Image Denoising

… Not What You Think

Michael Elad

Computer Science Department
The Technion - Israel Institute of Technology
Haifa 32000, Israel

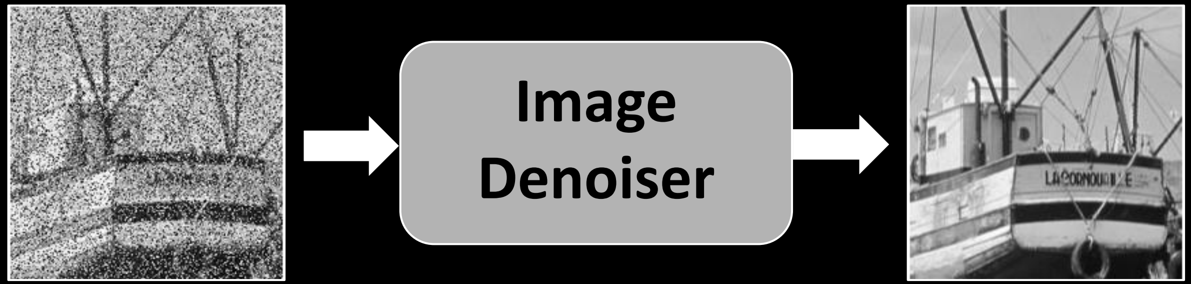
In dedication to **Gene H. Golub**
(February 29, 1932 - November 16, 2007)
The occasion of his 90th birthday

The Third
International Workshop on Matrix Computations

Lanzhou University, Lanzhou, P.R. China
April 15–19, 2022

This Lecture is About …

Image Denoising



Noisy image

Denoised image

Removal of noise from images is a heavily studied problem in image processing

In this talk we expand on recent discoveries and developments around this seemingly dead topic

Fields/Topics appearing in this Talk:

- ❑ Image processing - Restoration
- ❑ Optimization
- ❑ Machine Learning
- ❑ Deep Learning
- ❑ Sparse Representation
- ❑ Linear Algebra
- ❑ Probability Theory & Statistics
- ❑ Stochastic Differential Equations

A Disclaimer:
This talk is self-contained

Michael Elad
The Computer-Science Department
The Technion

Our Agenda

Part I

1. Brief Introduction & History
2. Image Denoising: The Classic Era
3. The Deep Learning Revolution
4. Synergy: Classic + Deep Learning

Part II

5. Our Focus Today: Denoising for …
    - Solving general inverse problems
    - Image Synthesis
    - High perceptual quality recovery
6. Summary

Introduction & History

So, Let’s Talk About …

Image Denoising

or more accurately

Removal of **White Additive Gaussian Noise** from an Image

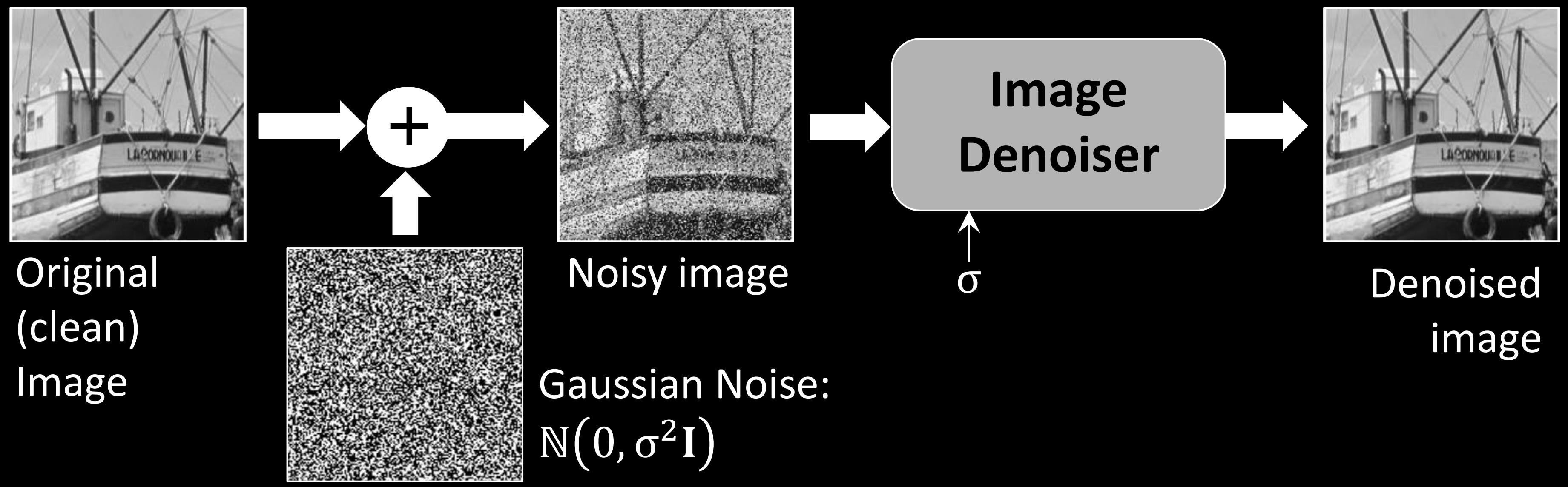
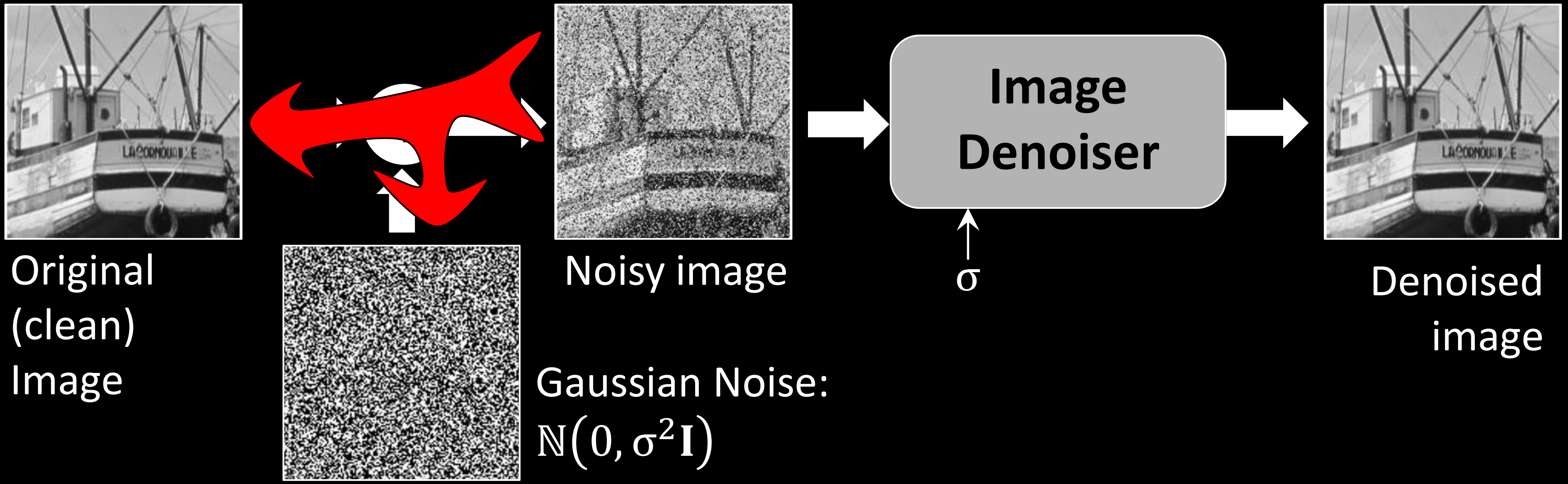


Image Denoising is Challenging

Image denoising is far from trivial task! Why?

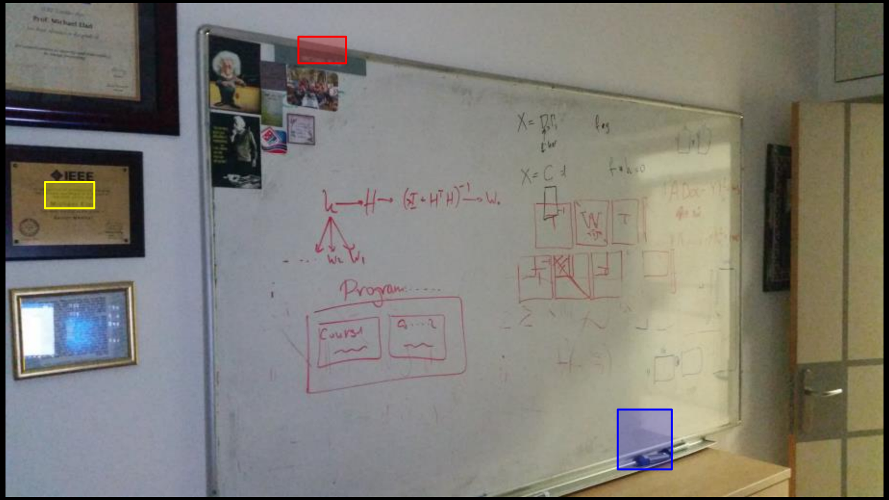
- Because our goal is to remove noise as much as possible while preserving the details in the image
- Denoising is essentially a highly ill-posed separation task

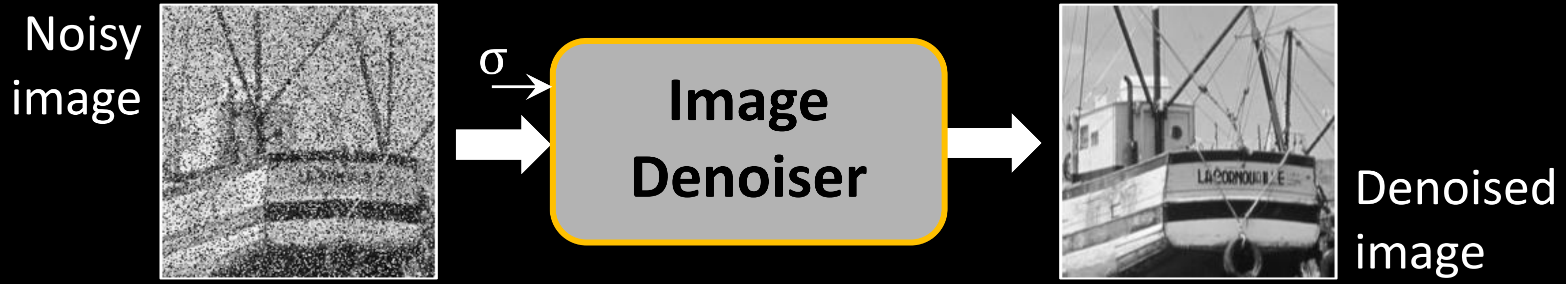


Michael Elad
The Computer-Science Department
The Technion

Why Work on Image Denoising?

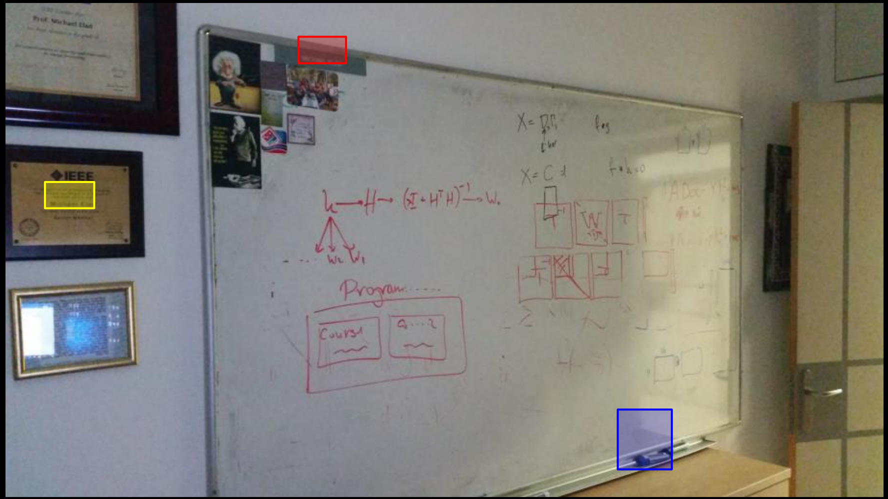
1. **Practical:** It is a real-world problem, arising in all cameras,

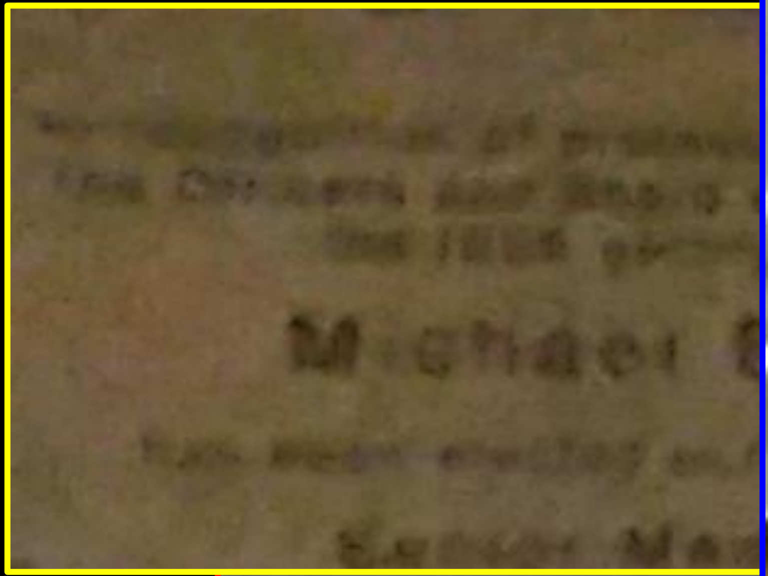




Why Work on Image Denoising?

1. Practical: It is a real-world problem, arising in all cameras,





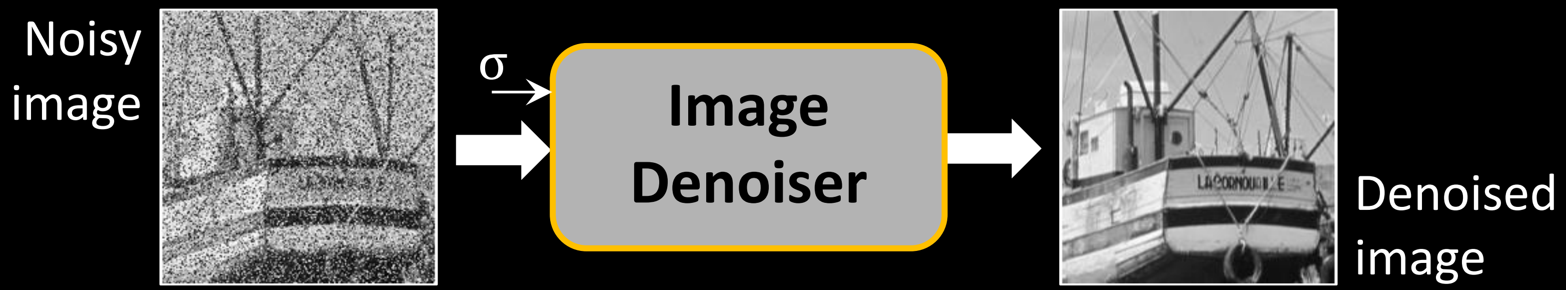
Noisy image

σ

Ima
Deno

Why Work on Image Denoising?

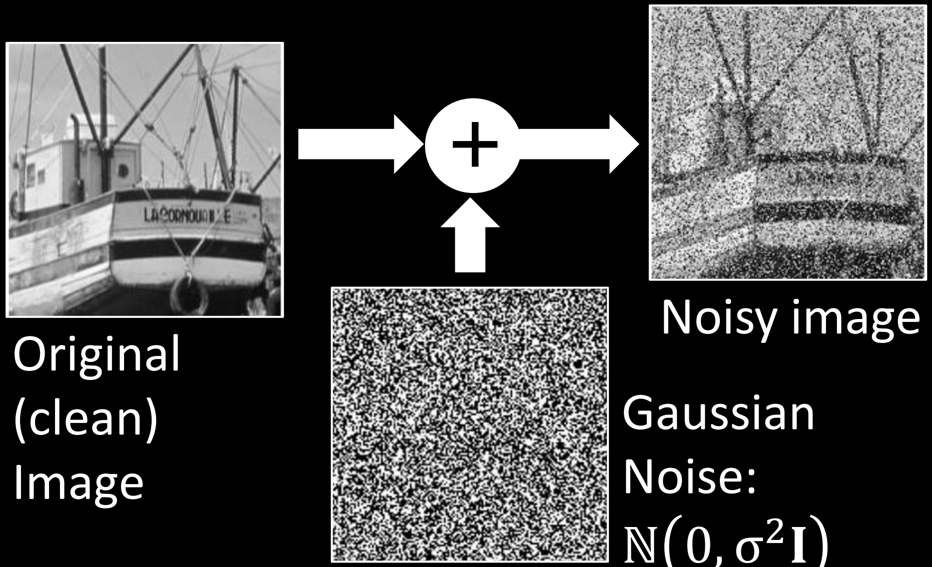
1. Practical: It is a real-world problem, arising in all cameras,
2. Front-Gate to Image Processing: Being the simplest inverse problem, it is a platform for assessing new ideas in our field, &
3. Other Uses for the Denoiser Engine: Recent work has shown that given a denoiser, there are other fascinating uses for it that go far beyond noise removal



Michael Elad
The Computer-Science Department
The Technion

Why Assume Gaussian Noise?

- The Gaussian case is more common and much more important
- When considering a Poisson noise,
  - High count of photons – The distribution gets closer and closer to the Gaussian case
  - Low-count Poisson-distributed image can be converted to a Gaussian-noisy one by Anscomb - Variance Stabilizing Transform
- Many of the developed ideas for the Gaussian case can be converted to other noise models
- MMSE denoisers for the Gaussian case are of extreme theoretical value (see later)



Original (clean) Image

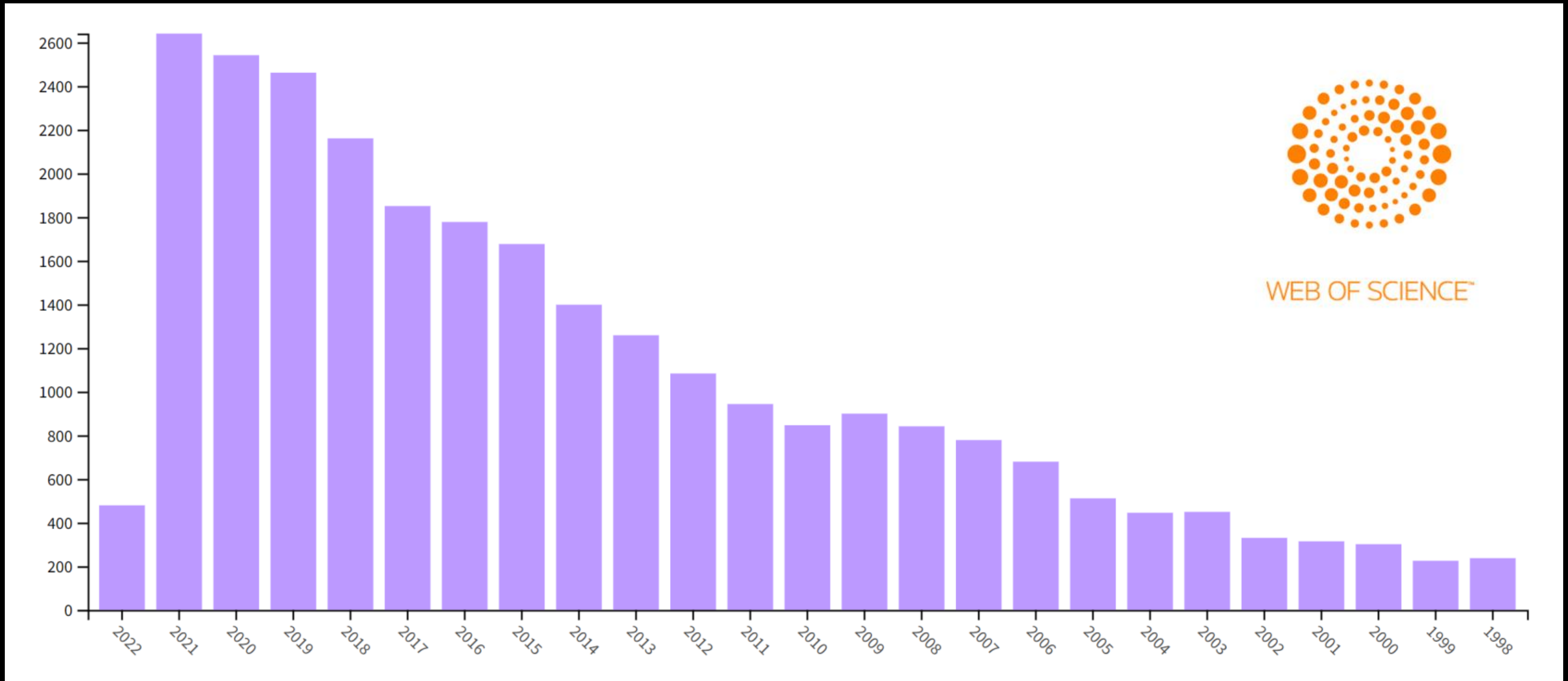
Noisy image

Gaussian Noise: $\mathbb{N}(0, \sigma^2 \mathbf{I})$

Michael Elad
The Computer-Science Department
The Technion

Image Denoising: Little bit of History

Roughly speaking, there are ~28,000 papers* on this subject, offering algorithms, theoretical analysis and so much more

