Image Denoising

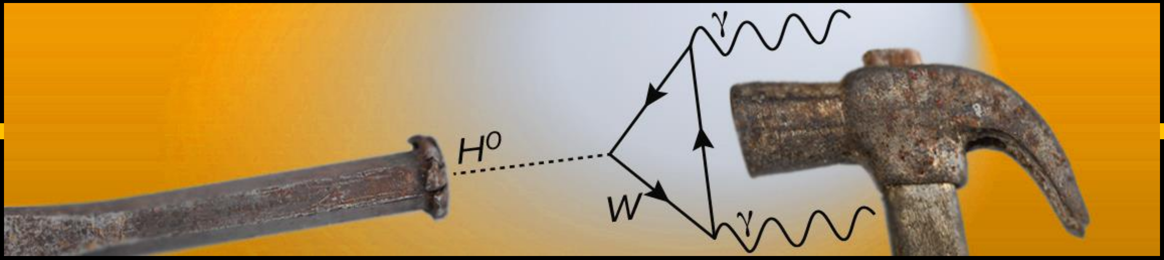
… Not What You Think

Michael Elad

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

Verily Research

November 1st 2023



This Lecture is About …

Image Denoising

Noisy image

Image Denoiser

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

Michael Elad
The Computer-Science Department
The Technion

Our Agenda

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

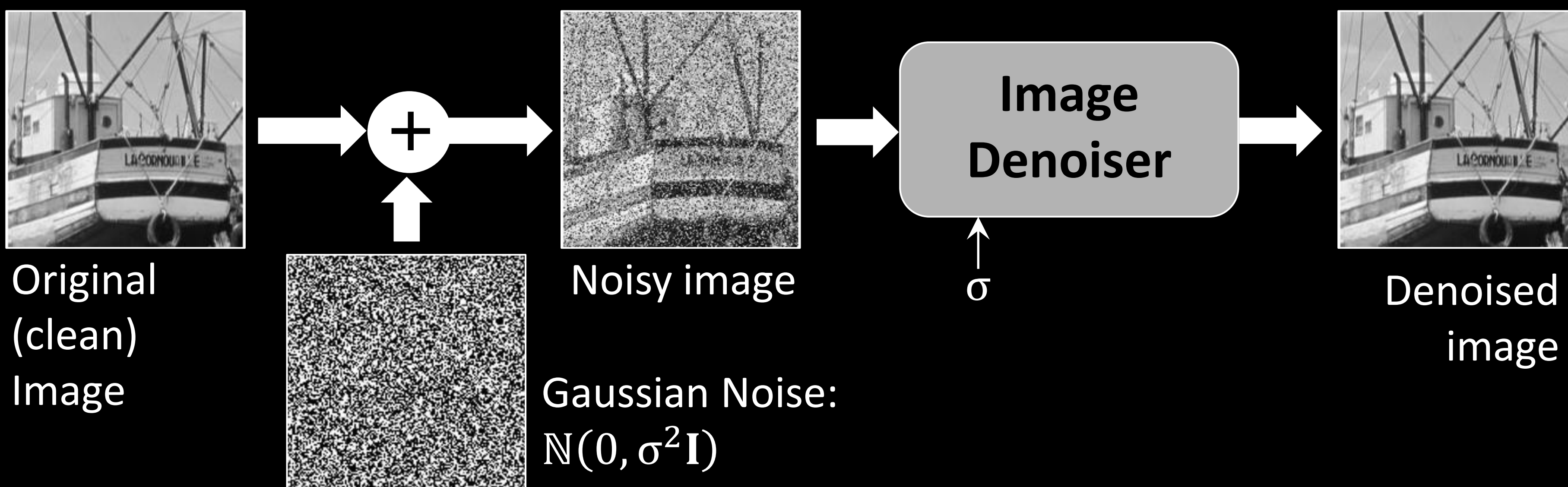
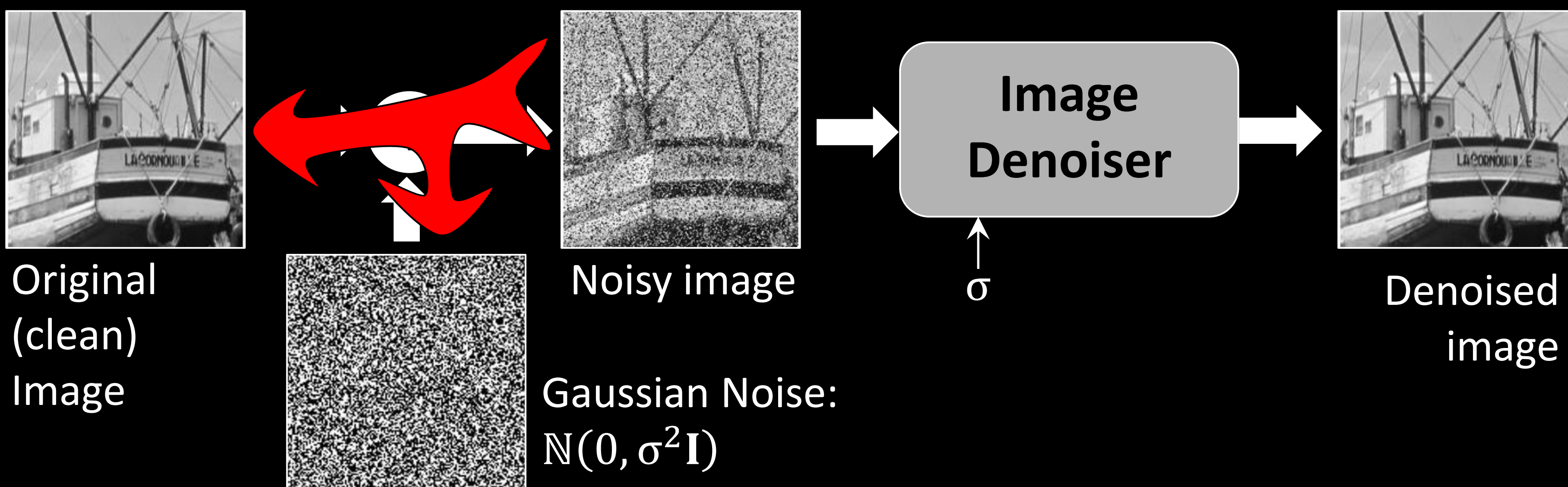


