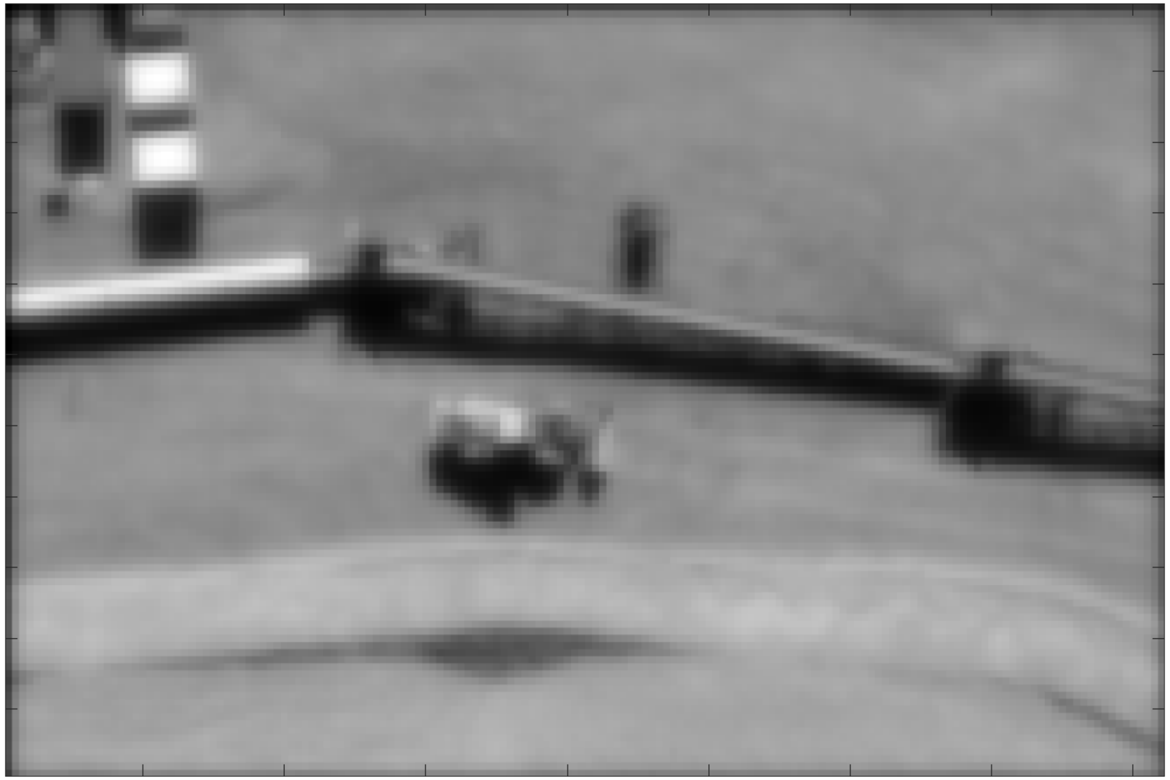
- Method: **denoiser** followed by **deconvolution** using "HyperLaplace prior"
  - Two denoising options: wavelets and **NafNet** deep neural network
  - **Hyper-Laplace prior** does not penalize heavy-tailed distribution of gradients
    - Iterative "alternating projections" method: one projection is done via FFT, the other has analytical solution → relative low complexity
- Input: bicubic interpolation from A and B images
- This method has just one parameter λ that determines the strength of the regularizer

D. Krishnan, R. Fergus, "Fast Image Deconvolution using Hyper-Laplacian Priors", Proc. NIPS 2009
L. Chen et al., "Simple baselines for image restoration", ECCV 2022

Effect of PSF

Image A vs “ground truth” image at target 2x resolution (no PSF/noise)