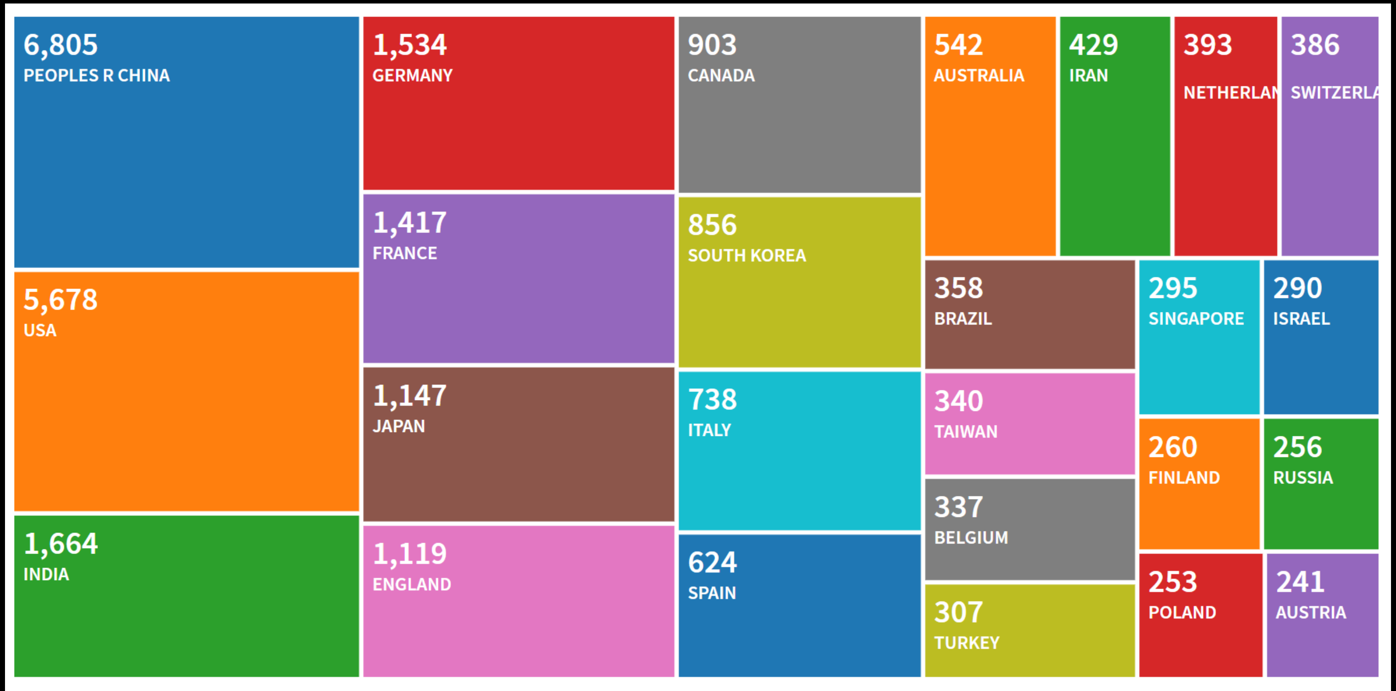


* Search done on April 9th 2022 in WoS, topic: ((image or video) and (denoising or (noise and remov) or clean))

Image Denoising: Little bit of History

Citing Articles:	
USA:	140524
China:	45284
Germany:	29272
France:	35585
England:	24090
Canada:	18325
Spain:	17880
Israel:	13988
Australia:	13358
Switz.:	12504
Japan:	12389
Italy:	11754
Netherland:	10455
India:	8830
Finland:	7842
Korea:	7558
Belgium:	5027
Singapore:	4964
Brazil:	4849
Taiwan:	4134
Iran:	3112
Russia:	2595

This research comes from all over the globe



… and it is heavily cited

Michael Elad
The Computer-Science Department
The Technion

The Classic Era

Design of Image Denoising Algorithms

How can we design a denoiser?

The classic Bayesian approach (1960-2014):

- Model image content with a **prior** expression (e.g., forcing smoothness, sparsity, low-rank, self-similarity, …), and
- Formulate the denoising task as an optimization problem

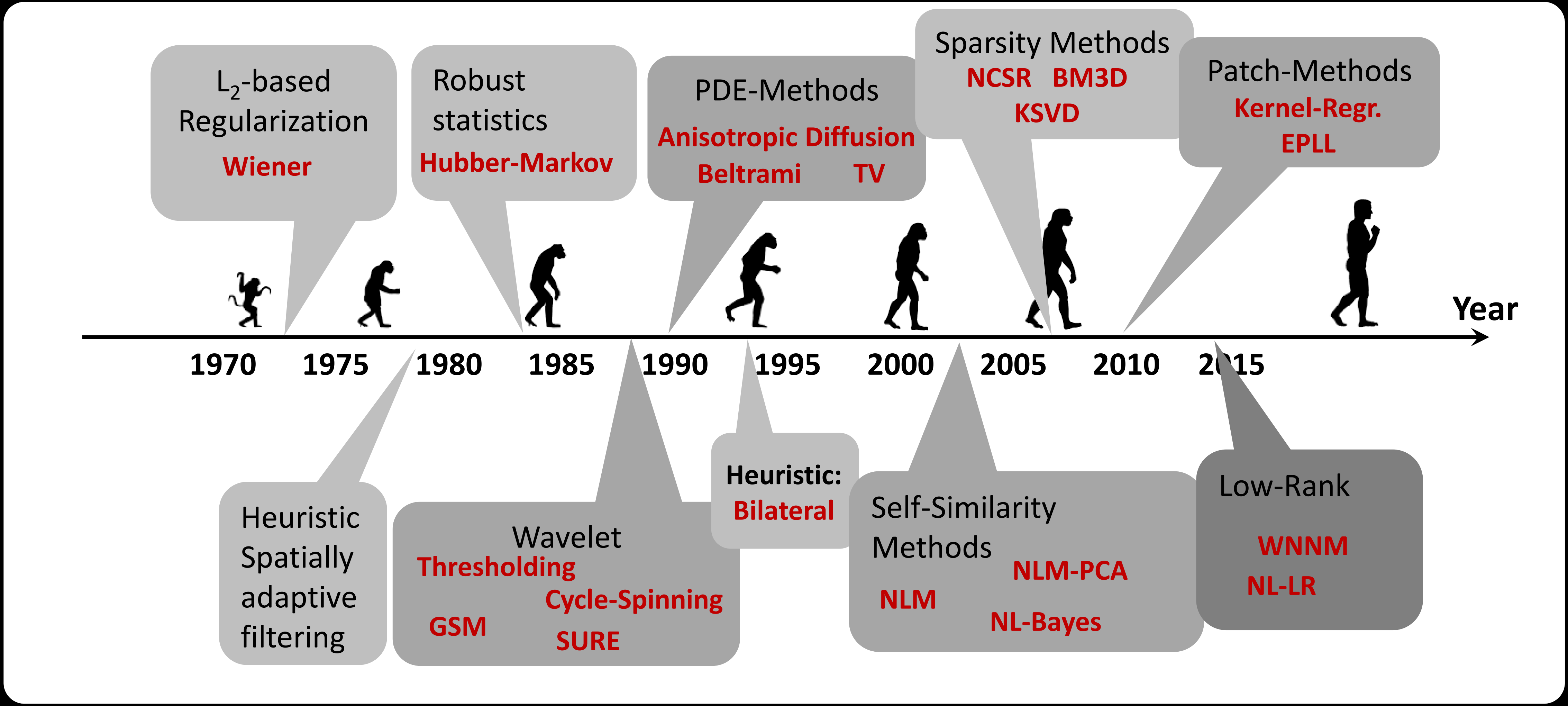
$$\hat{x} = \min_{x} \underbrace{\|x - y\|^2}_{\text{Likelihood}} + \underbrace{\rho(x)}_{\text{Prior}}$$

y : Given noisy image
$\hat{x}$: Denoised result

Design an iterative or a direct algorithm for getting $\hat{x}$ from y

Michael Elad
The Computer-Science Department
The Technion

Image Denoising: Evolution



End of an Era?

This evolution of algorithms and the tendency of different methods to perform very similarly has led to a feeling that "Denoising is Dead"

IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 19, NO. 4, APRIL 2010 895

Is Denoising Dead?

Priyam Chatterjee, *Student Member, IEEE*, and Peyman Milanfar, *Fellow, IEEE*

***Abstract*—Image denoising has been a well studied problem in the field of image processing. Yet researchers continue to focus attention on it to better the current state-of-the-art. Recently proposed methods take different approaches to the problem and yet their denoising performances are comparable. A pertinent question then to ask is whether there is a theoretical limit to denoising performance and, more importantly, are we there yet? As camera manufacturers continue to pack increasing numbers of pixels per unit area, an increase in noise sensitivity manifests itself in the form of a noisier image. We study the performance bounds for the image denoising problem. Our work in this paper estimates a lower bound on the mean squared error of the denoised result and compares the performance of current state-of-the-art denoising methods with this bound. We show that despite the phenomenal recent progress in the quality of denoising algorithms, some room for improvement still remains for a wide class of general images, and at certain signal-to-noise levels. Therefore, image denoising is not dead—yet.**

***Index Terms*—Bayesian Cramér–Rao lower bound (CRLB), bias, bootstrapping, image denoising, mean squared error.**

I. INTRODUCTION

IMAGE denoising has been a well-studied problem in the image processing community and continues to attract

erature on such performance limits exists for some of the mor complex image processing problems such as image registratio [7], [8] and super-resolution [9]–[12]. Performance limits t object or feature recovery in images in the presence of poin wise degradation has been studied by Treibitz *et al.* [13]. I their work, the authors study the effects of noise among othe degradations and formulate expressions for the optimal filterin parameters that define the resolution limits to recovering an given feature in the image. While their study is practical, does not define statistical performance limits to denoisin general images. In [14], Voloshynovskiy *et al.* briefly analyz the performance of MAP estimators for the denoising problem However, our bounds are developed in a much more genera setting and, to the best of our knowledge, no comparable stud currently exists for the problem of denoising. The preser study will enable us to understand how well the state-of-the-a denoising algorithms perform as compared to these limit From a practical perspective, it will also lead to understandin the fundamental limits of increasing the number of sensors i the imaging system with acceptable image quality being mad possible by noise suppression algorithms.

Before we analyze image denoising statistically, we first de

CVPR 2011

Natural Image Denoising: Optimality and Inherent Bounds

Anat Levin and Boaz Nadler
Department of Computer Science and Applied Math
The Weizmann Institute of Science

Abstract

The goal of natural image denoising is to estimate a clean version of a given noisy image, utilizing prior knowledge on the statistics of natural images. The problem has been studied intensively with considerable progress made in recent years. However, it seems that image denoising algorithms are starting to converge and recent algorithms improve over previous ones by only fractional dB values. It is thus important to understand how much more can we still improve natural image denoising algorithms and what are the inherent limits imposed by the actual statistics of the data. The challenge in evaluating such limits is that constructing proper models of natural image statistics is a long standing and yet unsolved problem.

To overcome the absence of accurate image priors, this paper takes a non parametric approach and represents the distribution of natural images using a huge set of 10^{10} patches. We then derive a simple statistical measure which provides a lower bound on the optimal Bayesian minimum mean square error (MMSE). This imposes a limit on the best possible results of denoising algorithms which utilize a

ever, it seems that the performance of denoising algorithms is starting to converge. Recent techniques typically improve over previous ones by only fractional dB values. In some cases the difference between the results of competing algorithms is so small and inconclusive, that one actually has to successively toggle between images on a monitor to visually compare their denoising quality. This raises the question of whether the error rates of current denoising algorithms can be reduced much further, or whether there are inherent limitations imposed by the statistical structure of natural images? The goal of this paper is to derive a *lower bound* on the best possible denoising error under a well defined statistical framework. Such a bound can help us understand if there is hope to significantly improve the current state-of-the-art image denoising with even better algorithms, or whether we have nearly approached the fundamental limits.

Understanding the limits of natural image denoising is also important as an instance of a more fundamental computer and human vision challenge: modeling the statistics of natural images and understanding the inherent limits of their statistical power. Several works attempted to estimate the entropy of natural images [15, 4]. However, there is

End of an Era?

And so, somewhere around 2010-2012, the general feeling in our community was that …

We are currently touching the ceiling in denoising performance and chances of improving them are very slim

⇓

There is no point in devising new denoising methods

⇓

Work in this field has diminishing returns

Well, We Were Wrong !

End of an Era?

Wrong ? How?

The past decade has taught us that image denoising is still

very much alive and kicking

due to several branches of novel activity on:

- Obtaining better performing denoisers with deep learning
- New frontiers in denoising:
  - Better adaptation to image content
  - Denoising strategies that go beyond PSNR
  - Identifying alternative methods for designing/training denoisers
  - Extending the denoising task to realistic noise, and
- Discovering new ways for leveraging denoisers for other needs

The Deep Learning Revolution

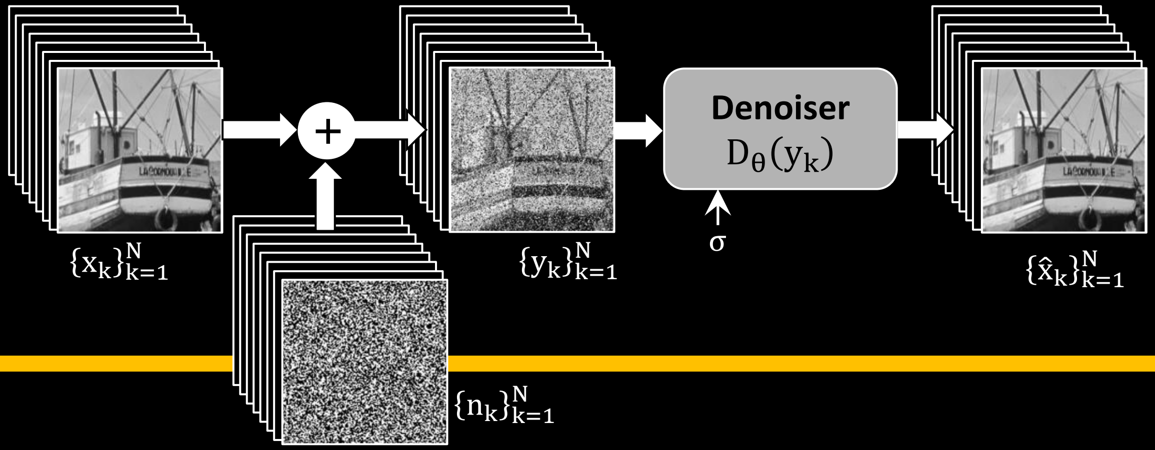
Design of Algorithms: Take 2

How can we **ALTERNATIVELY** design a denoiser?

The machine learning approach (2012-Now):

- Gather a LARGE dataset of clean images $\{x_k\}_{k=1}^N$
- Add AWGN these images: $\{y_k = x_k + n_k\}_{k=1}^N$
- Define a parametric denoising machine $D_\theta(y)$
- Train $D_\theta(\blacksquare)$ by setting its parameters θ:

$$\min_\theta \sum_{k=1}^{N} \|x_k - \underbrace{D_\theta(y_k)}_{\hat{x}_k}\|^2$$

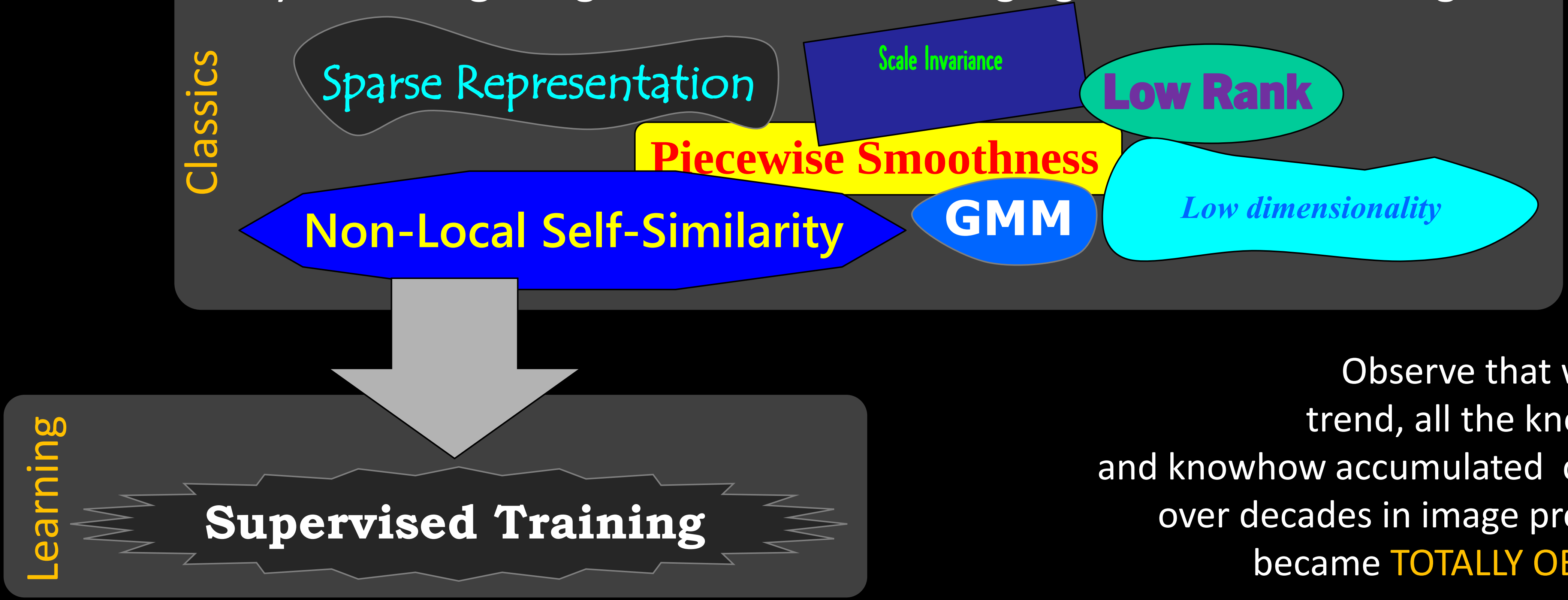


Michael Elad
The Computer-Science Department
The Technion

Image Denoising: A Paradigm Shift

How can we design a denoiser?