Image Denoising is Challenging

Image denoising is far from trivial task! Why?

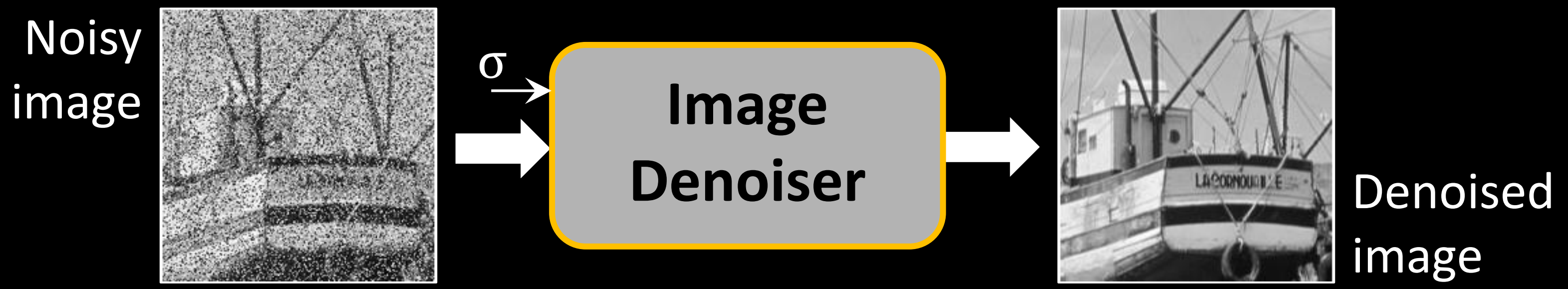
- Because our goal is to remove noise as much as possible while preserving the details in the image
- Denoising is essentially a highly ill-posed separation task