Image A

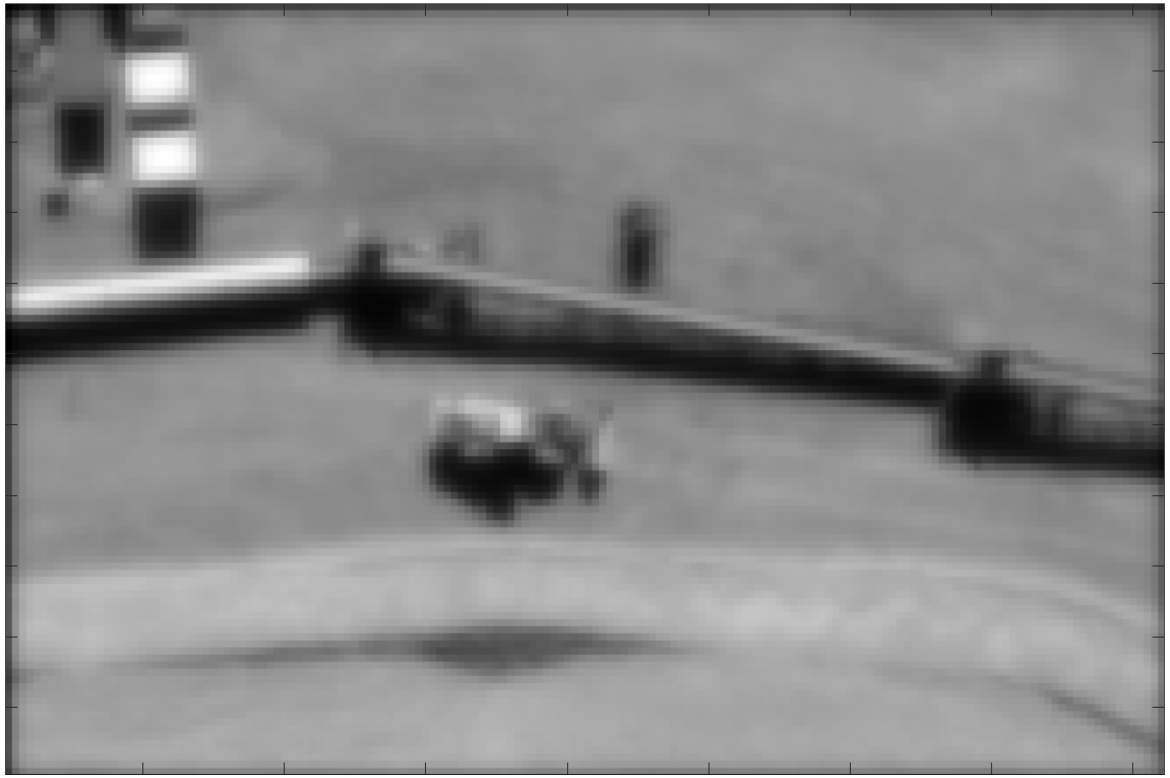


«Ground truth» (GT) image

Combining two images

- Reconstruction using bicubic interpolation

Image A

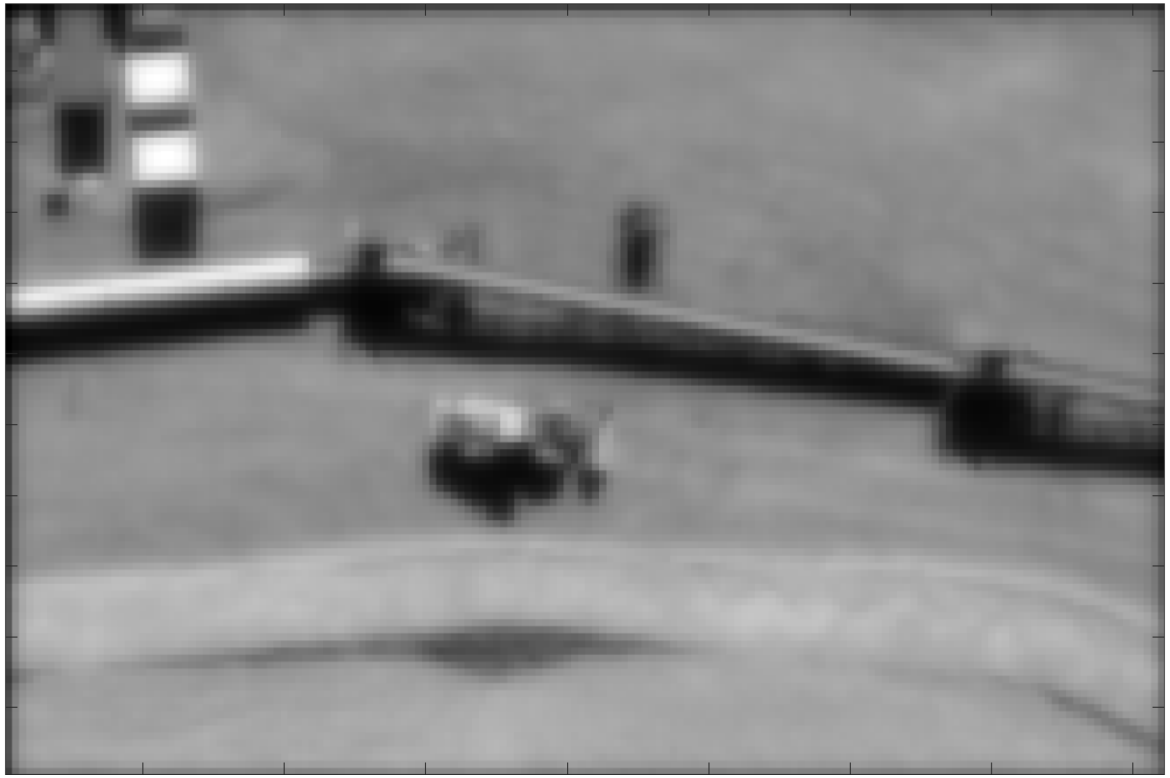


Bicubic interpolation

Combining two images

- Reconstruction using NafNet denoiser + Hyper Laplace deconvolution

Image A



Deconvolution result

Speed and memory

- Reconstruction of a 32000x32000 image divided into overlapping tiles
  - CPU: AMD Ryzen 7 3700X (8C/16T, 3.6/4.4 GHz base/boost clocks)
  - RAM: 64GB DDR4 3200MHz
  - GPU: Nvidia Quadro RTX 6000 (Turing generation, 24GB VRAM)
- NAFnet Denoising + HyperLaplace Deconvolution (CPU only)
  - **NAFnet denoising runtime: 59 min 24 sec**
  - **Deconvolution runtime: 6 min 27 sec**
- NAFnet Denoising + HyperLaplace Deconvolution (GPU+CPU)
  - **NAFnet denoising runtime: 2 min 4 sec**
  - **Deconvolution runtime: 6 min 27 sec**

Approach 2: supervised deep learning

- For deep learning we need a dataset…
- We do not have a paired datasat of low- and high-res images
  - Train an image restoration network on a **large dataset of satellite images** (e.g., Sentinel, SPOT, Landsat, …)
    - We use the available images as if they were high-res, and simulate PSF and downsampling
    - This assumes that the learned upscaling process is scale-invariant
  - **Apply directly** to target images, or…
  - **Fine-tune** the network using a small dataset of target images (or their likes) if available
- Open issues: Effect of domain gap between training and test images
- Selected architecture: NafNet
  - Input: interleaved A and B images with missing pixels at zero