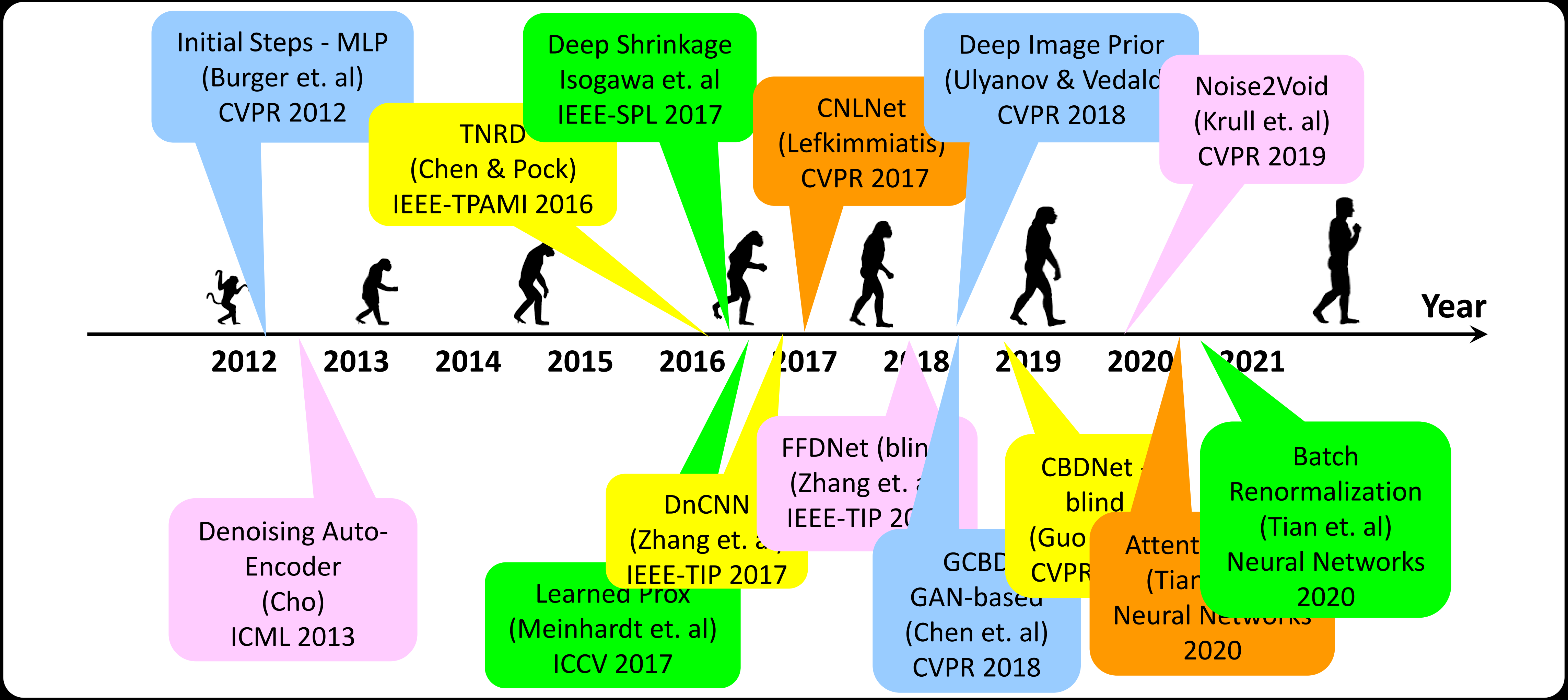
By **modeling** image content and leveraging it for noise filtering:



Observe that with this trend, all the knowledge and knowhow accumulated carefully over decades in image processing became TOTALLY OBSOLETE

Michael Elad
The Computer-Science Department
The Technion

Image Denoising: Recent Evolution

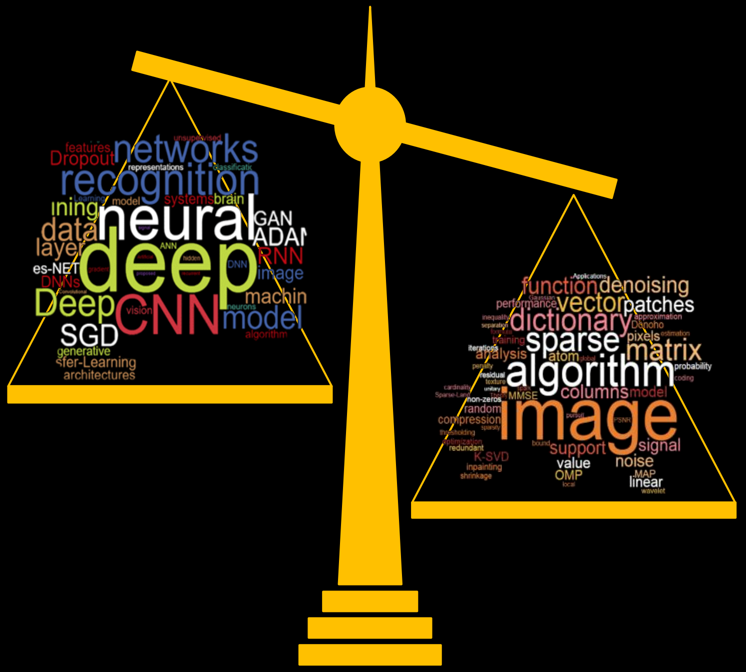


Michael Elad
The Computer-Science Department
The Technion

Synergy:
Classics + Deep Learning

Image Denoising: Return of the Classics

- In recent years deep learning is ruling the image denoising domain, pushing aside all the classical methods, along with their great achievements
- Recently, however, we do see a synergy between the two paradigms
- Recall: In building a supervised deep learning denoiser solution, we operate along the following lines:



Gather training data to use → Define an architecture for the Denoiser → Define a cost function (loss) to optimize → Train and hope for good generalization

Michael Elad
The Computer-Science Department
The Technion

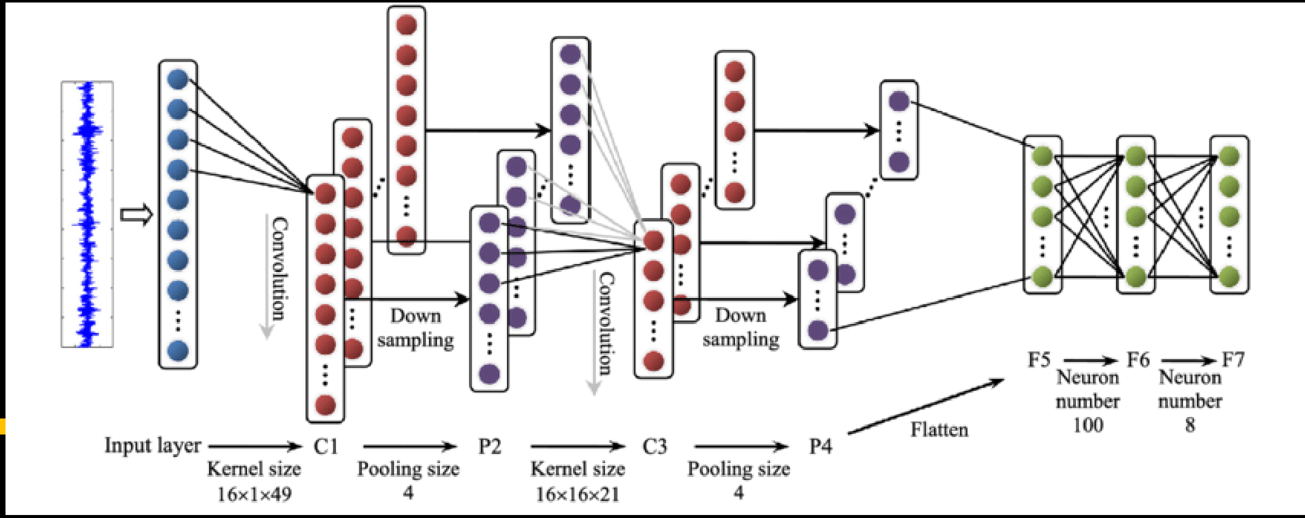
Image Denoising Architectures

So, how do we choose an architecture for a given task?

Option 1 - Copy an existing network that has shown good results in earlier work (VGG, U-Net, ...), and slightly modify it

Option 2 – Pile and Guess a series of steps that mix known pieces such as convolutions, fully connected layer, batch-norm, ReLU, pooling, stride, skips, upscale/downscale, connections, ... and add new “tricks”

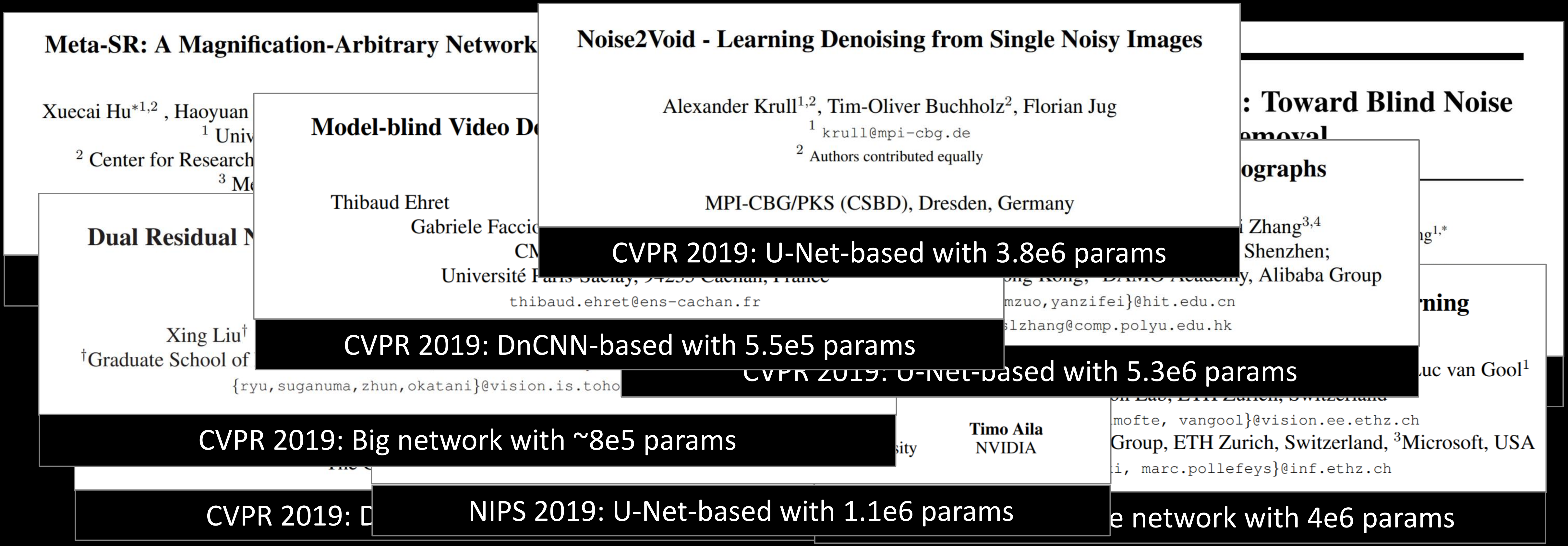
Option 3 – Neural Architecture Search (NAS)



Michael Elad
The Computer-Science Department
The Technion

Image Denoising Architectures

Here are several paper examples from CVPR/NIPS 2019
that illustrate these architectures



Alternative Architecture Design

- Message: One can do far better in choosing architectures by relying on unfolding algorithms from the classics of image processing
- The benefits in such architectures:
  - They are far more concise yet just as effective as leading methods
  - They are easier to train because they are lighter
  - They have the potential to break current performance barriers
  - They may bring better understanding and explainability
  - They enable better adaptation to out of distribution images
- Here are few representative examples:
  - Rethinking the CSC Model [Simon & Elad, NeurIPS `19]
  - LIDIA: Non-Local & Multi-Scale Denoising [Vaksman, Milanfar & Elad, CVPR (NTIRE) `20]
  - Deep KSVD Denoising [Scetbon, Milanfar & Elad, IEEE-TIP `21]
  - PatchCraft: Non-Local Video Denoising [Vaksman, Elad & Milanfar, ICCV `21]

Paper #1: Deep K-SVD

- In 2006, we developed a new and highly effective image denoising alg. by relying on sparsity of image patches and a learned dictionary
- This was state-of-the-art for a while, until beaten by competition (BM3D, NCSR, TNRD, WNNM, …)
- Over the years, various improvements of it came up – e.g. exploiting joint sparsity [Mairal et. al. `09] or leveraging the EPLL [Sulam et. al. `15]
- … And recently we decided to revisit this method …

3736 IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 15, NO. 12, DECEMBER 2006

Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries

Michael Elad and Michal Aharon

***Abstract*—We address the image denoising problem, where zero-mean white and homogeneous Gaussian additive noise is to be removed from a given image. The approach taken is based on sparse and redundant representations over trained dictionaries. Using the K-SVD algorithm, we obtain a dictionary that describes the image content effectively. Two training options are considered: using the corrupted image itself, or training on a corpus of high-quality image database. Since the K-SVD is limited in handling small image patches, we extend its deployment to arbitrary image sizes by defining a global image prior that forces sparsity over patches in every location in the image. We show how such Bayesian treatment leads to a simple and effective denoising algorithm. This leads to a state-of-the-art denoising performance, equivalent and sometimes surpassing recently published leading alternative denoising methods.**

we intend to concentrate on one specific approach towards the image denoising problem that we find to be highly effective and promising: the use of *sparse and redundant representations over trained dictionaries*.

Using redundant representations and sparsity as driving forces for denoising of signals has drawn a lot of research attention in the past decade or so. At first, sparsity of the unitary wavelet coefficients was considered, leading to the celebrated shrinkage algorithm [1]–[9]. One reason to turn to redundant representations was the desire to have the shift invariance property [10]. Also, with the growing realization that regular separable 1-D wavelets are inappropriate for handling images, several new tailored multiscale and directional redundant

Paper #1: Deep K-SVD

- In 2006, we developed a new and highly effective image denoising alg. by relying on sparsity of image patches and a learned dictionary
- This was state-of-the-art for a while, until beaten by competition (BM3D, NCSR, TNRD, WNNM, …)
- Over the years, various improvements of it came up – e.g. exploiting joint sparsity [Mairal et. al. `09] or leveraging the EPLL [Sulam et. al. `15]
- … And recently we decided to revisit this method …

5944 IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 30, 2021

Deep K-SVD Denoising

Meyer Scetbon, Michael Elad, *Fellow, IEEE*, and Peyman Milanfar, *Fellow, IEEE*

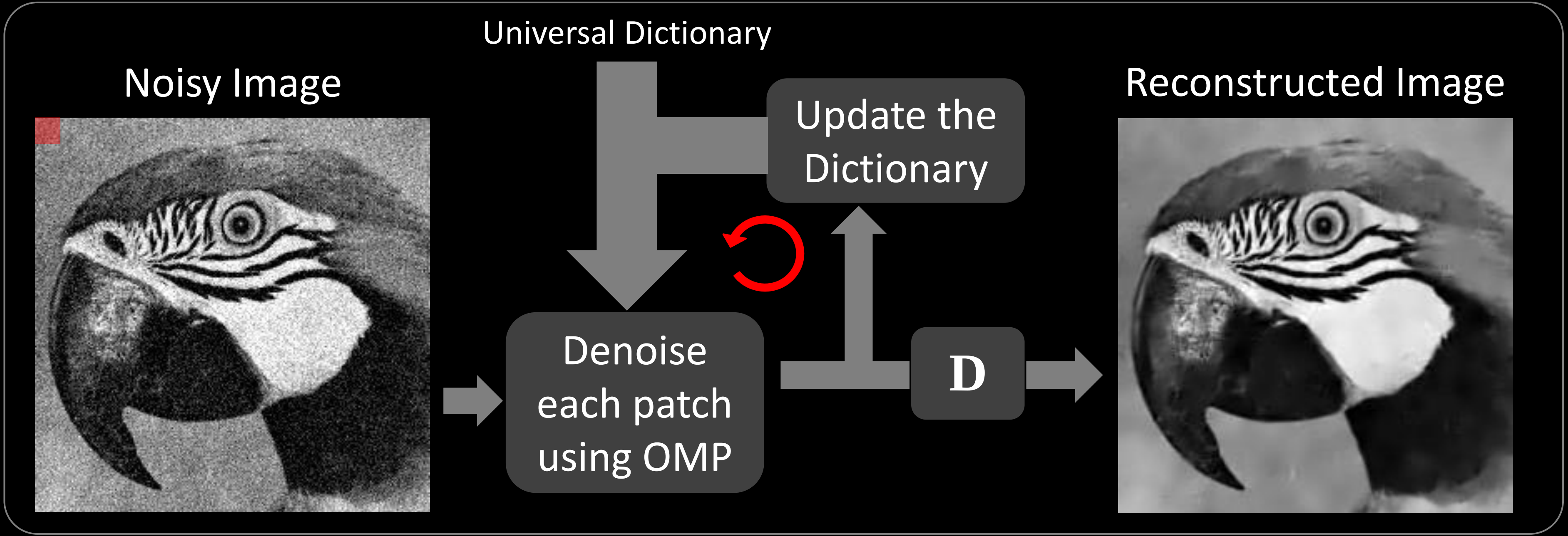
***Abstract*—This work considers noise removal from images, focusing on the well-known K-SVD denoising algorithm. This sparsity-based method was proposed in 2006, and for a short while it was considered as state-of-the-art. However, over the years it has been surpassed by other methods, including the recent deep-learning-based newcomers. The question we address in this paper is whether K-SVD was brought to its peak in its original conception, or whether it can be made competitive again. The approach we take in answering this question is to redesign the algorithm to operate in a supervised manner. More specifically, we propose an end-to-end deep architecture with the exact K-SVD computational path, and train it for optimized denoising. Our work shows how to overcome difficulties arising in turning the K-SVD scheme into a differentiable, and thus learnable, machine. With a small number of parameters to learn and while preserving the original K-SVD essence, the proposed architecture is shown to outperform the classical K-SVD algorithm substantially, and getting closer to recent state-of-the-art learning-based denoising methods. Adopting a broader context, this work touches on themes around the design of deep-learning solutions for image processing tasks, while paving a bridge between classic methods and novel deep-learning-based ones.**

further broadening the applicability of image denoising algorithms [1], [34]. Due to its practical importance and the fact that it is the simplest inverse problem, image denoising has become the entry point for many new ideas brought over the years to the realm of image processing. Over a period of several decades, many image denoising algorithms have been proposed and tested, forming an evolution of methods with gradually improved performance.

A common and systematic approach for the design of novel denoising algorithms is the Bayesian point of view. This calls for image priors, used as regularizers within the Maximum a Posteriori (MAP) or the Minimum Mean Squared Error (MMSE) estimators. In this paper we concentrate on one specific regularization approach, as introduced in [11]: the use of sparse and redundant representation modeling of image patches – this is the K-SVD denoising algorithm, which stands at the center of this paper. The authors of [11] defined a global image prior that forces sparsity over patches in every location

Paper #1: Deep K-SVD

So, how does the original K-SVD denoiser work?