Michael Elad
The Computer-Science Department
The Technion

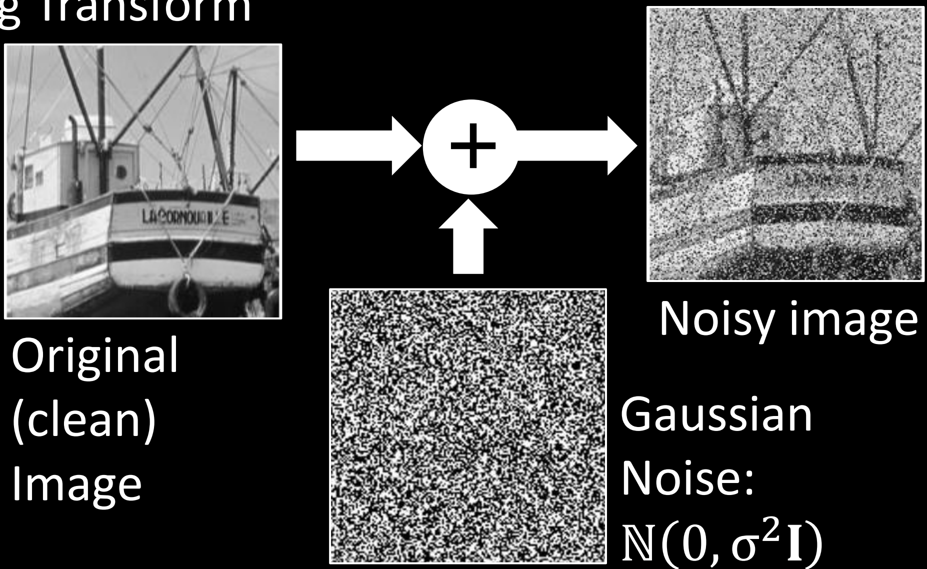
Why Work on Image Denoising?