Training process

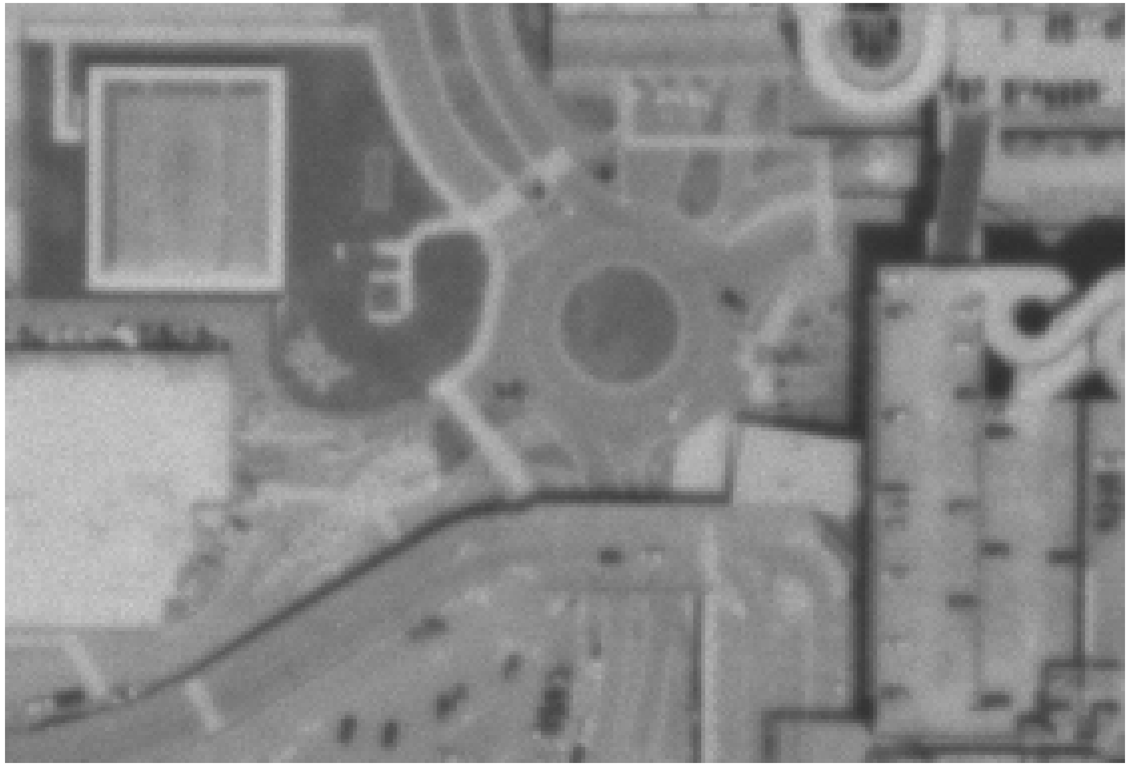
- **Datasets**:
  - **DIV2K**: 800 natural images, various contents and pixel resolutions
  - **USGS** Landsat (+ **Hexagon** aerial images for finetuning):
    - 776 images @30 m after 8x augmentation (mirroring, rotation)
    - 80 resampled Hexagon images after 8x augmentation
  - **WorldStrat**: 3924 images, 1.5m pixel resolution (SPOT 6/7)
- **Patch size**: (192, 192)