Core idea: Assume that all patches obey sparse modeling

$$\min_{\alpha} \|\alpha\|_0 \ \text{s.t.} \ \|\mathbf{D}\alpha - \mathbf{R}_i y\|_2 \leq T$$

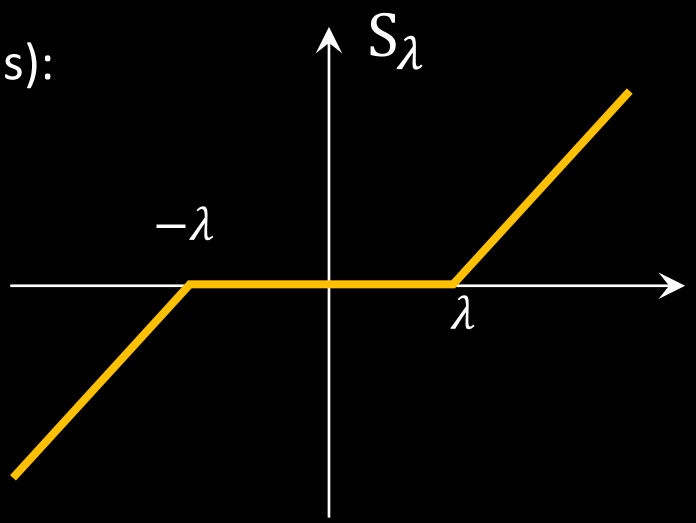
Paper #1: Deep K-SVD

Unfolding this Algorithm:

- All patches (with full overlaps) go through the same “pursuit” in parallel
- OMP problematic (L_0, greedy) $\rightarrow$ Use LISTA [Gregor & LeCun `00] (7 iterations):

$$\min_{\alpha} \|\alpha\|_1 + \lambda\|\mathbf{D}\alpha - \mathbf{R}_i y\|_2^2$$

$$\rightarrow \quad \alpha_{k+1} = S_\lambda\{\alpha_k + c\mathbf{D}^T(\mathbf{D}\alpha_k - \mathbf{R}_i y)\}$$ [ISTA]

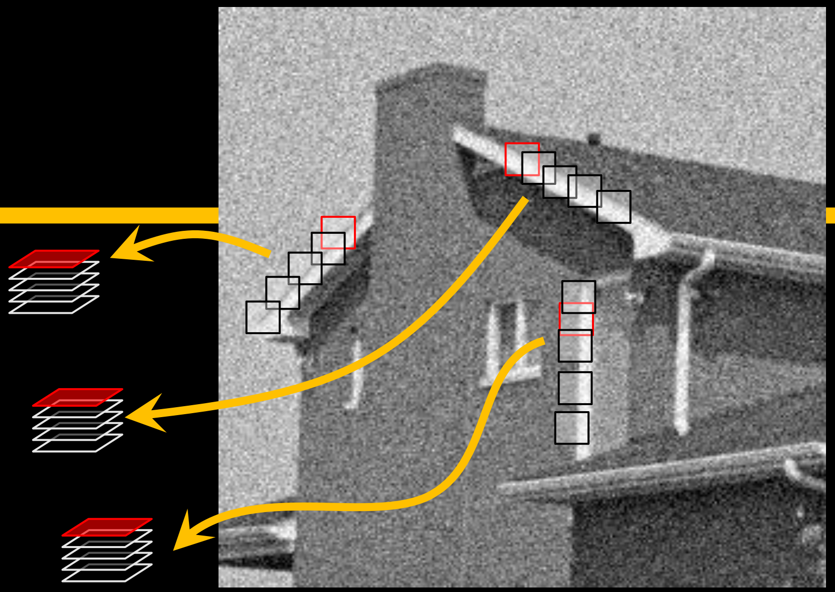


- Each patch should get a dynamic # of non-zeros
  $\rightarrow$ Get an adaptive λ by another small network

Bottom Line:

- The dictionary and few other parameters are learned in a supervised fashion
- Our reference: DnCNN (550K params) [Zhang, `17]. Using 45K params only, this elementary method improve substantially over its original version, and gets very close to DnCNN

Paper #2: Non-Local and Multi-Scale



- Two key forces that the previous work has totally failed to use are (i) self-similarity and (ii) multi-scale connections
- BM3D [Dabov et. al 2007]: A highly effective denoiser based on sparsity and self-similarity
- Core idea: Gather similar patches to 3D blocks & sparse code them jointly
- Our idea: Unfold this algorithm, augment it with a multi-scale treatment, and design it via supervised learning
- LIDIA: our recent work with this unfolding approach

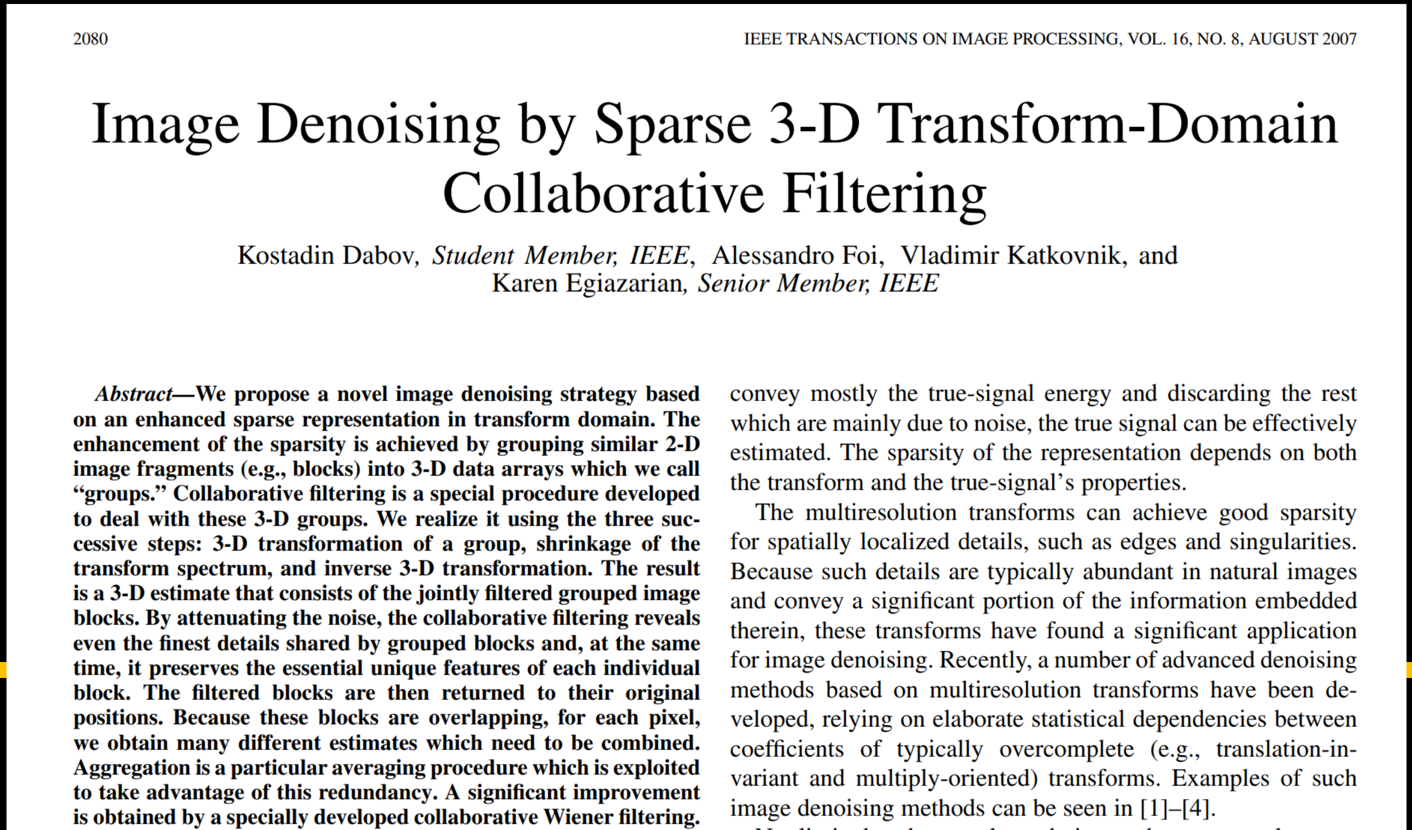


Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering

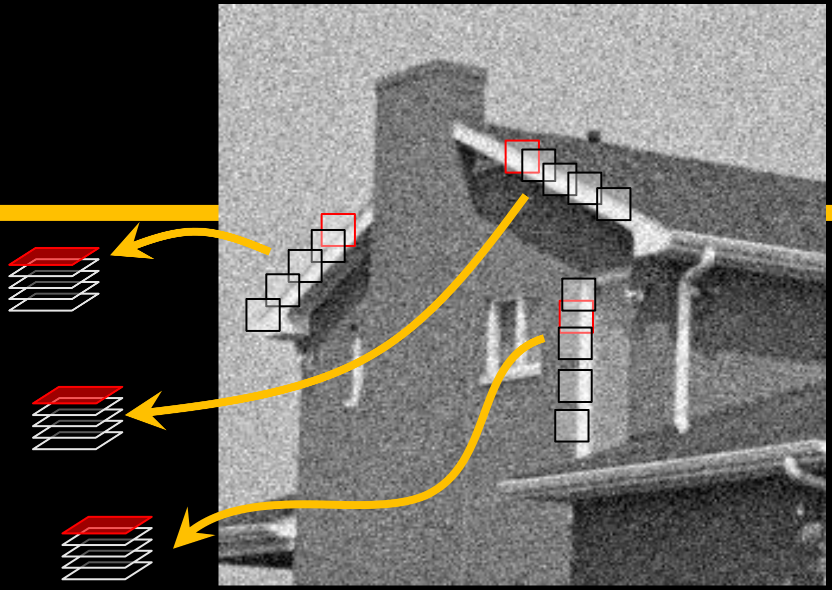
Kostadin Dabov, *Student Member, IEEE*, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian, *Senior Member, IEEE*

***Abstract*—We propose a novel image denoising strategy based on an enhanced sparse representation in transform domain. The enhancement of the sparsity is achieved by grouping similar 2-D image fragments (e.g., blocks) into 3-D data arrays which we call "groups." Collaborative filtering is a special procedure developed to deal with these 3-D groups. We realize it using the three successive steps: 3-D transformation of a group, shrinkage of the transform spectrum, and inverse 3-D transformation. The result is a 3-D estimate that consists of the jointly filtered grouped image blocks. By attenuating the noise, the collaborative filtering reveals even the finest details shared by grouped blocks and, at the same time, it preserves the essential unique features of each individual block. The filtered blocks are then returned to their original positions. Because these blocks are overlapping, for each pixel, we obtain many different estimates which need to be combined. Aggregation is a particular averaging procedure which is exploited to take advantage of this redundancy. A significant improvement is obtained by a specially developed collaborative Wiener filtering.**

convey mostly the true-signal energy and discarding the rest which are mainly due to noise, the true signal can be effectively estimated. The sparsity of the representation depends on both the transform and the true-signal's properties.

The multiresolution transforms can achieve good sparsity for spatially localized details, such as edges and singularities. Because such details are typically abundant in natural images and convey a significant portion of the information embedded therein, these transforms have found a significant application for image denoising. Recently, a number of advanced denoising methods based on multiresolution transforms have been developed, relying on elaborate statistical dependencies between coefficients of typically overcomplete (e.g., translation-invariant and multiply-oriented) transforms. Examples of such image denoising methods can be seen in [1]–[4].

Paper #2: Non-Local and Multi-Scale



- ❑ Two key forces that the previous work has totally failed to use are (i) self-similarity and (ii) multi-scale connections
- ❑ BM3D [Dabov et. al 2007]: A highly effective denoiser based on sparsity and self-similarity
- ❑ Core idea: Gather similar patches to 3D blocks & sparse code them jointly
- ❑ Our idea: Unfold this algorithm, augment it with a multi-scale treatment, and design it via supervised learning
- ❑ LIDIA: our recent work with this unfolding approach

2080 IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 16, NO. 8, AUGUST 2007

Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering

LIDIA: Lightweight Learned Image Denoising with Instance Adaptation

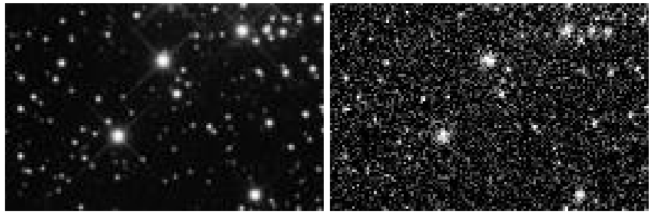
Gregory Vaksman
CS Department - The Technion
grishav@campus.technion.ac.il

Michael Elad
Google Research
melad@google.com

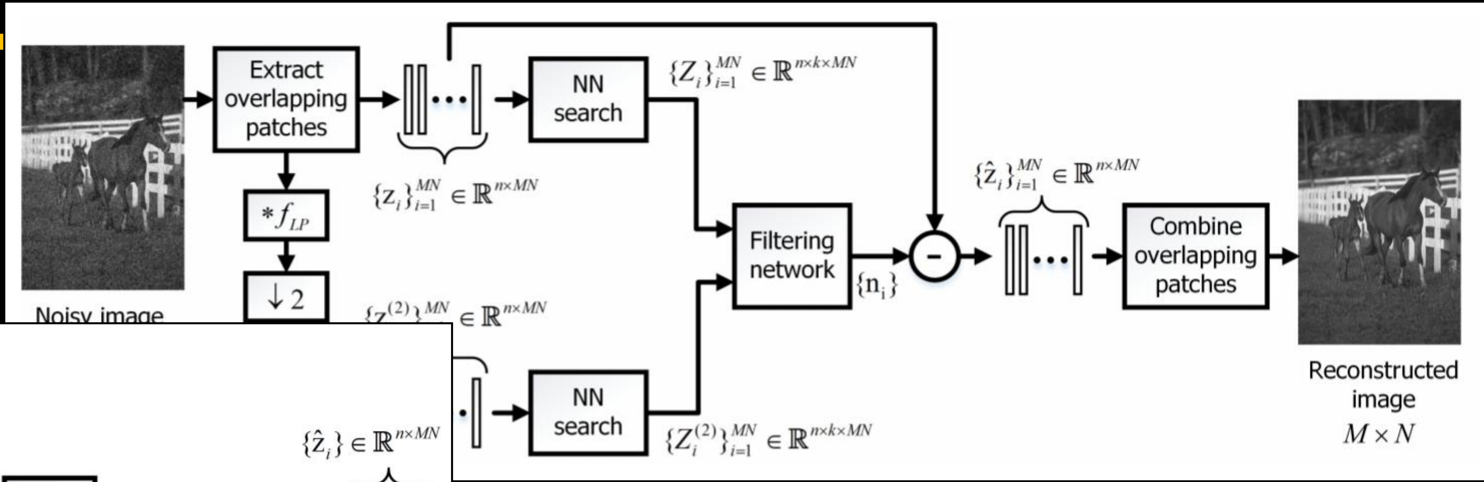
Peyman Milanfar
Google Research
milanfar@google.com

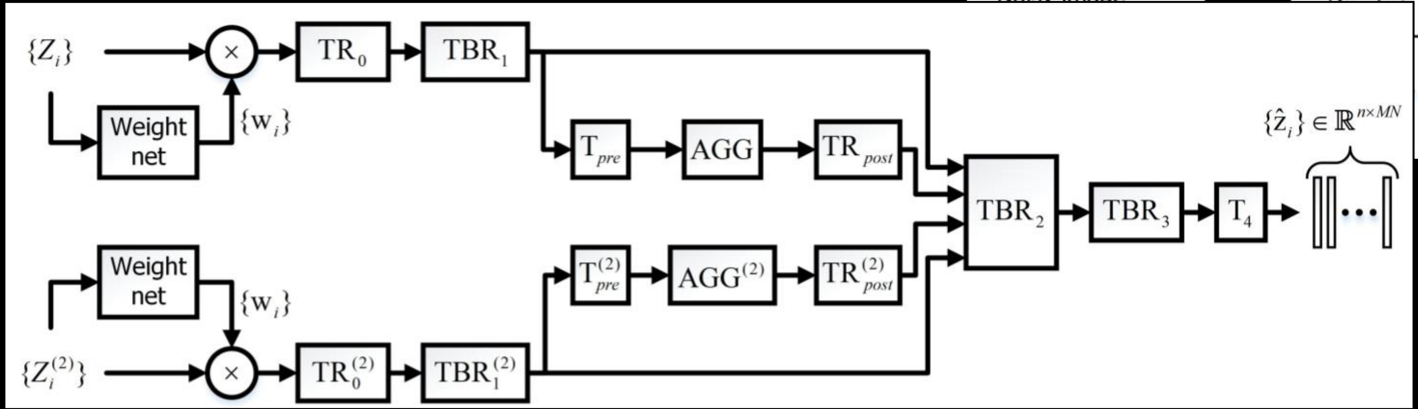
Abstract

Image denoising is a well studied problem with an extensive activity that has spread over several decades. Leading classical denoising methods are typically designed to exploit the inner structure in images by modeling local overlapping patches, and operating in an unsupervised fashion.

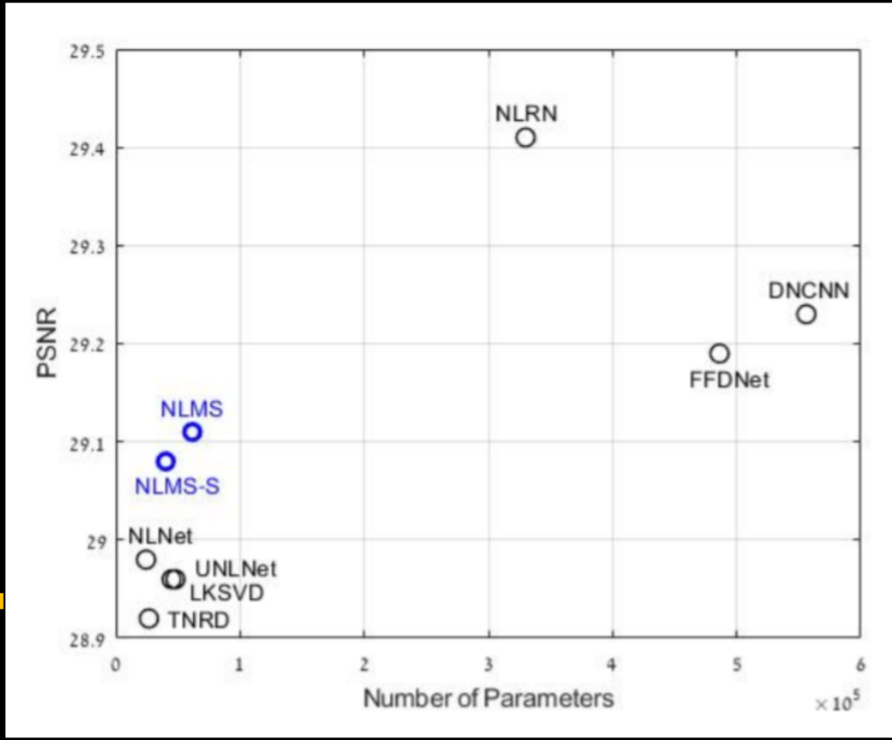


Paper #2: Non-Local and Multi-Scale





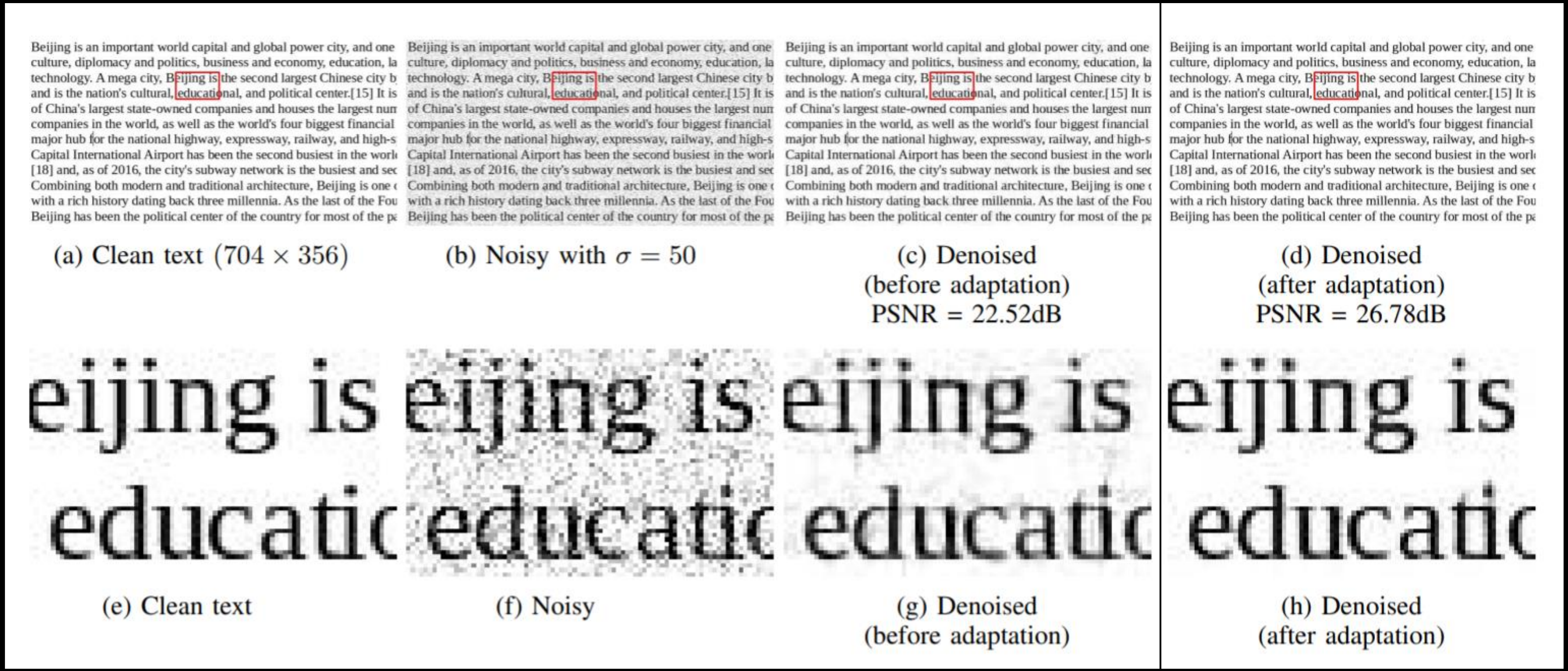
- The proposed architecture implements the ideas mentioned above in a simple a direct way
- This illustrates the performance vs. #of parameters for various networks