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



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

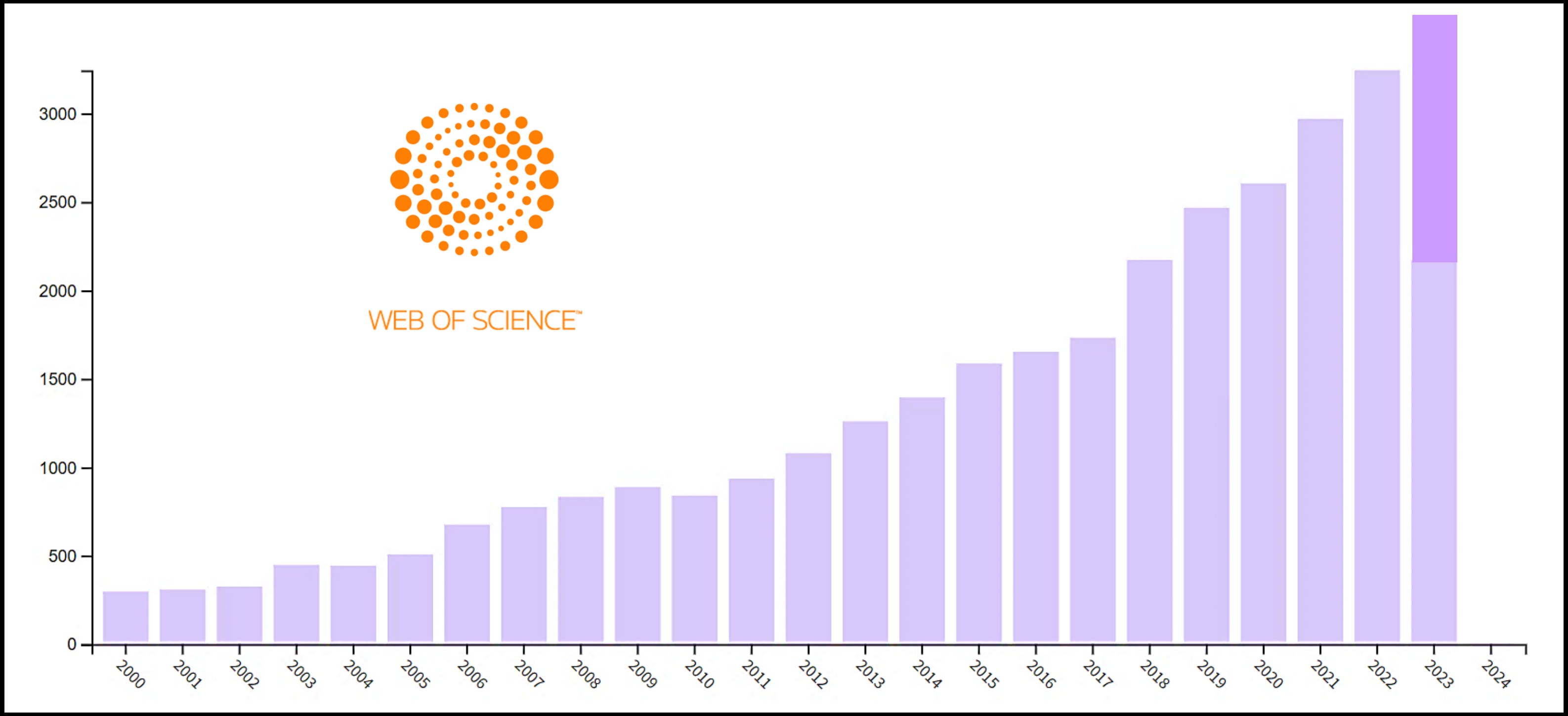
Noisy image

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 ~33,000 papers* on this subject, offering algorithms, theoretical analysis and so much more

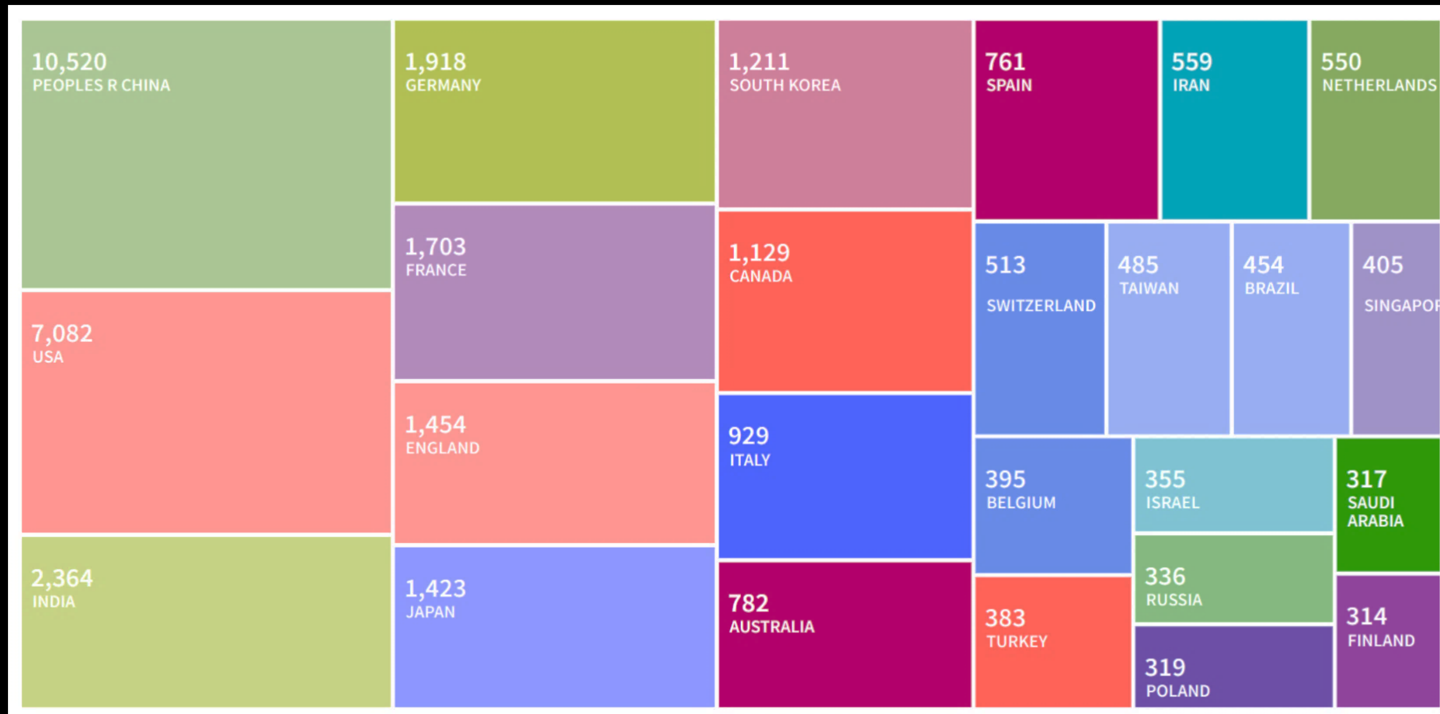


* Search done on October 26th 2023 in WoS, topic: ((image or video) and (denoising or (noise and remov) or clean))