Test images

- Left: GT airport image, no PSF/noise
- Right: GT Hexagon image

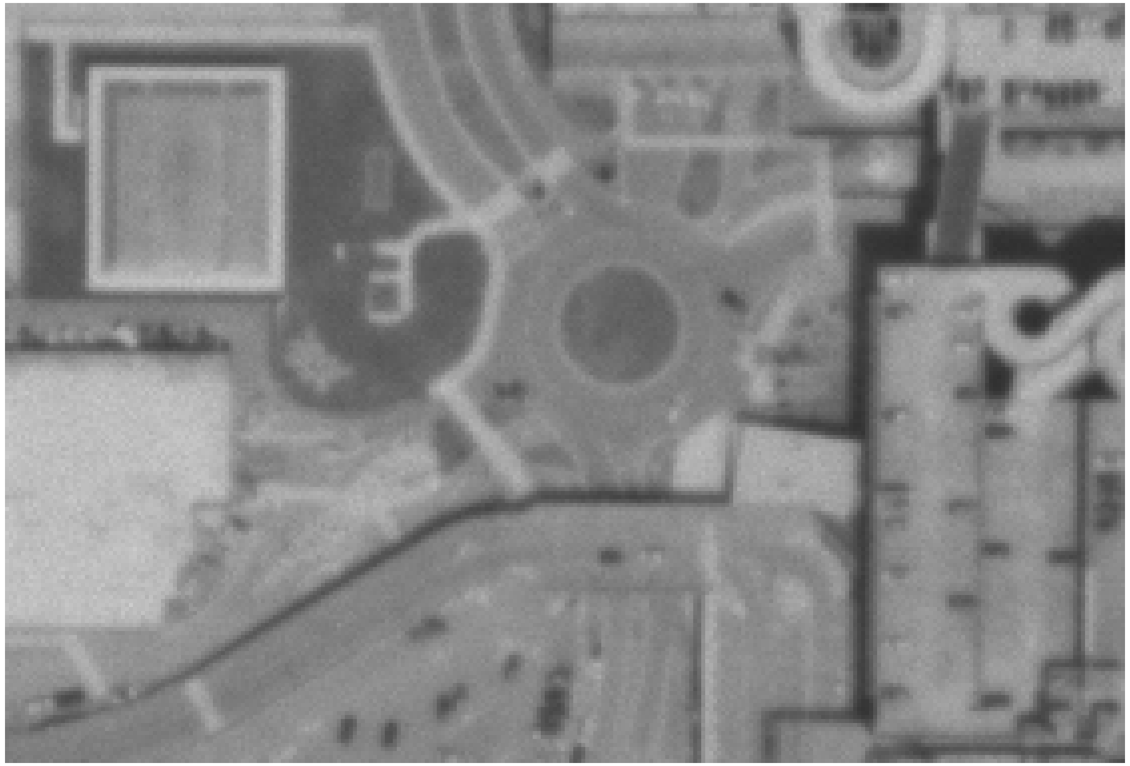
Test images

- Left: GT airport image, no PSF/noise
- Right: airport image “A”