Michael Elad
The Computer-Science Department
The Technion

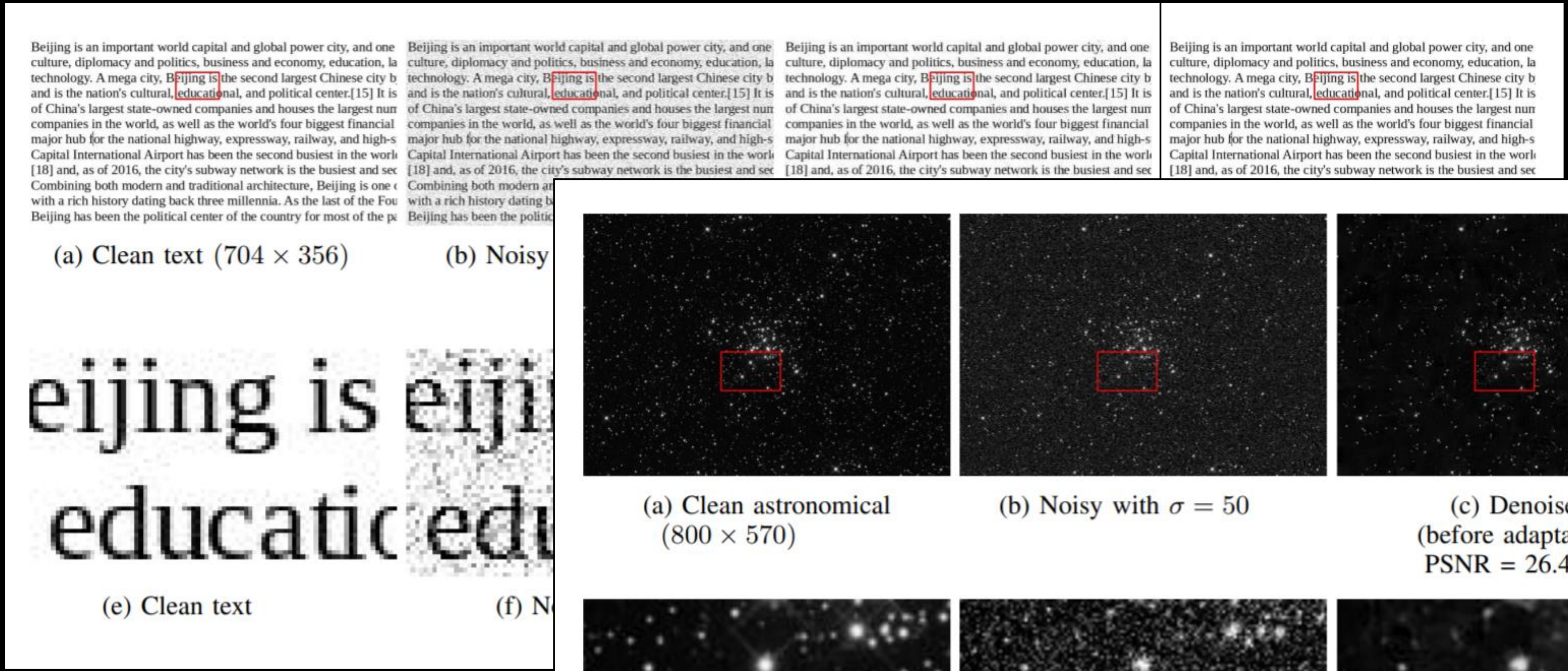
Paper #2: Non-Local and Multi-Scale

- An additional and unexpected benefit: Fast and effective adaptation capability



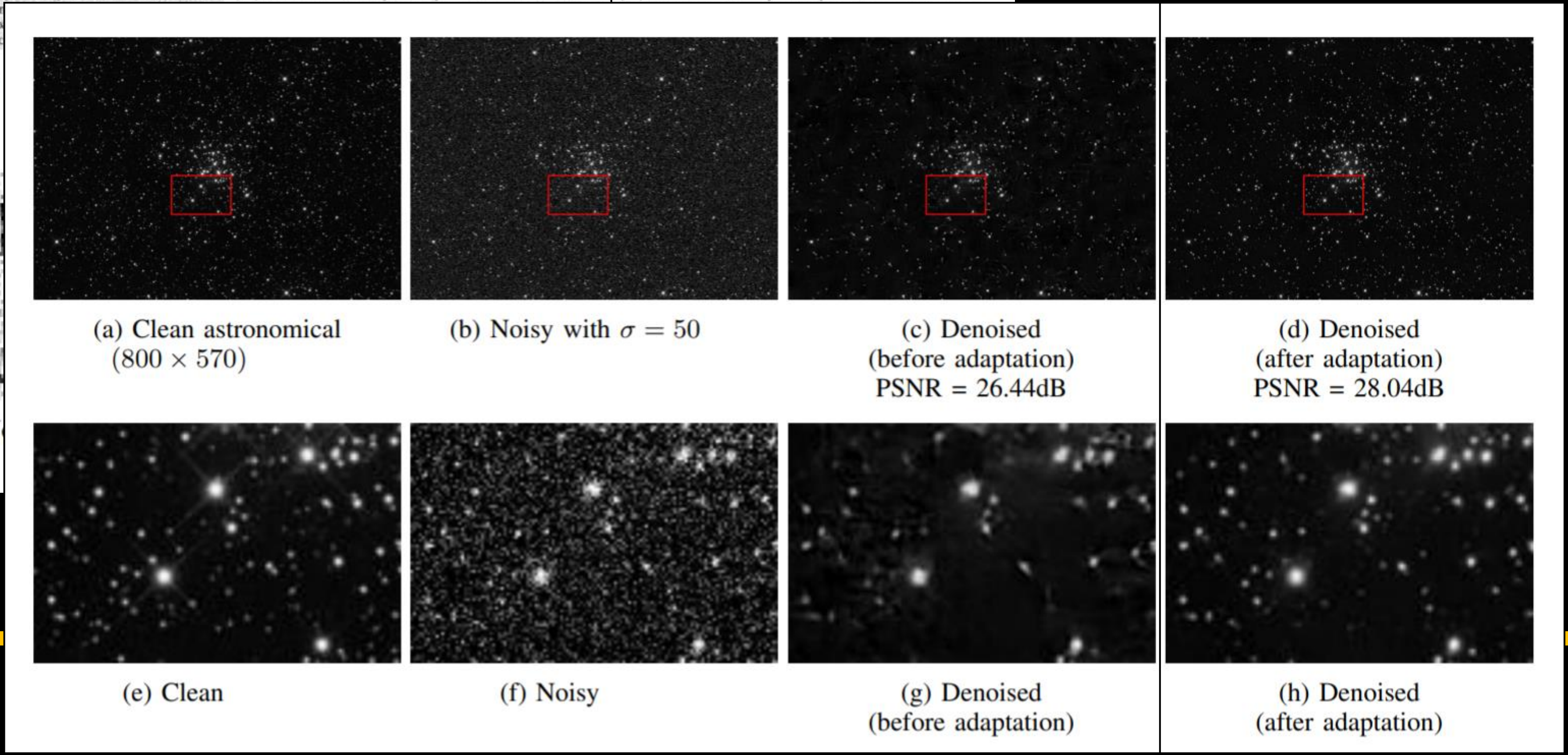
Paper #2: Non-Local and Multi-Scale

- An additional and unexpected benefit: Fast and effective adaptation capability



(a) Clean text (704 × 356) (b) Noisy

(e) Clean text (f) N



(a) Clean astronomical (800 × 570)

(b) Noisy with $\sigma = 50$

(c) Denoised (before adaptation) PSNR = 26.44dB

(d) Denoised (after adaptation) PSNR = 28.04dB

(e) Clean

(f) Noisy

(g) Denoised (before adaptation)

(h) Denoised (after adaptation)

Summary – Synergy of Classics and Deep Does Exist

- The right way to build solutions to imaging tasks goes as follows:

Modeling the Data → **Algorithm Development** → **Unfold this to a Network** → **Apply Supervised Learning**

- What should be taken into account for the algorithm' design?
  - The degradation and noise statistics ("the physics")
  - Prior on the image: (i) Non-Local self similarity; (ii) multi-scale connections; & (iii) Sparsity or other form of simplicity (e.g. low-rank)
  - The objective (e.g., MMSE)

- More broadly, sparse modeling of data could be key
  - In explaining existing deep-learning architectures
  - In creation of new ones
  - In bringing theoretical understanding to deep-learning

Still Unanswered

Open Questions:

- ❑ When designing an algorithm (and thus a network) for solving inverse problems, should we consider MMSE or MAP?
- ❑ It will be great to see this advocated rationale breaking existing performance barriers – this is yet to happen
- ❑ What about using this rationale for supporting unsupervised solutions? Recall the K-SVD denoising with an adapted dictionary
- ❑ Denoising is a regression task, like many others in computational imaging. However, what about recognition or synthesis tasks?

Our Focus Today: Recent Discoveries

Our Focus Today

Recent findings on using denoisers for other tasks:

- ❑ Discovery 1: Solving general inverse problems [2013-]
- ❑ Discovery 2: Image Synthesis [2019-]
- ❑ Discovery 3: High perceptual quality recovery [2021-]

We turn to describe these results

Noisy image y

Image Denoiser
$D(y, \sigma)$

$\uparrow$
σ

Denoised image $\hat{x}$

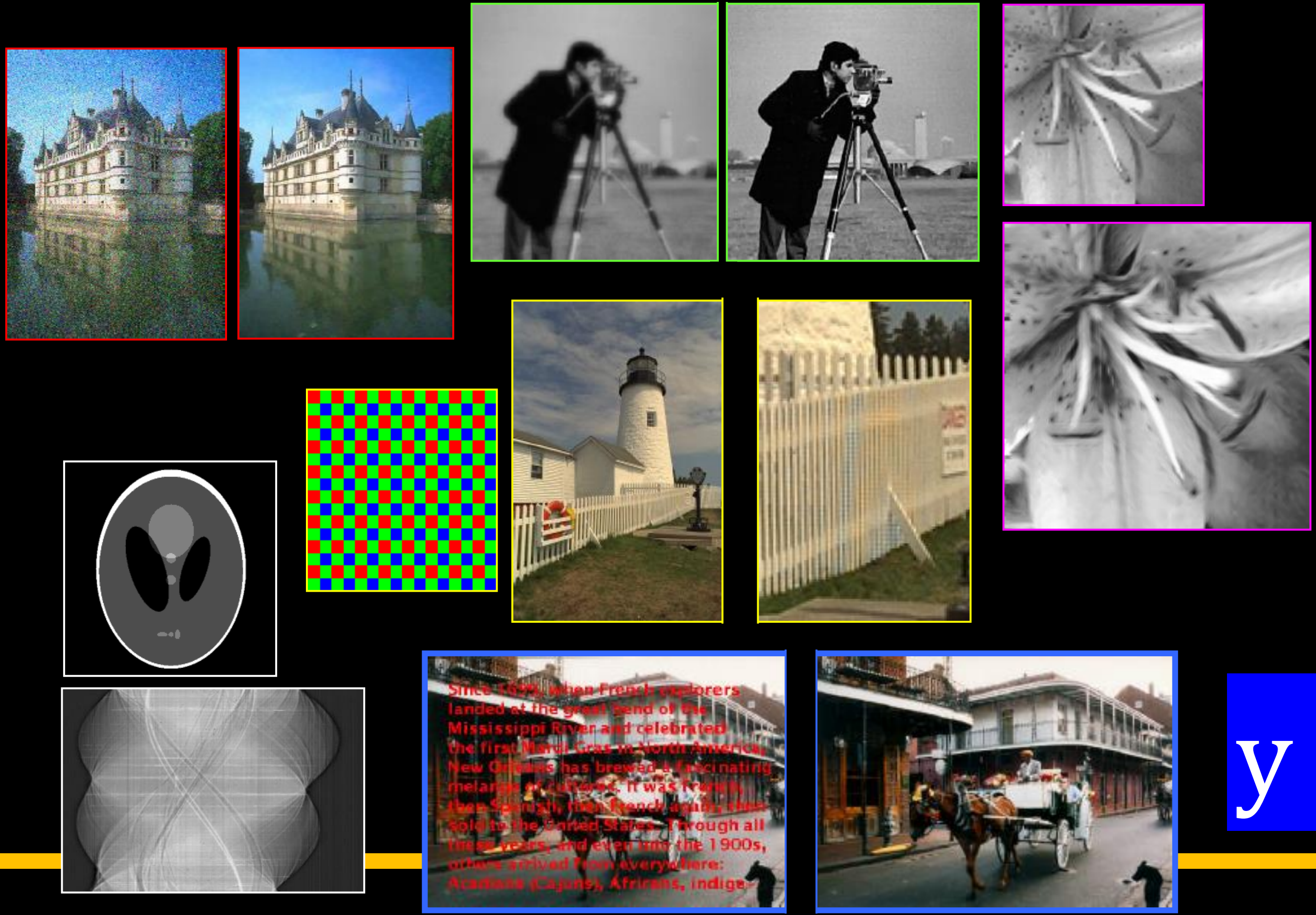
Michael Elad
The Computer-Science Department
The Technion

Discovery 1: Solving Inverse Problems

Inverse Problems: Recovery of images from corrupted measurements

- ❑ De-Noising
- ❑ De-Blurring
- ❑ In-Painting
- ❑ De-Mosaicing
- ❑ Tomography
- ❑ Image Scale-Up & super-resolution

… and more …



$$y = Hx + n$$

Michael Elad
The Computer-Science Department
The Technion

Discovery 1: Solving Inverse Problems

How can we solve inverse problems?

We can return to the classic Bayesian approach:

- Model image content with a prior expression (e.g., forcing smoothness, sparsity, low-rank, self-similarity, …), and
- Formulate the inversion task as an optimization problem

$$\hat{x} = \min_{x} \underbrace{\|Hx - y\|^2}_{\text{Likelihood}} + \underbrace{\rho(x)}_{\text{Prior}}$$

y : Given measurements
$\hat{x}$: Denoised result

- This is known as MAP estimation
- It is an extension of the classic denoising, tailoring methods for inverse problems
- This approach leads to iterative algorithm for getting $\hat{x}$ from y
- Is there a supervised learning alternative? Definitely!

Discovery 1: Solving Inverse Problems

Question: Given a denoiser $D(y, \sigma)$
how can one solve inverse problems with it?

Plug-and-play priors for model based reconstruction SV Venkatakrishnan, CA Bouman, B Wohlberg 2013 IEEE Global Conference on Signal and Information Processing, 945-948	526	2013
The little engine that could: Regularization by denoising (RED) Y Romano, M Elad, P Milanfar SIAM Journal on Imaging Sciences 10 (4), 1804-1844	373	2017

Answer: Use $D(y, \sigma)$ as a ***regularizer***

Practical Implication: Iterated use of $D(\cdot, \sigma)$

y → Simple Operation → $D(\cdot, \sigma)$ → Simple Operation → $D(\cdot, \sigma)$ → Simple Operation → $D(\cdot, \sigma)$ → Simple Operation

Michael Elad
The Computer-Science Department
The Technion

Discovery 1: Solving Inverse Problems

Here is (roughly) the PnP Perspective in a nutshell:

- Recall: Inverse problems can be formulated as optimization tasks:

$$\hat{x} = \min_{x} \frac{1}{2}\|Hx - y\|^2 + \rho(x)$$

- Let's do something "stupid" and split the unknown:

$$\hat{x} = \min_{x,v} \frac{1}{2}\|Hx - y\|^2 + \rho(v) \quad \text{s.t.} \quad x = v$$

- Now, turn the constraint into a penalty*

$$\hat{x} = \min_{x,v} \frac{1}{2}\|Hx - y\|^2 + \rho(v) + \beta\|x - v\|^2$$

- And solve by alternating between x and v
  - Least-Squares: $\hat{x} = \min_{x} \frac{1}{2}\|Hx - y\|^2 + \beta\|x - v\|^2$
  - A denoiser: $\hat{v} = \min_{v} \rho(v) + \beta\|x - v\|^2$

… and this way we got an iterated algorithm that keeps calling to a denoiser, for solving the inverse problem

* The PnP uses the Augmented Lagrange which is more accurate and less sensitive to the choice of β

Michael Elad
The Computer-Science Department
The Technion

Discovery 1: Solving Inverse Problems

Here is the RED Perspective in a nutshell:

Let's start again with the formulated optimization task, and suggest a very specific regularization term:

$$x^T(\mathbf{I}-\mathbf{S})x$$

$$\hat{x} = \min_x \frac{1}{2}\|Hx-y\|^2 + \rho(x) = \min_x \frac{1}{2}\|Hx-y\|^2 + \underbrace{\boxed{\lambda x^T[x - D(x,\sigma)]}}$$

Let's use the Steepest Descent

Under mild conditions* the gradient of this is $[x - D(x,\sigma)]$

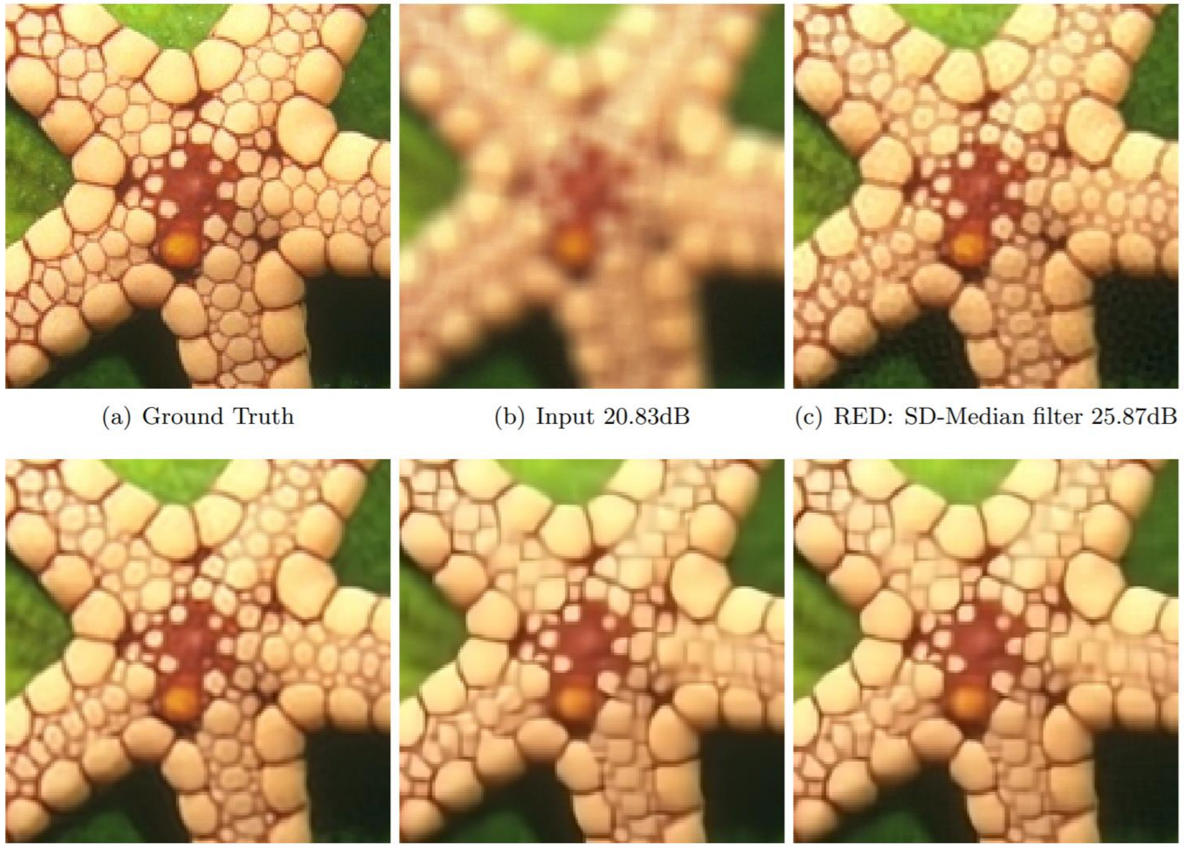
$$\hat{x}_{k+1} = \hat{x}_k - \mu\left[H^T(H\hat{x}_k - y) + \lambda[\hat{x}_k - D(\hat{x}_k,\sigma)]\right]$$

… and this way we got an iterated algorithm that keeps calling to a denoiser, and is guaranteed to achieve the minimum

* Differentiability, local homogeneity, passivity and symmetric Jacobian (MMSE)