Image Denoising: Little bit of History

Citing Articles:	
USA:	208188
China:	92111
France:	46506
Germany:	41808
England:	36459
Canada:	27033
Spain:	25317
Australia:	23502
Israel:	17711
India:	17604
Switz.:	17601
Japan:	17427
Italy:	16393
Nether.:	15609
Korea:	14119
Finland:	11969
Singapore:	9695
Belgium:	8383
Brazil:	7637
Taiwan:	6802
Iran:	5697
Russia:	4607

This research comes from all over the globe



… and it is heavily cited

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 $\rho(x)$ (e.g., forcing smoothness, sparsity, low-rank, self-similarity, …), and
- Formulate the denoising task as an optimization problem

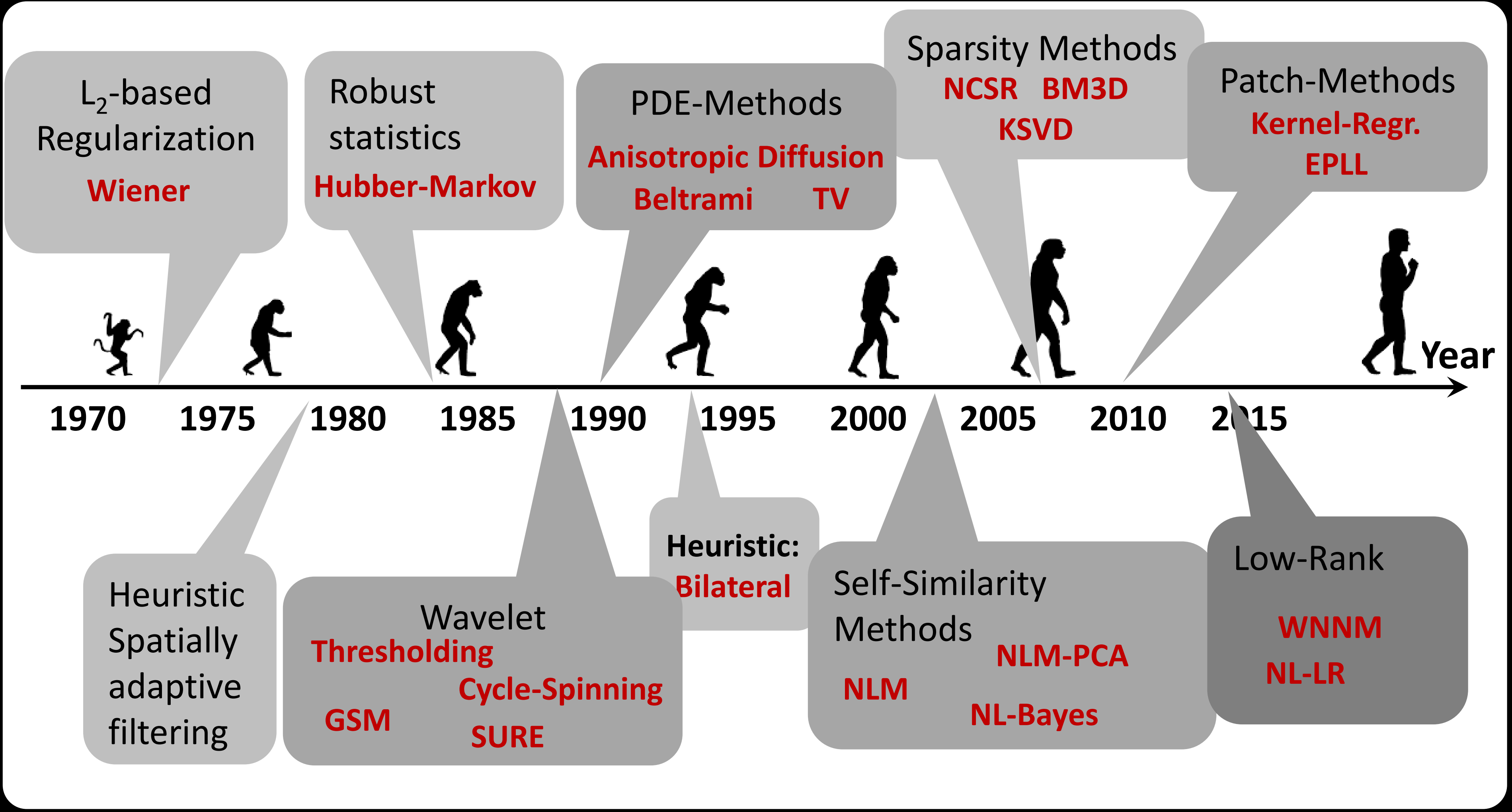
$$\hat{x} = \min_{x} \underbrace{\|x - y\|^2}_{\text{Likelihood}} + \underbrace{\rho(x)}_{\text{Prior}}$$