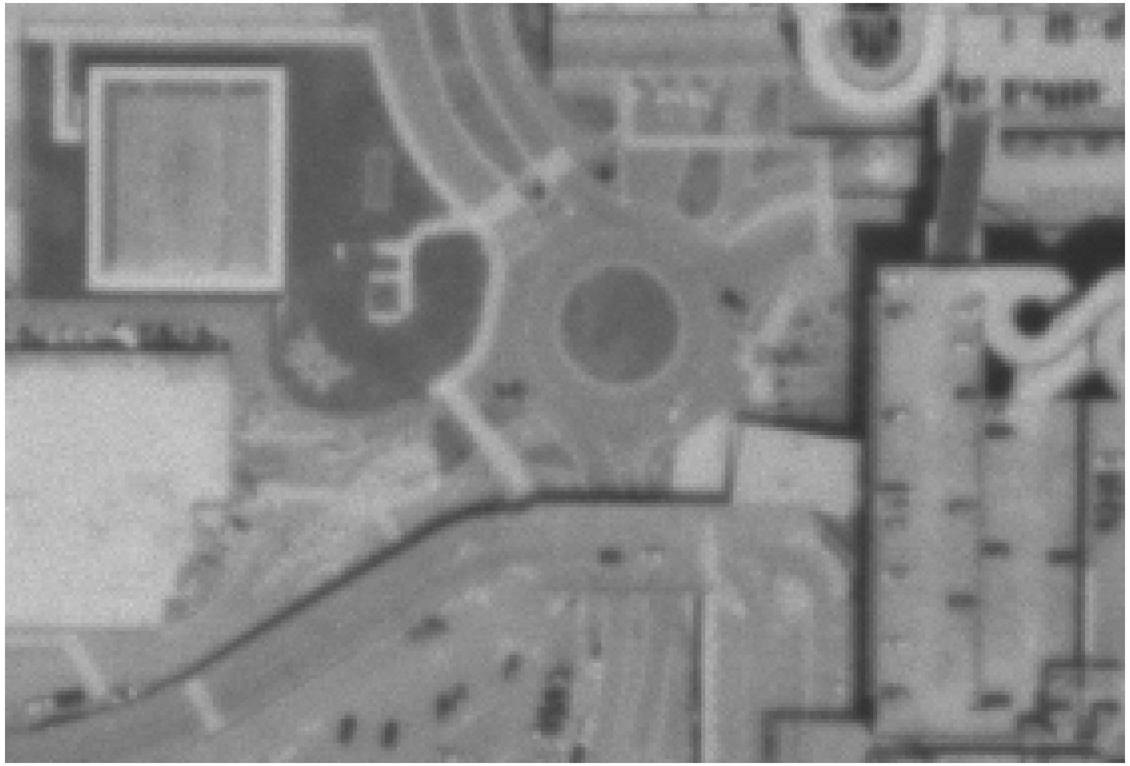
Sample results

- Left: image A
- Right: model-based deconvolution



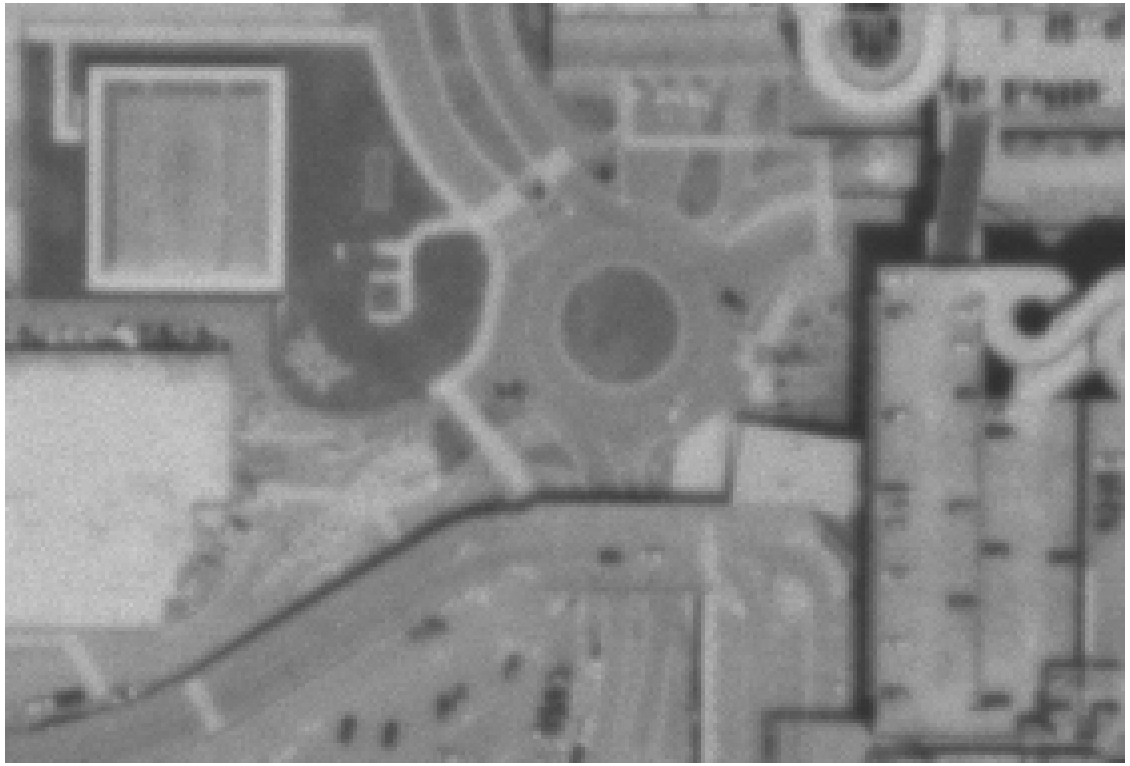
Sample results

- Left: image A
- Right: DIV2K



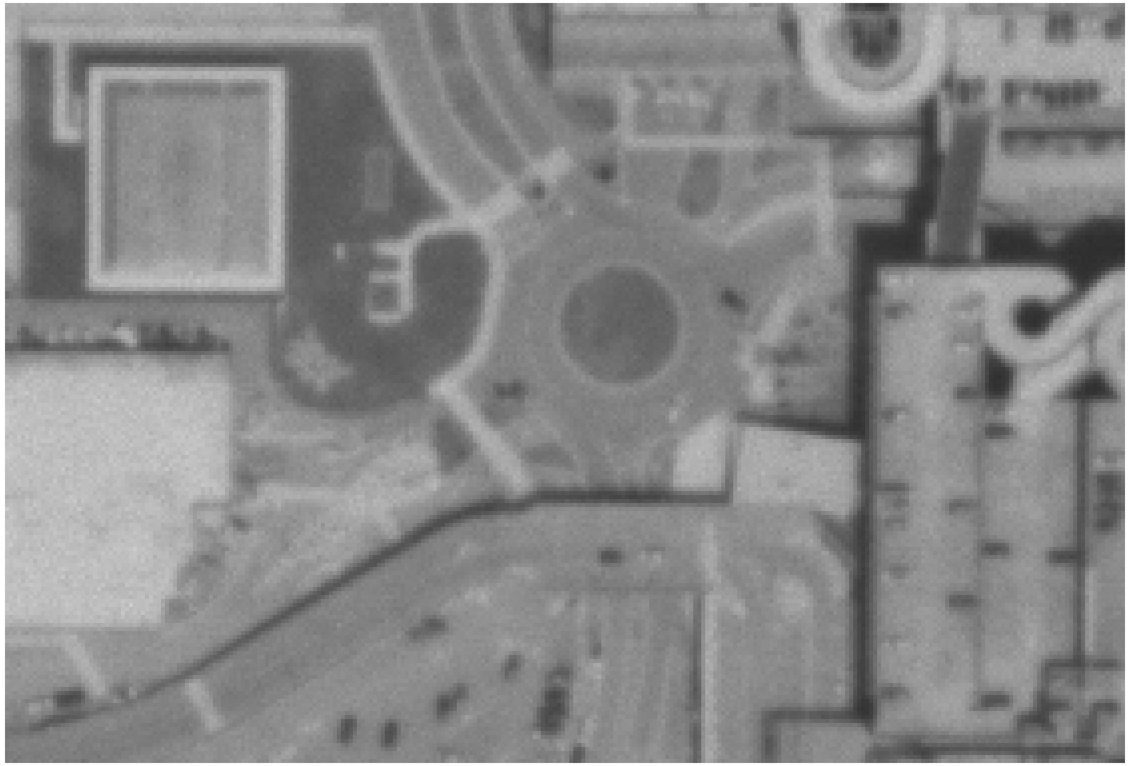
Sample results

- Left: image A
- Right: USGS+Hexagon



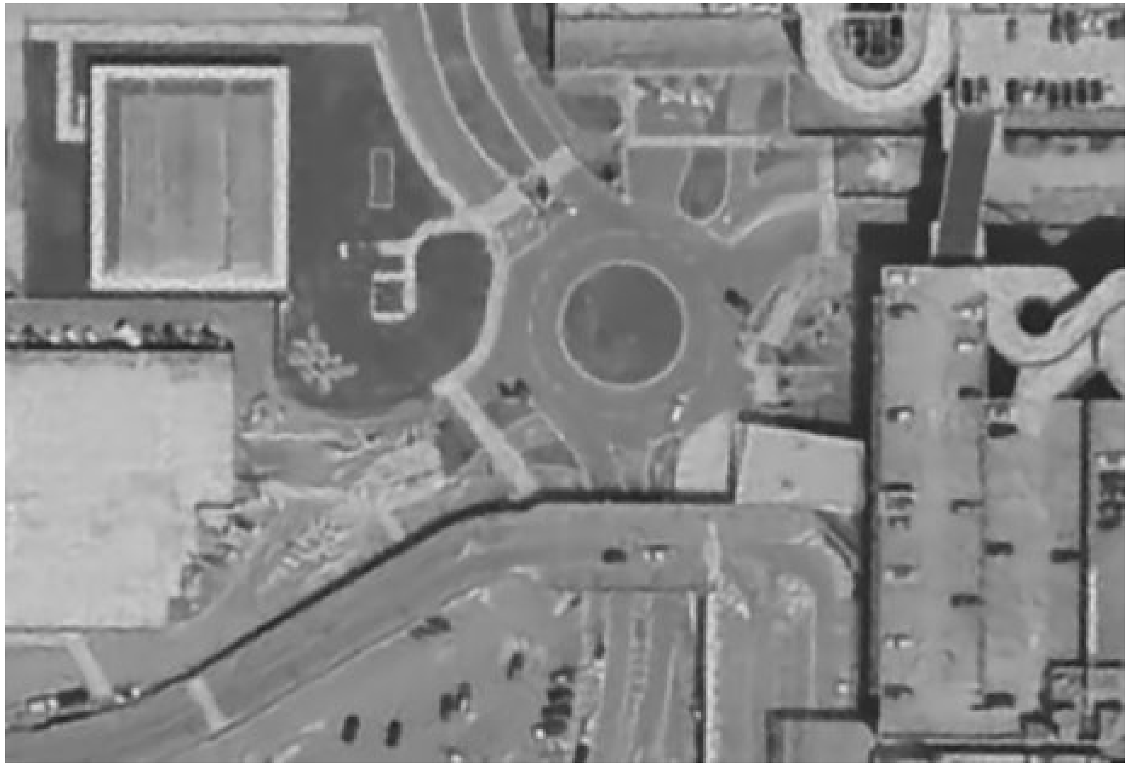
Sample results

- Left: image A
- Right: WorldStrat



Sample results

- Left: USGS, no fine-tuning
- Right: USGS+Hexagon



Sample results

- Left: model-based deconvolution
- Right: Worldstrat

Quality metrics

Metrics computed with respect to ground truth image

- **SNR**: signal-to-noise ratio between reconstructed and ground truth image, $SNR(x,y) = \frac{\sum_i x_i^2}{\sum_i (x_i - y_i)^2}$