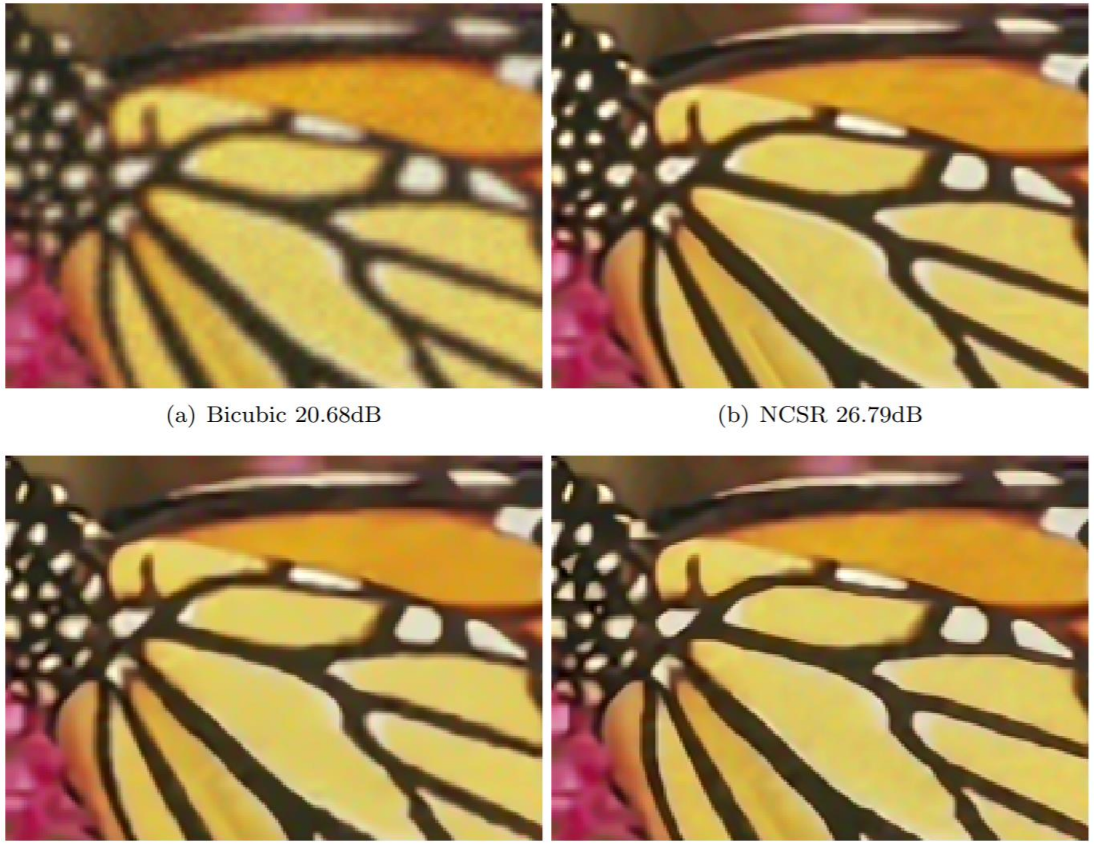
Discovery 1: Solving Inverse Problems

Here are some results for Deblurring and Super-Resolution



(a) Ground Truth (b) Input 20.83dB (c) RED: SD-Median filter 25.87dB

(d) NCSR 28.39dB (e) P^3-TNRD 28.43dB (f) RED: FP-TNRD 28.82dB



(a) Bicubic 20.68dB (b) NCSR 26.79dB

(c) P^3-TNRD 26.61dB (d) RED: SD-TNRD 27.39dB

Discovery 1: Solving Inverse Problems

Plug-and-play priors for model based reconstruction SV Venkatakrishnan, CA Bouman, B Wohlberg 2013 IEEE Global Conference on Signal and Information Processing, 945-948	526	2013
The little engine that could: Regularization by denoising (RED) Y Romano, M Elad, P Milanfar SIAM Journal on Imaging Sciences 10 (4), 1804-1844	373	2017

- PnP and RED are heavily cited and extensively studied, owing to their generality and elegance
- Follow-up work focuses on
  - Proving convergence to the desired solution and tying these to properties of the permissible denoisers (e.g. MMSE …)
  - Deployment in various applications
  - Creation of new variants of these two methods … and …
- Unfolding all over again: PnP/RED can be used to define well-motivated architectures for solving general inverse problems, built around a core learned denoising engine

Discovery 2: Image Synthesis

- In recent years, and with the deep-learning revolution, there is a growing interesting is synthesizing images “out of thin air”
- The popular tool of interest is called GAN – Generative Adversarial Network, built of two competing networks – a generator and a critique
- Why synthesize? Because
  - We can, and it is fascinating
  - It testifies that we have seized the distribution of images, and
  - It could be used for other needs
- Could we synthesize images differently?

thispersondoesnotexist.com

Michael Elad
The Computer-Science Department
The Technion

Discovery 2: Image Synthesis

Question: Given a denoiser $\mathrm{D}(y, \sigma)$
how can one synthesize images with it?

Generative modeling by estimating gradients of the data distribution Y Song, S Ermon Advances in Neural Information Processing Systems 32	235	2019
Improved techniques for training score-based generative models Y Song, S Ermon Advances in neural information processing systems 33, 12438-12448	88	2020
Stochastic Solutions for Linear Inverse Problems using the Prior Implicit in a Denoiser Z Kadkhodaie, EP Simoncelli Advances in Neural Information Processing Systems 34	2	2021

Answer: Use $\mathrm{D}(y, \sigma)$ as a ***Projector*** onto the image manifold

Practical Implication: Iterated use of $\mathrm{D}(\cdot, \sigma)$ with varying σ

noise → Simple Operation → $\mathrm{D}(\cdot, \sigma_1)$ → Simple Operation → $\mathrm{D}(\cdot, \sigma_2)$ → Simple Operation → $\mathrm{D}(\cdot, \sigma_3)$ →

Michael Elad
The Computer-Science Department
The Technion

Discovery 2: Image Synthesis

Here is the core idea in a nutshell:

Our goal: draw a sample from the distribution of images $\mathrm{P}(\mathrm{x})$

- Start with a random noise image $\hat{\mathrm{x}}_0$
- Climb to a more probable image by the iterative equation:

$$\hat{\mathrm{x}}_{\mathrm{k}+1} = \hat{\mathrm{x}}_{\mathrm{k}} + \mathrm{a} \cdot \underbrace{\boxed{\nabla \log \mathrm{P}(\hat{\mathrm{x}}_{\mathrm{k}})}} + \mathrm{b} \cdot \mathrm{z}_{\mathrm{k}}$$ (Langevin Dynamics)

This is known as the Score Function and it is approximately proportional to $[\hat{\mathrm{x}}_{\mathrm{k}} - \mathrm{D}(\hat{\mathrm{x}}_{\mathrm{k}}, \sigma)]$ for a small value of σ

⇔

This suggests an implicit relation between MMSE denoisers and Priors: $\mathrm{D}(\mathrm{x}, \sigma) \leftrightarrow \mathrm{P}(\mathrm{x})$

… and this way we got an iterated algorithm that keeps calling to a denoiser, and is guaranteed to obtain a sample from $\mathrm{P}(\mathrm{x})$

Discovery 2: Image Synthesis

In practice, instead of the plain Langevin with a fixed (and small) value of σ we use the Annealed Langevin Algorithm that considers a sequence of blurred priors:

$$P(x+v) \quad \text{for} \quad v \sim \mathbb{N}\left(0, \sigma_k^2 \mathbf{I}\right)$$

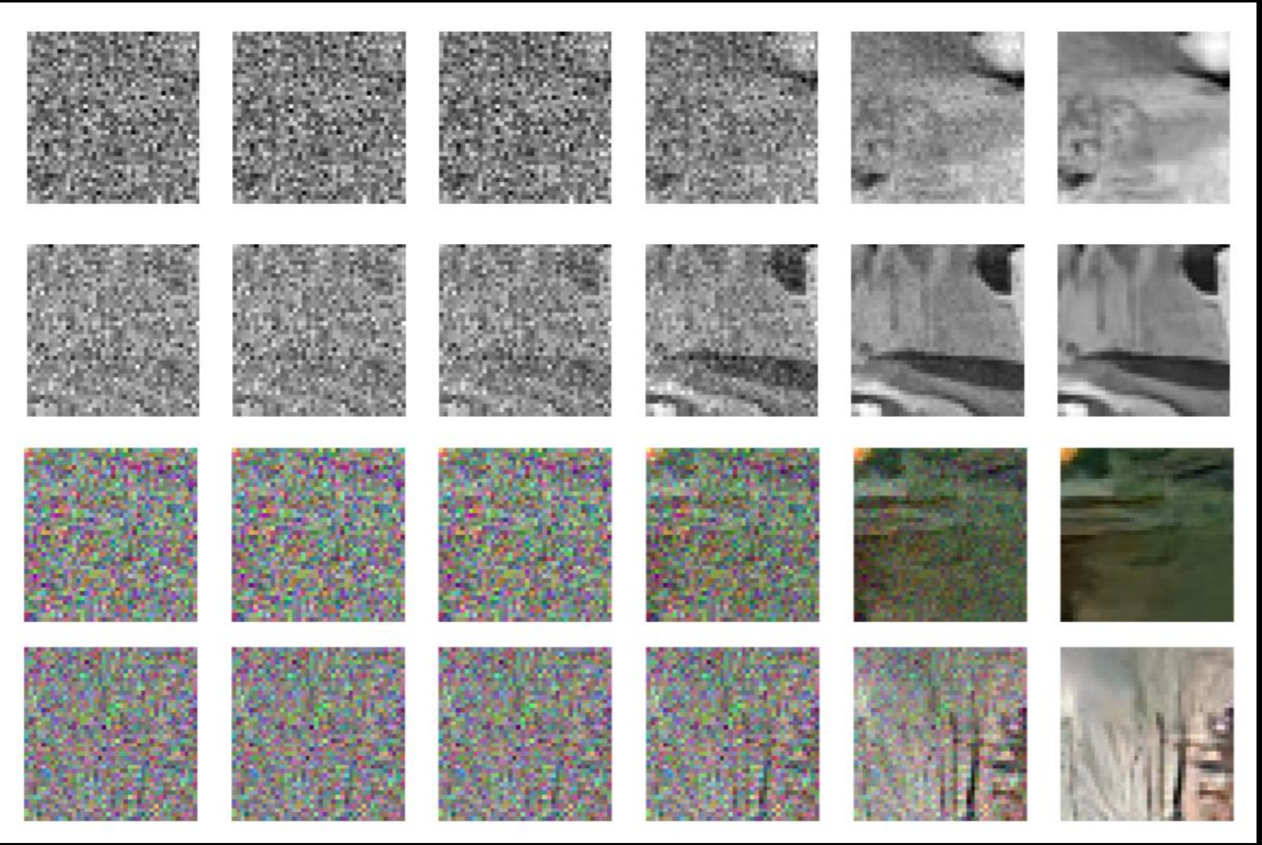
$$= P(x) \otimes c \cdot \exp\left\{-\frac{1}{2\sigma^2}\|x\|^2\right\}$$

$$\text{with} \quad \sigma_0 > \sigma_1 > \sigma_2 \quad \cdots \quad > \sigma_N > 0$$

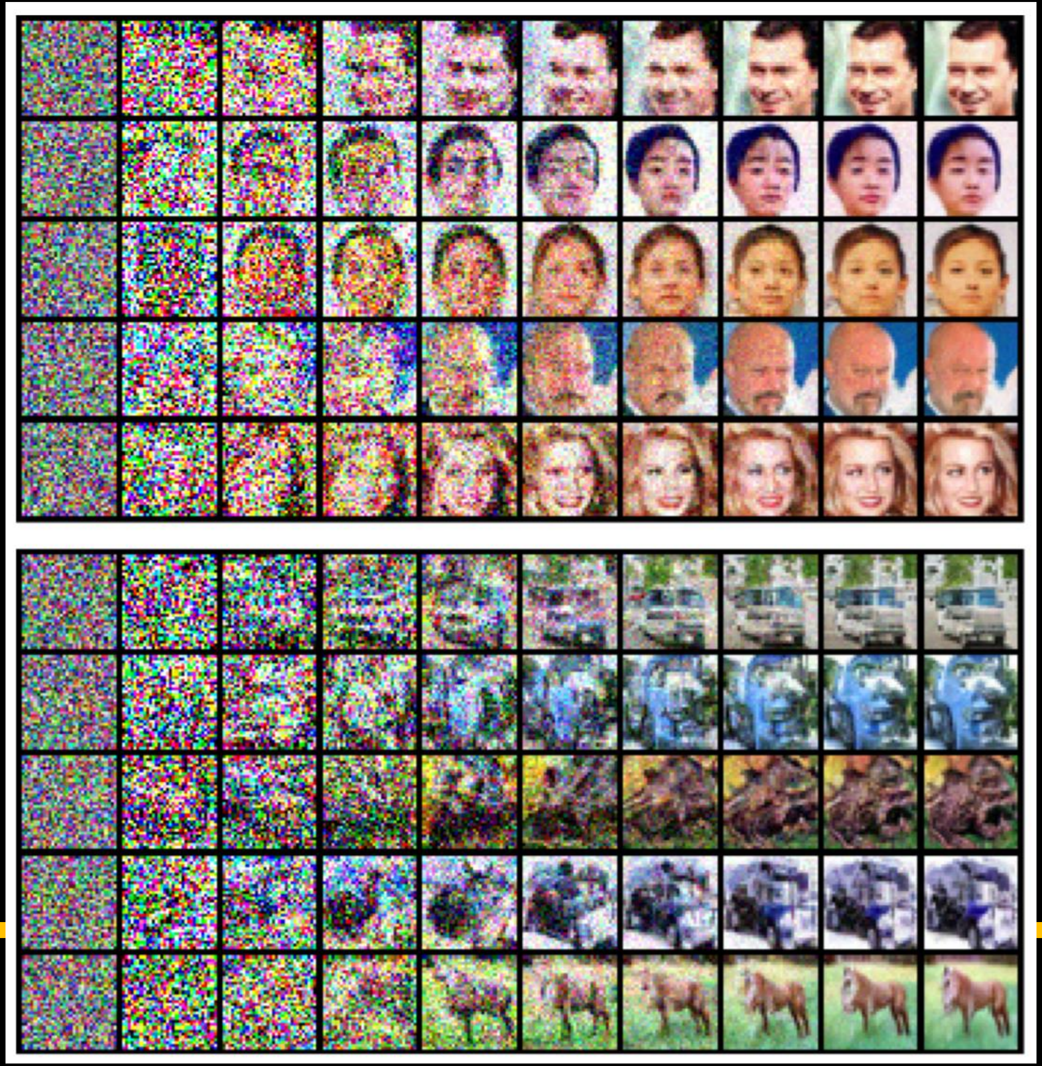
The core idea: start by drawing from a wider distribution and gradually narrow it, leading to a faster sampling performance

Discovery 2: Image Synthesis

Does it work? Here are some results



Kadkhodaie & Simoncelli

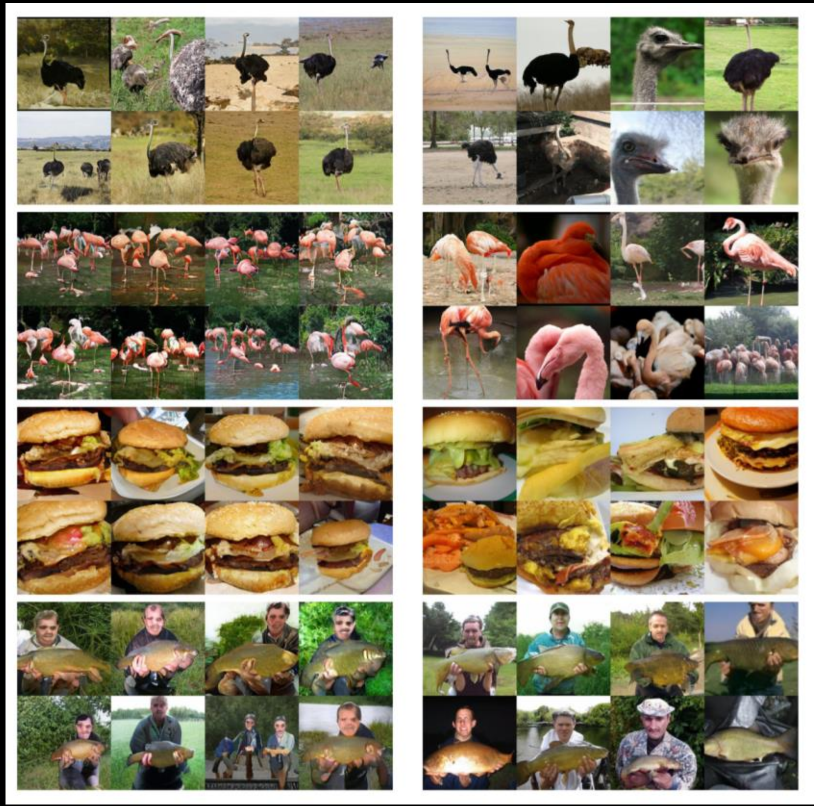


Song & Ermon

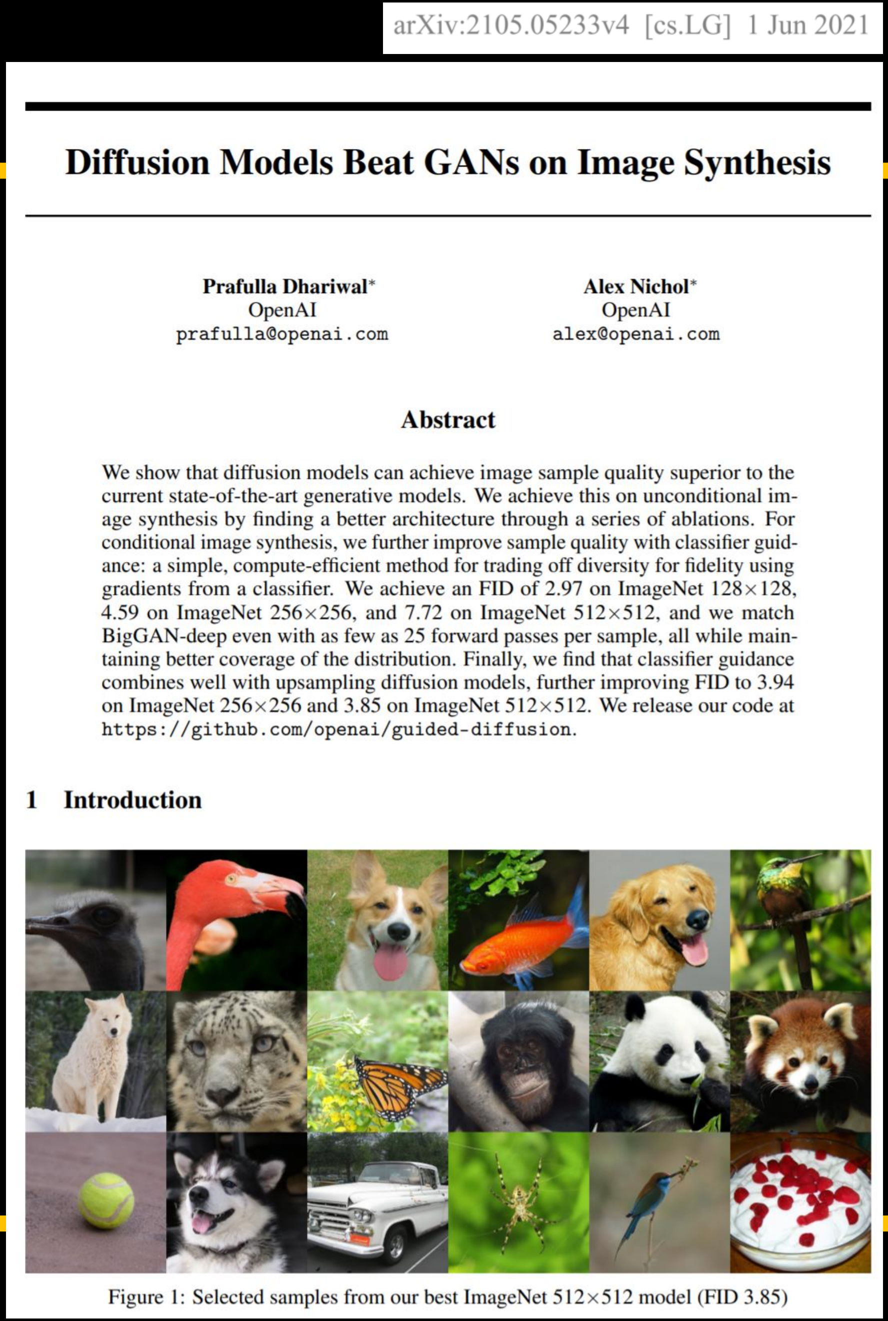
Michael Elad
The Computer-Science Department
The Technion

Discovery 2: Image Synthesis

Claim: diffusion-based methods are the best in image synthesis



BigGAN (FID 6.95) Diffusion (FID 4.59)



arXiv:2105.05233v4 [cs.LG] 1 Jun 2021

Diffusion Models Beat GANs on Image Synthesis

Prafulla Dhariwal*
OpenAI
prafulla@openai.com

Alex Nichol*
OpenAI
alex@openai.com

Abstract

We show that diffusion models can achieve image sample quality superior to the current state-of-the-art generative models. We achieve this on unconditional image synthesis by finding a better architecture through a series of ablations. For conditional image synthesis, we further improve sample quality with classifier guidance: a simple, compute-efficient method for trading off diversity for fidelity using gradients from a classifier. We achieve an FID of 2.97 on ImageNet 128×128, 4.59 on ImageNet 256×256, and 7.72 on ImageNet 512×512, and we match BigGAN-deep even with as few as 25 forward passes per sample, all while maintaining better coverage of the distribution. Finally, we find that classifier guidance combines well with upsampling diffusion models, further improving FID to 3.94 on ImageNet 256×256 and 3.85 on ImageNet 512×512. We release our code at https://github.com/openai/guided-diffusion.