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

$$P(x) = C \cdot \exp\{-\rho(x)\}$$

This is the MAP estimate, which leads to an iterative or a direct algorithm for getting $\hat{x}$ from y

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 De

Priyam Chatterjee, *Student M*

***Abstract*—Image denoising has been a well studied the field of image processing. Yet researchers continue tention on it to better the current state-of-the-art. Re posed methods take different approaches to the probl their denoising performances are comparable. A pert tion then to ask is whether there is a theoretical limit t performance and, more importantly, are we there yet? manufacturers continue to pack increasing numbers o unit area, an increase in noise sensitivity manifests itself of a noisier image. We study the performance bounds fo denoising problem. Our work in this paper estimates a l on the mean squared error of the denoised result and co performance of current state-of-the-art denoising me this bound. We show that despite the phenomenal rece in the quality of denoising algorithms, some room fo ment still remains for a wide class of general images, an signal-to-noise levels. Therefore, image denoising is not**

***Index Terms*—Bayesian Cramér–Rao lower boun bias, bootstrapping, image denoising, mean squared er**

I. Introduction

IMAGE denoising has been a well-studied p the image processing community and continue

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 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

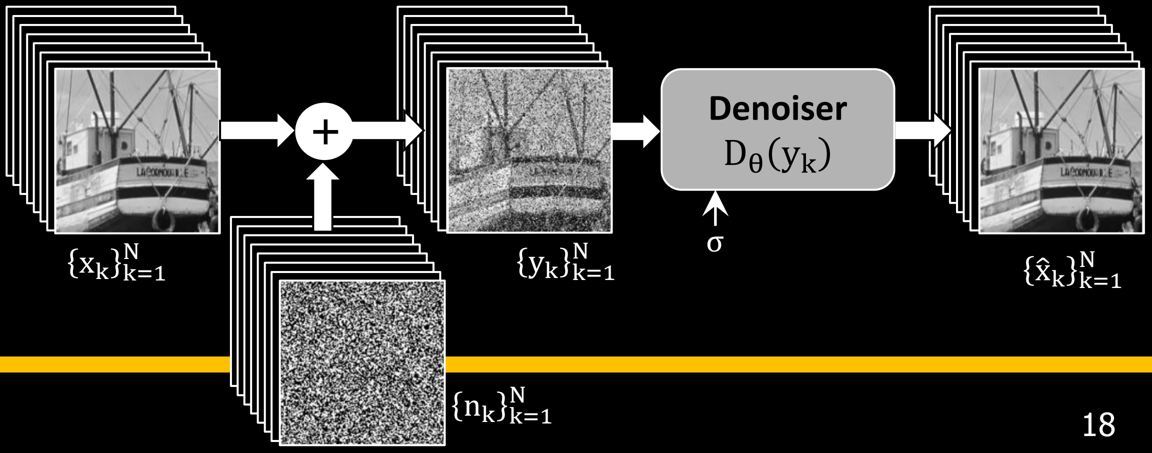
Design of Algorithms: Take 2

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:

