Learning Multi-Scale Local Conditional Probability Models of Images (top 25%)

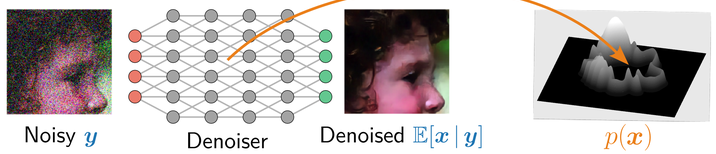
Zahra Kadkhodaie[1] Florentin Guth[2] Stéphane Mallat[3,4] Eero P. Simoncelli[4,5]

[1]CDS, New York University, zk388@nyu.edu [2]DI, ENS, CNRS, PSL University, Paris, France [3]Collège de France, Paris, France [4]Flatiron Institute, New York, USA [5]CNS, Courant, and CDS, New York University

Summary

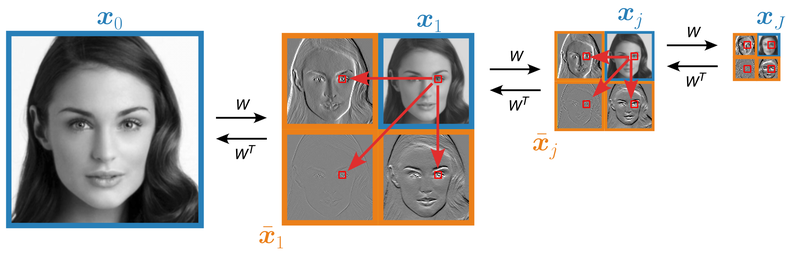
- How do score-based models manage to overcome the curse of dimensionality?
- Networks require global receptive fields to capture global spatial dependencies
- We show that these **global dependencies become local** after a multiscale factorization of the probability distribution
- We obtain high-resolution denoising, super-resolution, and synthesis results with **local conditional networks at each scale**

Denoising and Score-Matching



- A denoiser is a tool to learn priors: $\mathbb{E}[\boldsymbol{x} \mid \boldsymbol{y}] = \boldsymbol{y} + \sigma^2 \nabla \log p(\boldsymbol{y})$
  (Miyasawa 1961; Tweedie (via Robbins) 1956; Raphan & Simoncelli 2006; Vincent 2011)
- By using $\nabla \log p(\boldsymbol{y})$, we can synthesize images by doing gradient ascent on the log-probability
- This fails if the denoiser receptive field is smaller than the image size!

Wavelet Conditional Distributions

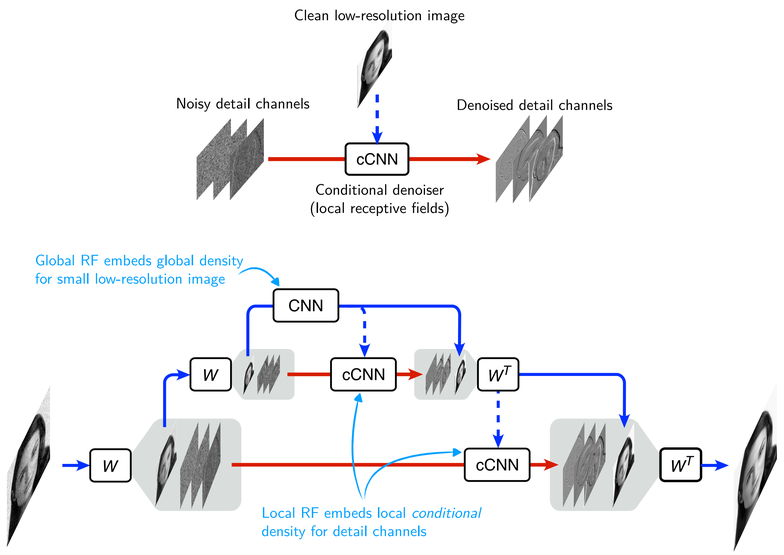


- The probability distribution of the entire image $\boldsymbol{x}_0$ can be factorized as a product of conditional probabilities at each scale:

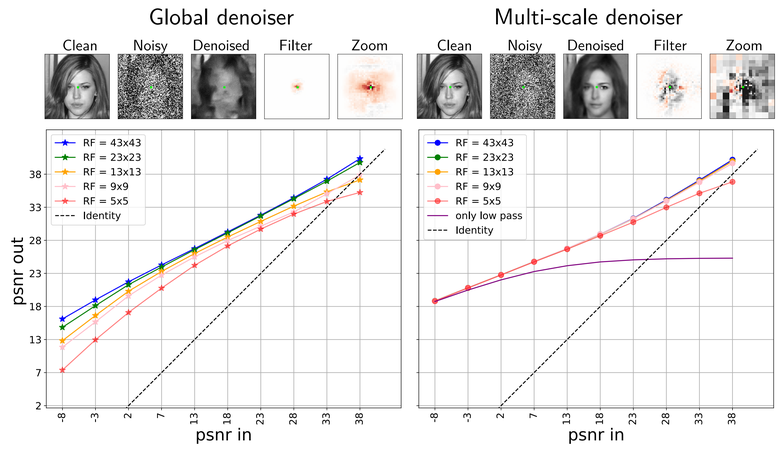
$$p(\boldsymbol{x}_0) = p(\boldsymbol{x}_1, \bar{\boldsymbol{x}}_1) = p(\boldsymbol{x}_1)\, p(\bar{\boldsymbol{x}}_1|\boldsymbol{x}_1) = p(\boldsymbol{x}_J) \prod_{j=1}^{J} p(\bar{\boldsymbol{x}}_j|\boldsymbol{x}_j).$$

- This suggests first generating the lowest-resolution image $\boldsymbol{x}_J$ and iteratively increasing the resolution by conditionally generating details $\bar{\boldsymbol{x}}_j$
- **Theorem:** Restricting the receptive fields of the denoisers is equivalent to enforcing a **Markov property** on $\bar{\boldsymbol{x}}_j$ given $\boldsymbol{x}_j$

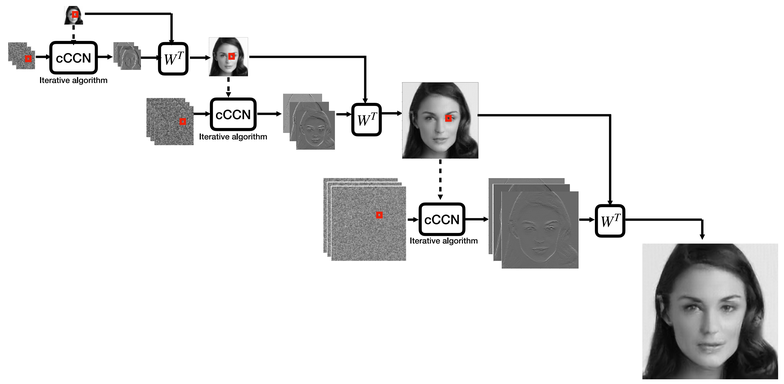
Multi-Scale Local Conditional Denoising



- We evaluate the denoising performance of the conditional denoisers
- **Their receptive fields can be reduced to** 9×9 without harming performance



Super-Resolution and Synthesis



- The conditional denoisers can be used to perform super-resolution and image synthesis
- The global structure is captured with a global prior on the small low-pass image