1 Introduction

Figure 1: Selected samples from our best ImageNet 512×512 model (FID 3.85)

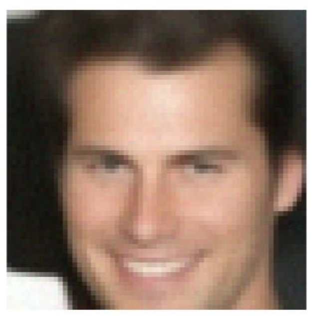
Discovery 3: Targeting Perceptual Quality

Suppose that we need to denoise the following image:

Noisy ($\sigma = 100$)

Image Denoiser

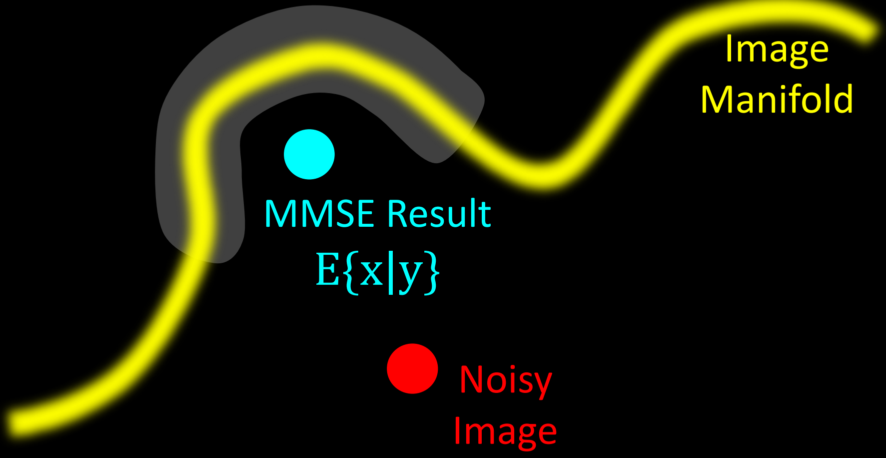
$\uparrow$
σ



Denoised image

Original

Should we be pleased with this result? It seems to be a bit … blurry, no? Why?



Minimum Mean-Squared-Error (MMSE) denoisers are great for MSE result, but their result falls outside the manifold

Michael Elad
The Computer-Science Department
The Technion

Discovery 3: Targeting Perceptual Quality

Question: How can we denoise an image while targeting "High Perceptual Quality"?

High perceptual quality image denoising with a posterior sampling cgan G Ohayon, T Adrai, G Vaksman, M Elad, P Milanfar Proceedings of the IEEE/CVF International Conference on Computer Vision …	3	2021
Stochastic image denoising by sampling from the posterior distribution B Kawar, G Vaksman, M Elad Proceedings of the IEEE/CVF International Conference on Computer Vision …	5	2021

Answer: Denoise by sampling from the posterior $\mathrm{P}(\mathrm{x}|\mathrm{y})$

Practical Implication: We consider 2 methods

- Training a deep denoiser via CGAN, or
- Iterated use of an MMSE denoiser $\mathrm{D}(\cdot, \sigma)$

These methods produce a multitude of possible solutions

Discovery 3: Targeting Perceptual Quality

Let's have a closer look at the Stochastic Image Denoiser:

Task: Draw a sample from $P(x|y)$ where $[y = x + n, n \sim \mathbb{N}(0, \sigma_0^2 \mathbf{I})]$

- Start with a random noise image $\hat{x}_0$
- Climb to a more probable image by the iterative equation:

$$\hat{x}_{k+1} = \hat{x}_k + a \cdot \underbrace{\nabla \log P(\hat{x}_k|y)}_{} + b \cdot z_k$$

Langevin with a conditional Score

Bayes rule

$$= \nabla \log P(\hat{x}_k) + \nabla \log P(y|\hat{x}_k)$$

$$= \underbrace{\hat{x}_k - D(\hat{x}_k, \sigma)}_{\text{Approx. Score}} + \underbrace{\nabla \log P(y|\hat{x}_k)}_{\text{Looks like a simple Gaussian Distribution}}$$

Michael Elad
The Computer-Science Department
The Technion

Discovery 3: Targeting Perceptual Quality

Let's have a closer look at the Stochastic Image Denoiser:

$$\nabla \log P(\hat{x}_k|y) = \hat{x}_k - D(\hat{x}_k, \sigma) + \nabla \log P(y|\hat{x}_k)$$

- As we use the Annealed Langevin algorithm, there are two noise signals to consider:
  - Measurement's noise: $n \sim \mathbb{N}(0, \sigma_0^2 \mathbf{I})$
  - Synthetic annealing noise: $v \sim \mathbb{N}(0, \sigma_k^2 \mathbf{I})$ for $\sigma_0 > \sigma_1 > \sigma_2 \ \cdots \ > \sigma_N > 0$
- Implication: We recover a sequence of gradually less noisy images $\hat{x}_k$ where their noise v is assumed to be a portion of n

$$\nabla \log P(\hat{x}_k|y) = \hat{x}_k - D(\hat{x}_k, \sigma_k) + \frac{y - \hat{x}_k}{\sigma_0^2 - \sigma_k^2}$$

Michael Elad
The Computer-Science Department
The Technion

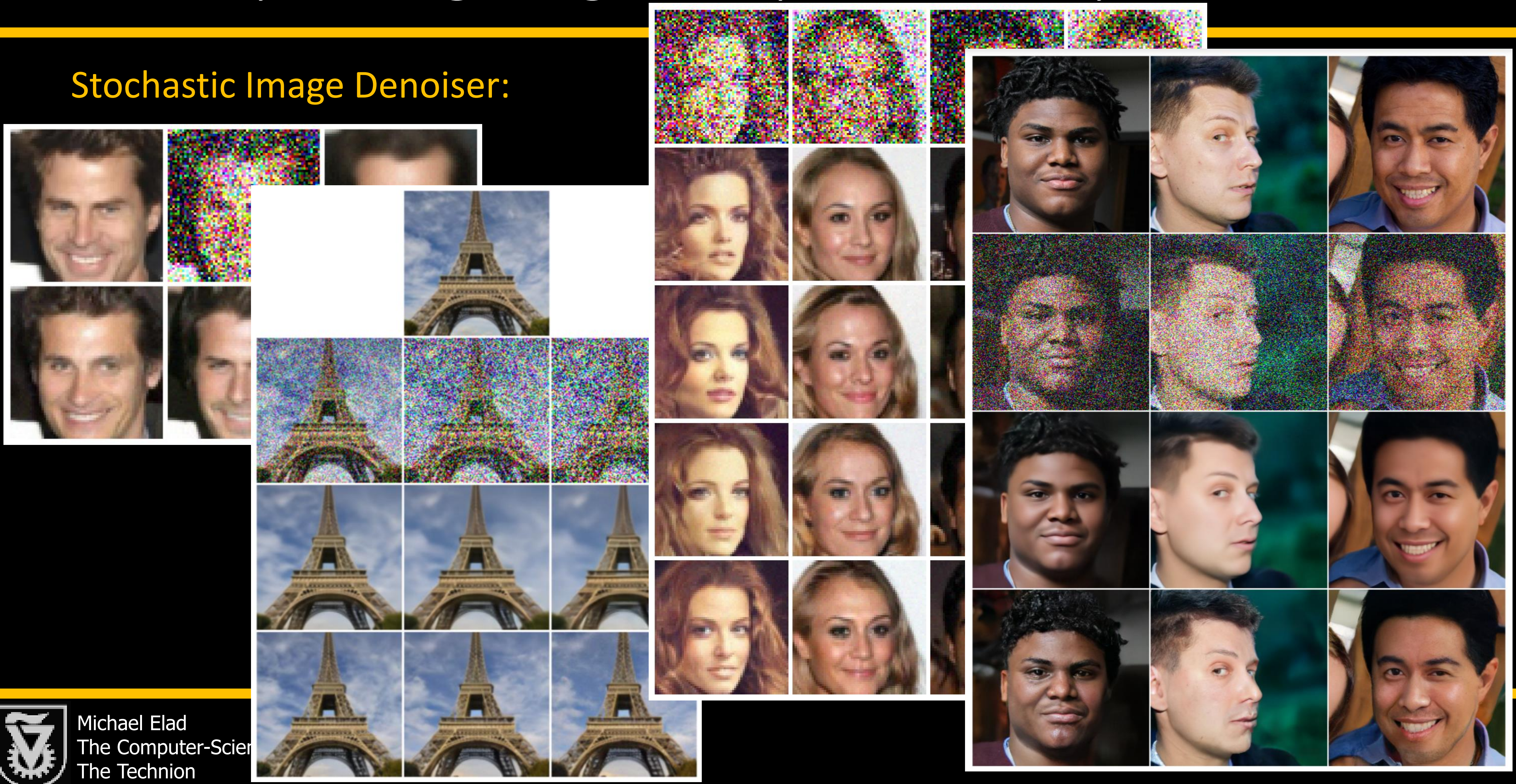
Discovery 3: Targeting Perceptual Quality

Stochastic Image Denoiser:

- We start from a noisy image ($\sigma \approx 150$ in this example)
- Then gradually denoise it using (conditional) annealed Langevin dynamics

Discovery 3: Targeting Perceptual Quality

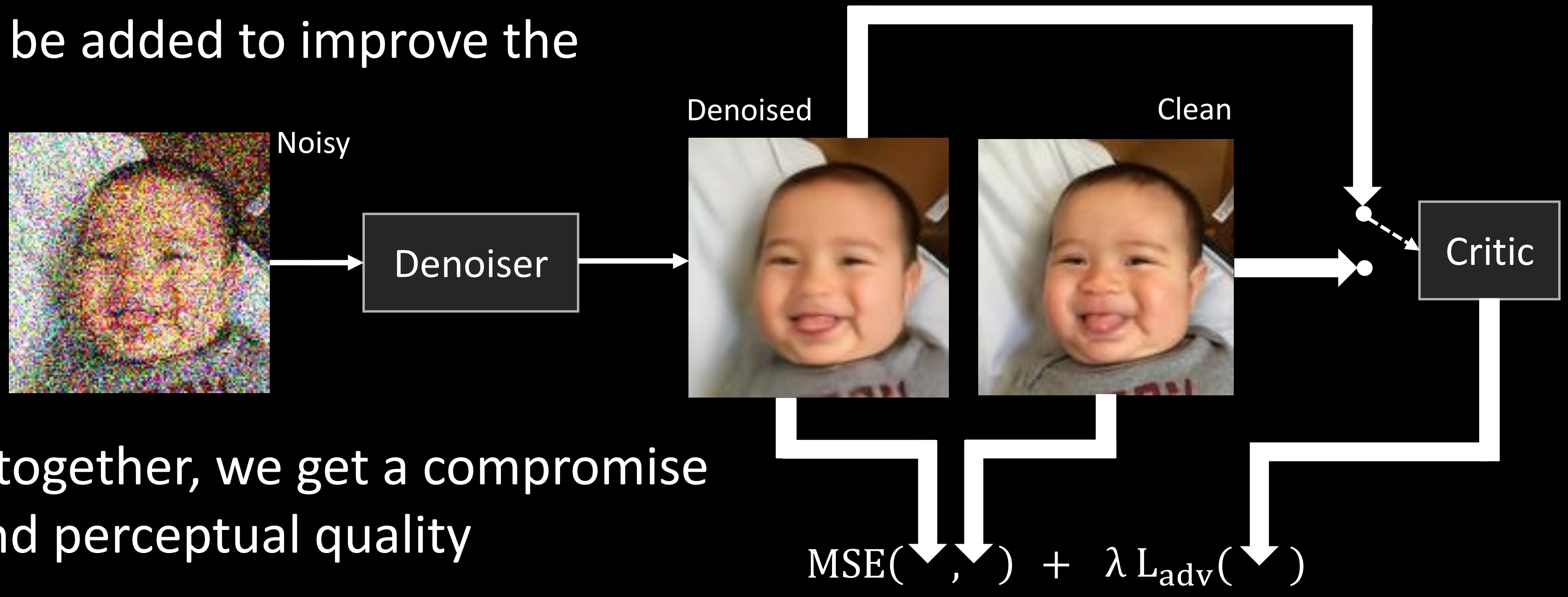
Stochastic Image Denoiser:



Discovery 3: Targeting Perceptual Quality

Let's have a closer look at the Conditional GAN Denoiser:

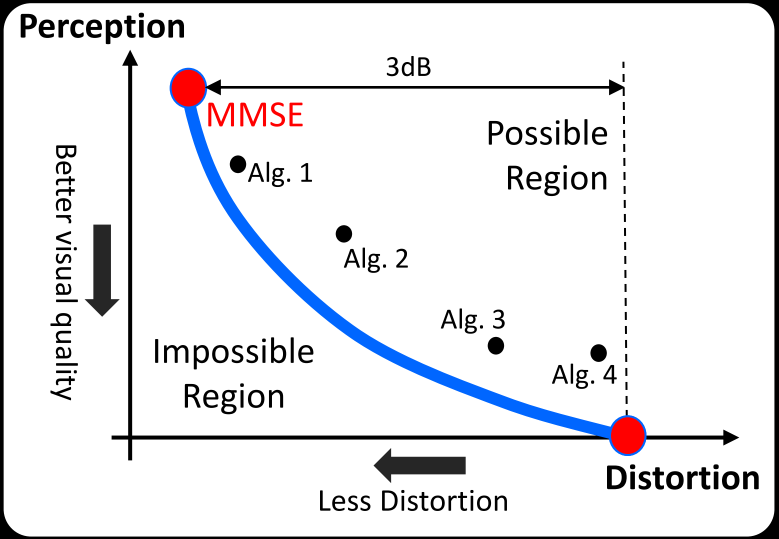
- Typical design approach: Optimize a distortion measure (e.g. MSE) between the denoised and the ideal images
- Adversarial loss could be added to improve the perceptual quality



- However, when used together, we get a compromise between distortion and perceptual quality

Discovery 3: Targeting Perceptual Quality

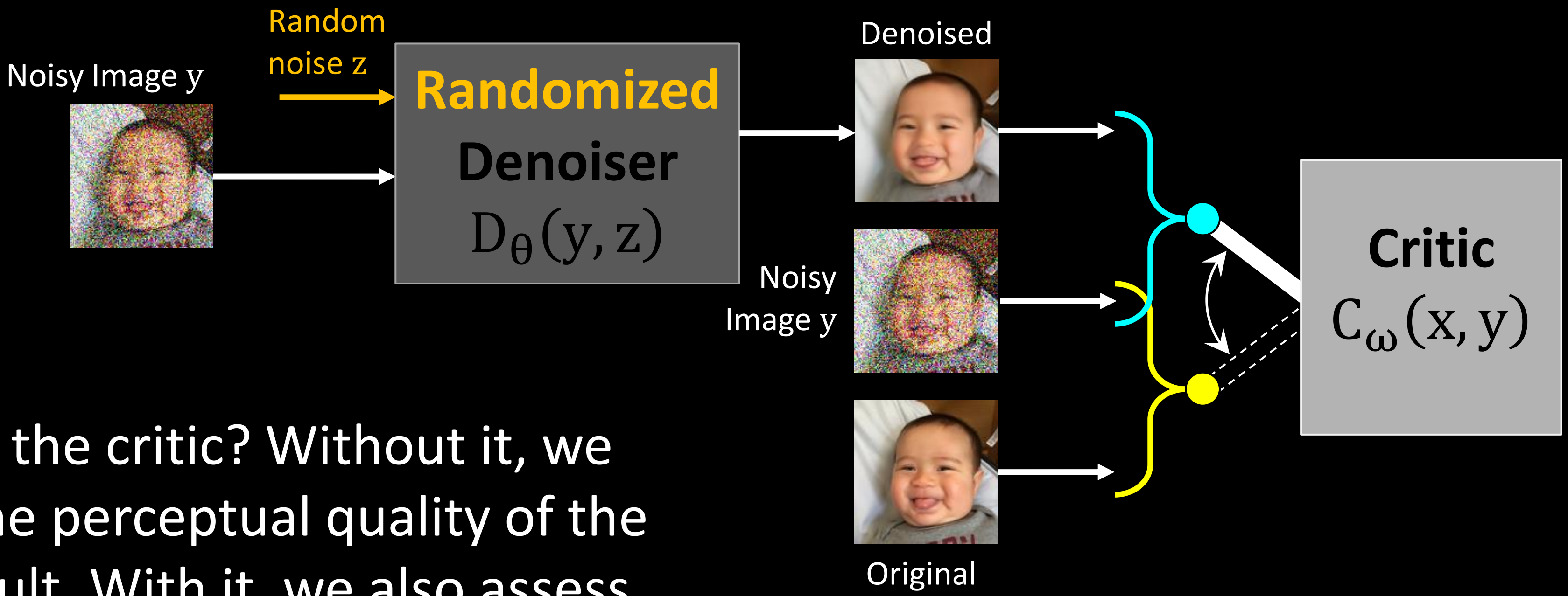
- For ill-posed restoration tasks, perceptual quality performance comes at the expense of its distortion [Blau & Michaeli 2017]
- We aim for best perceptual quality
- The posterior distribution attains perfect perceptual quality, compromising 3dB compared to the MMSE
- We propose to sample from the posterior via a Conditional GAN mechanism (PSCGAN)



$x \sim P_X$ $y \sim P_{Y|X=x}$ Samples from $P_{X|Y=y}$

Discovery 3: Targeting Perceptual Quality

The PSCGAN Architecture:



Why use y in the critic? Without it, we check only the perceptual quality of the denoised result. With it, we also assess its denoising validity

Michael Elad
The Computer-Science Department
The Technion

Discovery 3: Targeting Perceptual Quality

What about the Loss?