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

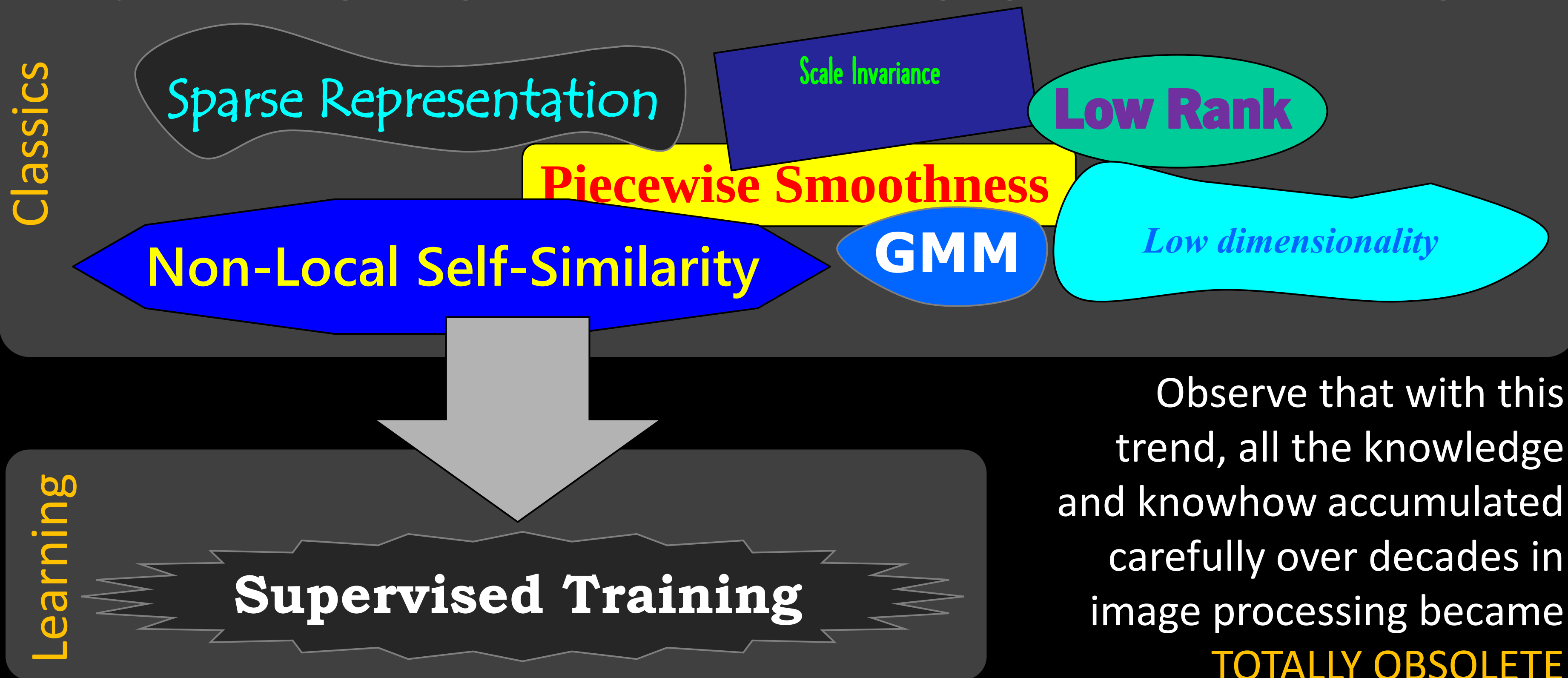


Michael Elad
The Computer-Science Department
The Technion

Image Denoising: A Paradigm Shift

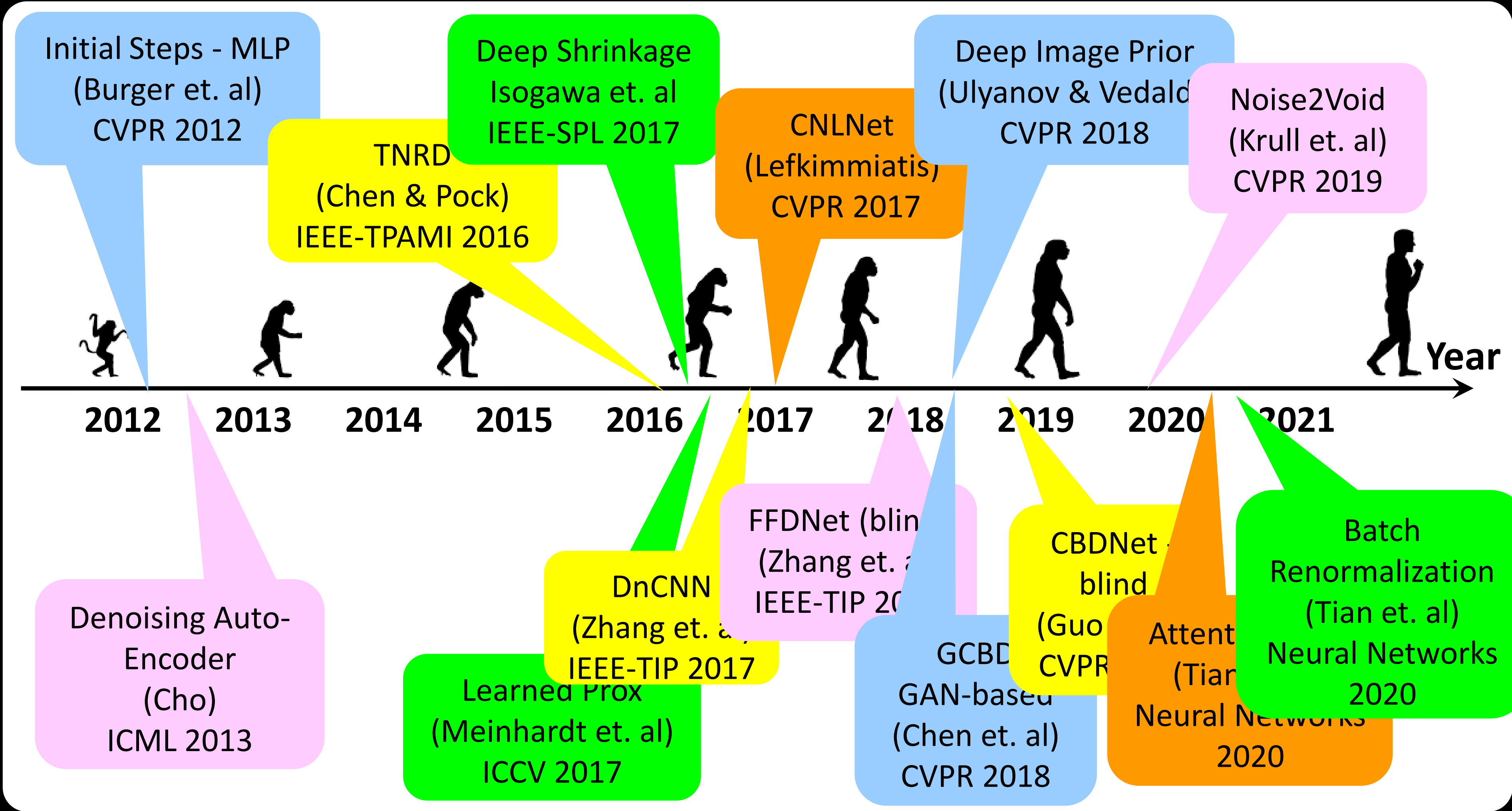
How can we design a denoiser?

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