- $\boldsymbol{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$
- Median Absolute Error: $\boldsymbol{MAE} = median_i(|x_i - y_i|)$
- Median Relative Error: $\boldsymbol{MRE} = median_i\left(\frac{|x_i - y_i|}{|x_i|}\right)$

Metric results - airport

- Deep learning methods clearly show better metrics than model-based methods
- This is consistent with the visual appearance of the restored images
- Very similar results on Hexagon image

	SNR	MAE	SSIM	MRE
Model-based method	20.421	88.348	0.684	4.303
NAFnet - DIV2K	21.749	70.547	0.754	3.477
NAFnet - USGS	22.005	**58.784**	**0.789**	**2.903**
NAFnet - USGS+Hexagon	21.940	60.068	0.784	2.950
NAFnet - WorldStrat	**22.067**	65.558	0.775	3.214

Conclusions

- **Supervised deep learning methods** are significantly better at increasing image contrast than model-based ones
  - Results are highly dependent on the training process
  - Their visual quality is better
  - Their accuracy and sharpness are better
  - Their running time is lower (because they can be accelerated on GPU)
- There is always a trade-off between **sharpness** and **noise/artifacts**
  - Using high-resolution images in the training set typically yields sharper images (Worldstrat, USGS+Hexagon)
  - Even the “less sharp” deep learning results are better than that achieved by model-based methods
- Results would be even better if the method could be trained on a large dataset of images similar to the target, or a paired high- and low- res dataset