- CGAN optimization leads to posterior sampling [Adler et al. 2018]:

$$\min_{\theta}\max_{\omega} \mathbb{E}_{X,Y}\left[C_{\omega}(x,y)\right] - \mathbb{E}_{D_{\theta},Y,Z}\left[C_{\omega}(D_{\theta},y)\right]$$

- However, this requires an unavailable balanced dataset to train on $\underbrace{\text{many } x\text{'s for each } y}_{\text{Impossible !!}}$ and $\underbrace{\text{many } y\text{'s for each } x}_{\text{Easy}}$

- On the other hand, we would like to avoid a penalty of the form: $\mathbb{E}_{X,Y,Z}\left[\|x - D_{\theta}(y,z)\|_2^2\right]$

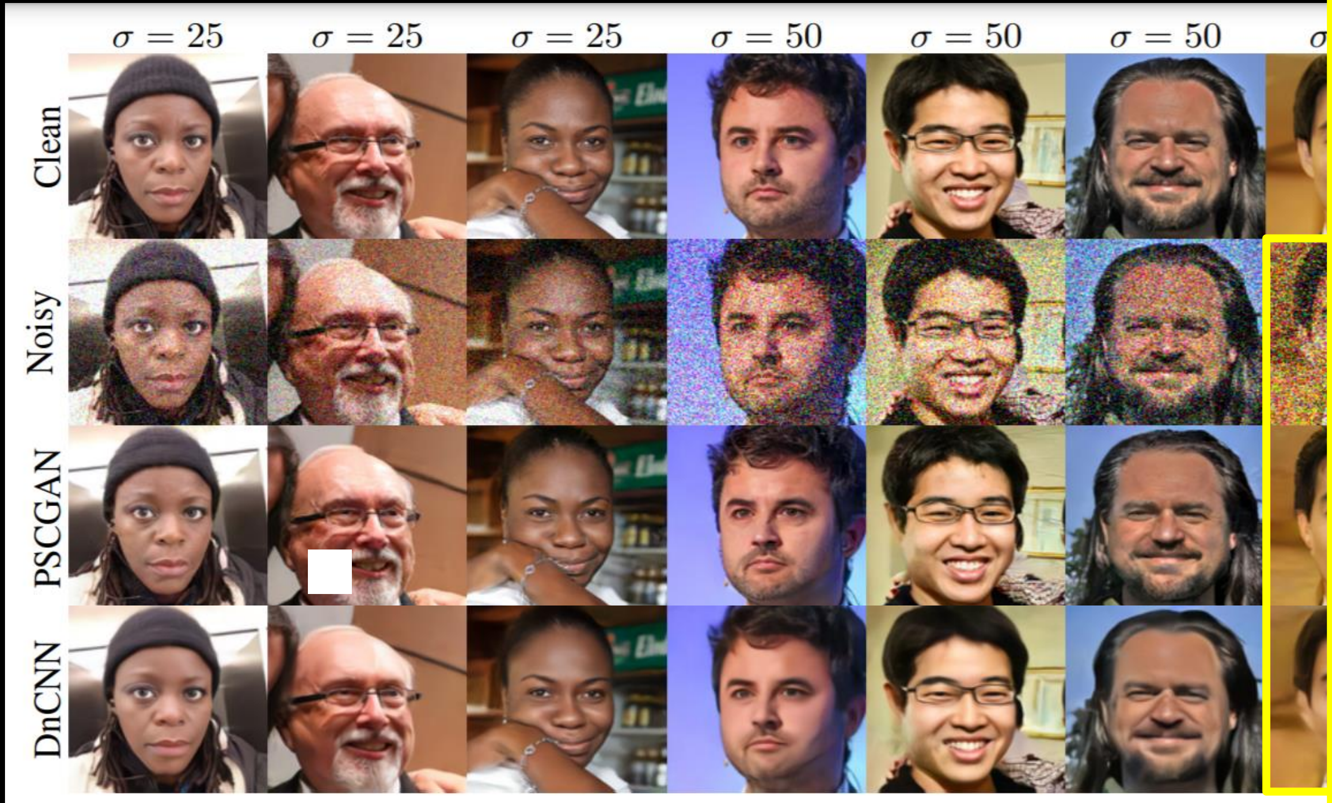
- Our remedy: adding an MMSE oriented penalty term of the form: $\mathbb{E}_{X,Y}\left[\|x - \mathbb{E}_z[D_{\theta}|y]\|_2^2\right]$

- This gives the MMSE result "for free" (averaging many instances)

Michael Elad
The Computer-Science Department
The Technion

Discovery 3: Targeting Perceptual Quality

CGAN:



Michael Elad
The Computer-Science Department
The Technion

Oh … and One Last Thing

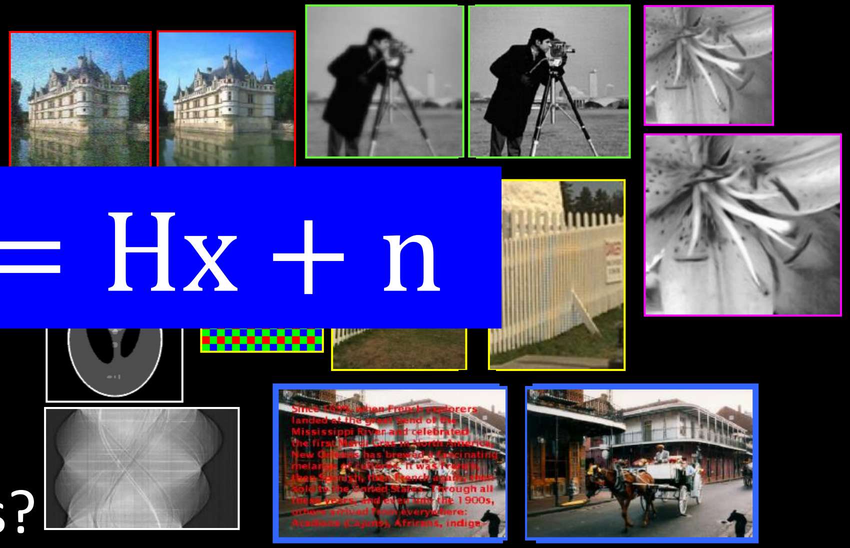
Back to Inverse Problems

- Goal: Recovery from corrupted measurements

  De-Noising De-Blurring

  In-Painting De-Mosaicing

  Tomography Image Scale-Up & super-resolution



$$y = Hx + n$$

- Can we suggest a "sampler" from $P(x|y)$ for handling all these problems, for getting "perfect looking" results?

- Answer: Yes! Use Langevin dynamics again, in an adapted form

Snips: Solving noisy inverse problems stochastically 4 2021
B Kawar, G Vaksman, M Elad
Advances in Neural Information Processing Systems 34

Denoising Diffusion Restoration Models 2022
B Kawar, M Elad, S Ermon, J Song
arXiv preprint arXiv:2201.11793

Back to Inverse Problems

- The idea is similar to our high-perceptual denoising, with necessary changes for considering the degradation operator H …
- Starting naively, using Bayes theorem, we need to calculate

$$\nabla \log \mathrm{P}(\mathrm{x_i}|\mathrm{y}) = \nabla \log \mathrm{P}(\mathrm{x_i}) + \nabla \log \mathrm{P}(\mathrm{y}|\mathrm{x_i})$$

- We know that $\mathrm{y} = \mathrm{Hx} + \mathrm{n}$ and thus:

$$\nabla \log\mathrm{P}(\mathrm{y}|\mathrm{x_i}) = \nabla \log\mathrm{P}(\mathrm{y} - \mathrm{Hx_i}|\mathrm{x_i}) =$$

$$\nabla \log\mathrm{P}(\mathrm{Hx} + \mathrm{n} - \mathrm{Hx} - \mathrm{Hv_i}|\mathrm{x_i}) = \nabla \log\mathrm{P}(\mathrm{n} - \mathrm{Hv_i}|\mathrm{x_i})$$

- However, … while $\mathrm{n} - \mathrm{Hv_i}$ is a simple Gaussian, it's dependency on $\mathrm{x_i}$ in non-trivial, so how do we proceed from here?

Michael Elad
The Computer-Science Department
The Technion

Back to Inverse Problems

- Step 1: Use SVD for decoupling the measurements $\mathrm{H} = \mathrm{U\Sigma V^T}$:

$$\mathrm{U^T y = U^T \underbrace{\left[U\Sigma V^T(x_i - v_i) + n\right]}_{y = Hx + n} = \Sigma V^T(x_i - v_i) + U^T n = \boxed{\Sigma\tilde{x}_T - \Sigma\tilde{v}_T + n_T = y_T}}$$

$$\Rightarrow \boxed{\mathrm{y_T[k] = \sigma_k \tilde{x}_T[k] - \sigma_k \tilde{v}_T[k] + n_T[k]}}$$

Decouple $\mathrm{\tilde{x}_T[k]} \leftrightarrow \mathrm{\tilde{v}_T[k]}$ by choosing $\mathrm{\tilde{v}_T[k]}$ to be a portion of $\mathrm{n_T[k]}$

- Thus, we can apply the Langevin dynamics algorithm on $\mathrm{\tilde{x}_T = V^T x_i}$ given $\mathrm{y_T = U^T y}$ and sample from the conditional
- Bottom line: An MMSE denoiser is used for a novel solution of inverse problems, this time targeting best perceptual quality

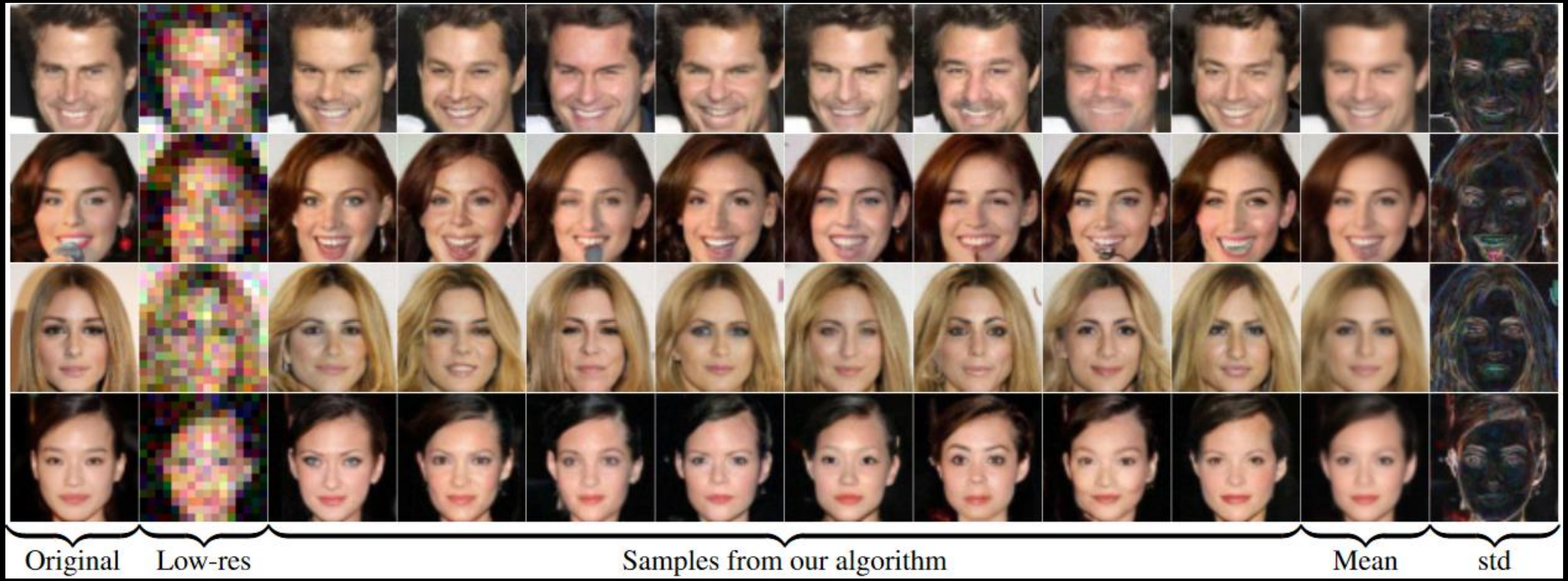
Back to Inverse Problems

Noisy Inpainting: A portion missing and noise with $\sigma_0 \approx 25$

Michael Elad
The Computer-Science Department
The Technion

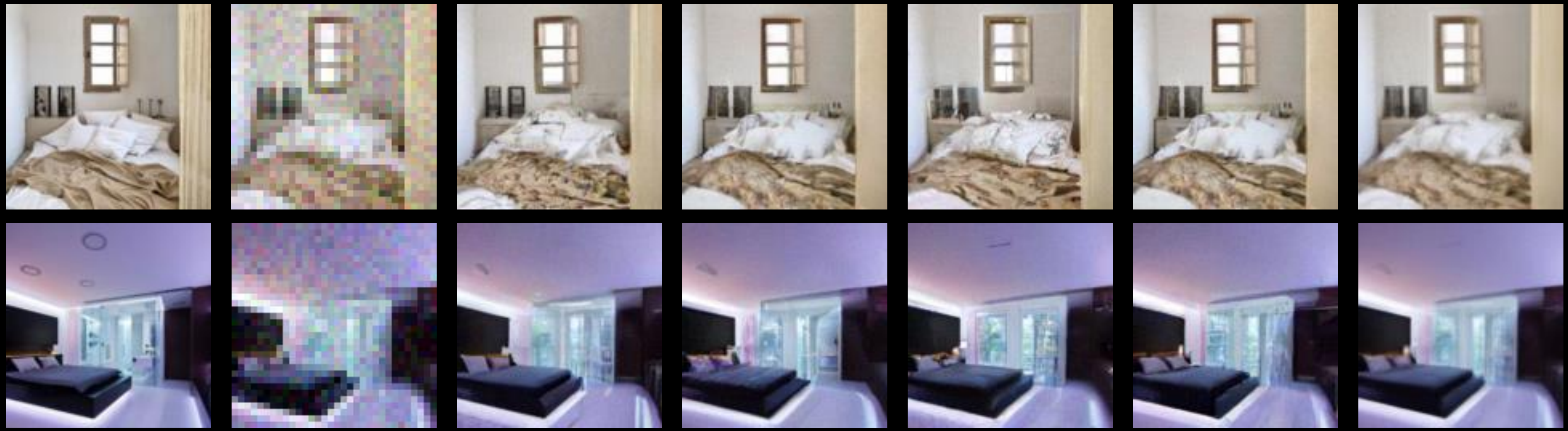
Back to Inverse Problems

Super resolution: downscaling by 4 with additive noise of $\sigma_0 \approx 25$



Back to Inverse Problems

Super resolution: downscaling by 4 with additive noise of $\sigma_0 \approx 12$

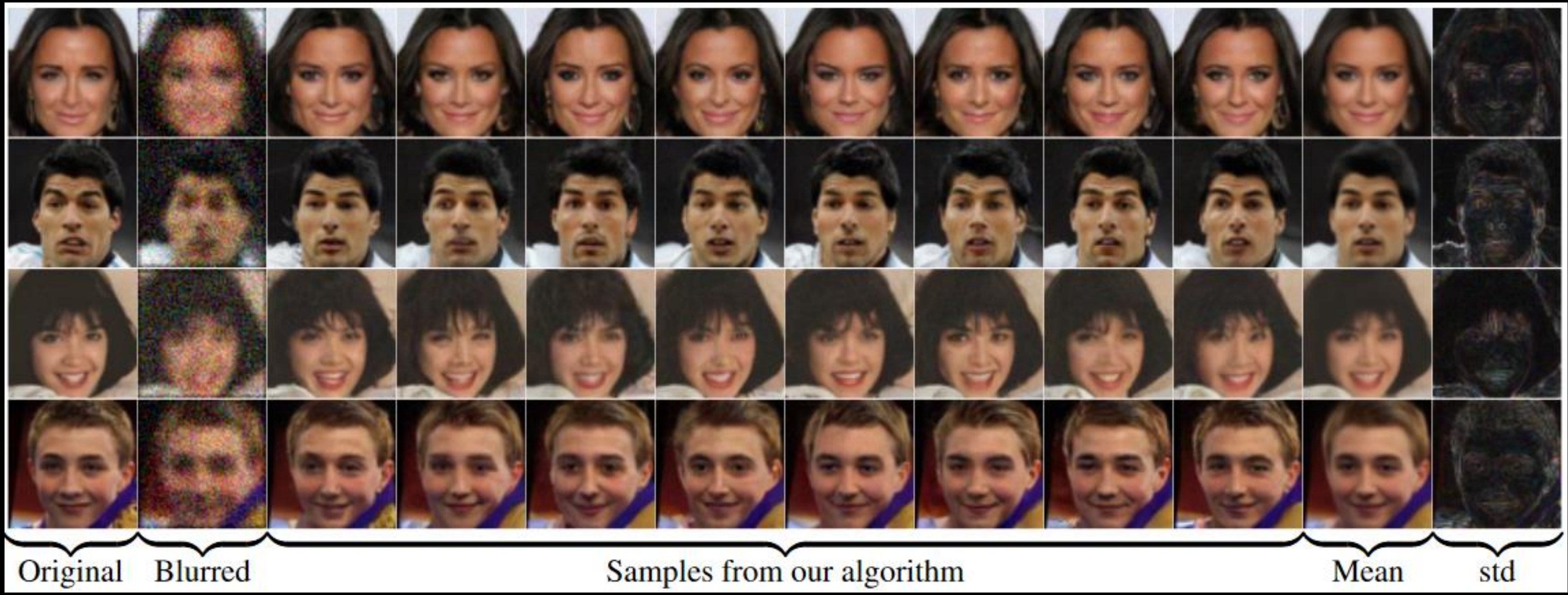


Original Degraded Samples from our algorithm Mean

Michael Elad
The Computer-Science Department
The Technion

Back to Inverse Problems

Deblurring: uniform 5×5 blur with additive noise of $\sigma_0 \approx 25$



Back to Inverse Problems

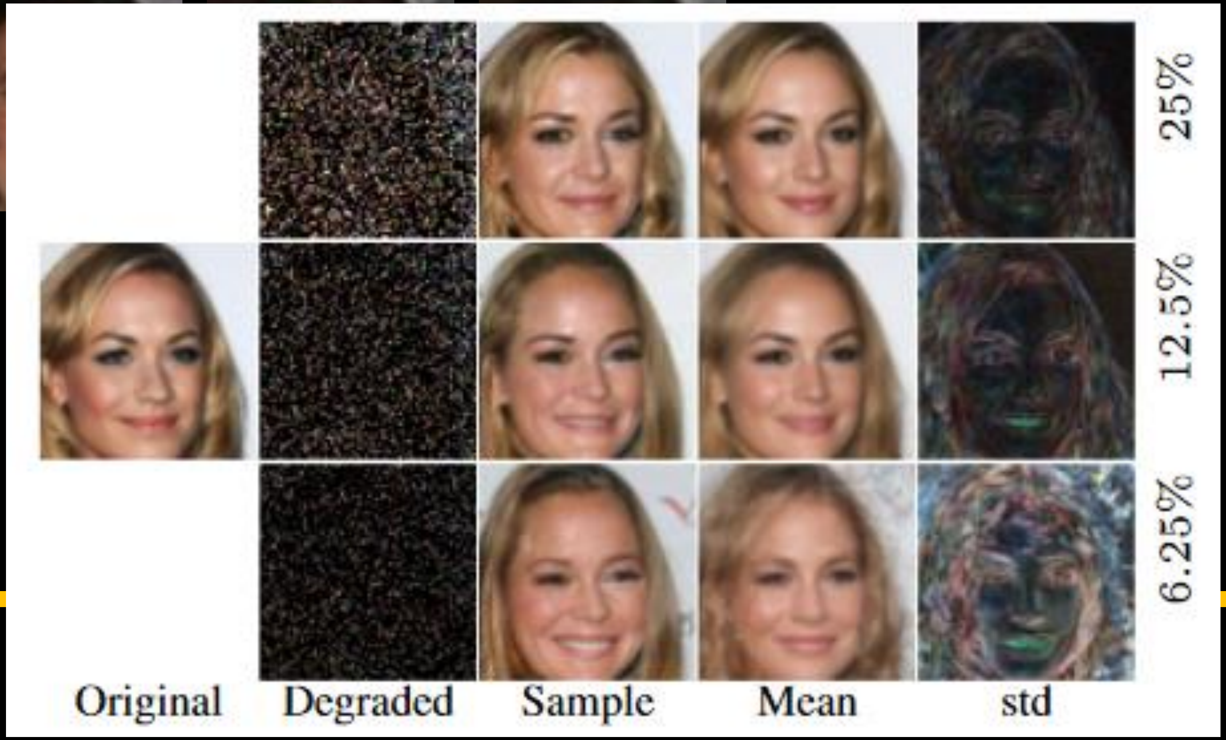
Compressive sensing (12.5%) with additive noise of $\sigma_0 \approx 25$

Original　Degraded　Samples from our algorithm　Mean

Back to Inverse Problems

Compressive sensing (12.5%) with additive noise of $\sigma_0 \approx 25$

Original Degraded Samples from our



Michael Elad
The Computer-Science Department
The Technion

Back to Inverse Problems

And just to remind you …

The proposed diffusion-based sampling scheme, while quite appealing, suffers from several key shortcomings:

- ❑ It is rather S L O W (many denoising activations)
- ❑ It is limited (as of now) to specific families of images
- ❑ Relying on SVD is cumbersome

See our recent work that answers (most of) these challenges:

Denoising Diffusion Restoration Models 2022
B Kawar, M Elad, S Ermon, J Song
arXiv preprint arXiv:2201.11793

Time to Summarize

What Have we Seen Today?

Suppose that we are given an MMSE denoiser $\mathrm{D}(\mathrm{y})$

WE CAN USE $\mathrm{D}(\mathrm{y})$ FOR ….

solving ANY inverse problem (PnP/RED)

synthesizing natural-looking images

denoising images while targeting high perceptual quality

solving ANY inverse problem with high perceptual quality

All the above are achieved by simply applying $\mathrm{D}(\mathrm{y},\sigma)$ iteratively

Summary

Image Denoising

… Not What You Think

1. There are so many opportunities and challenges in better understanding, designing, and proposing creative usage of image denoisers
2. Despite the fact that this has not been a talk about Deep-Learning, the presence of this field in the topics covered is prominent

Thank You

- The content of this lecture relies on ~10 papers that my group has worked on and published in the past several years
- Getting these results was enabled due to the amazing people I had the pleasure of collaborating with:

Peyman Milanfar, Yaniv Romano, Grisha Vaksman, Bahjat Kawar, Meyer Scetbon, Guy Ohayon, Theo Adrai

Thank You All for Attending

And special thanks to the organizers of this event:

- ❑ Zhong-Zhi Bai (Chinese Academy of Sciences, China)
- ❑ Walter Gander (ETH Zurich, Switzerland)
- ❑ Yu-Mei Huang (Lanzhou University, China)
- ❑ Lu Wang (Chinese Academy of Sciences, China)
- ❑ Wen-Ting Wu (Beijing Institute of Technology, China)

In dedication to **Gene H. Golub**
(February 29, 1932 - November 16, 2007)
The occasion of his 90th birthday