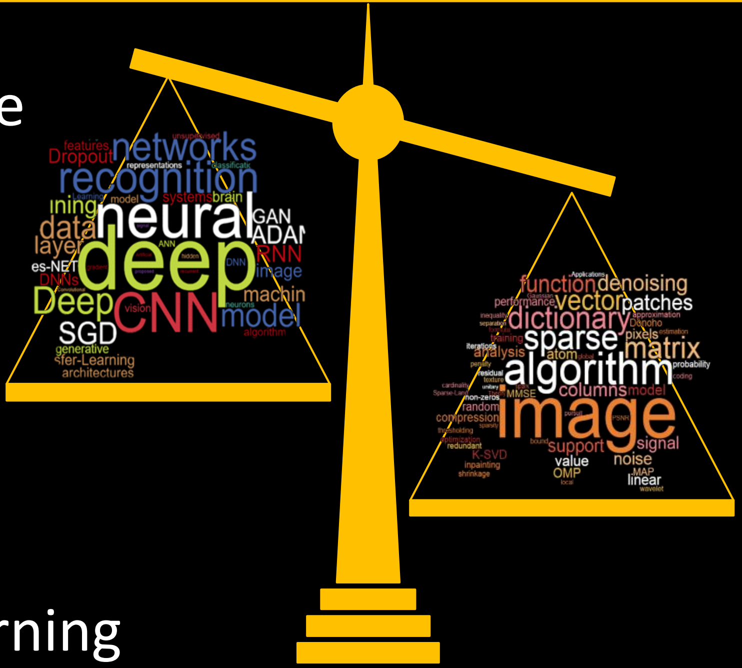
Observe that with this trend, all the knowledge and knowhow accumulated carefully over decades in image processing became TOTALLY OBSOLETE

Image Denoising: Recent Evolution



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

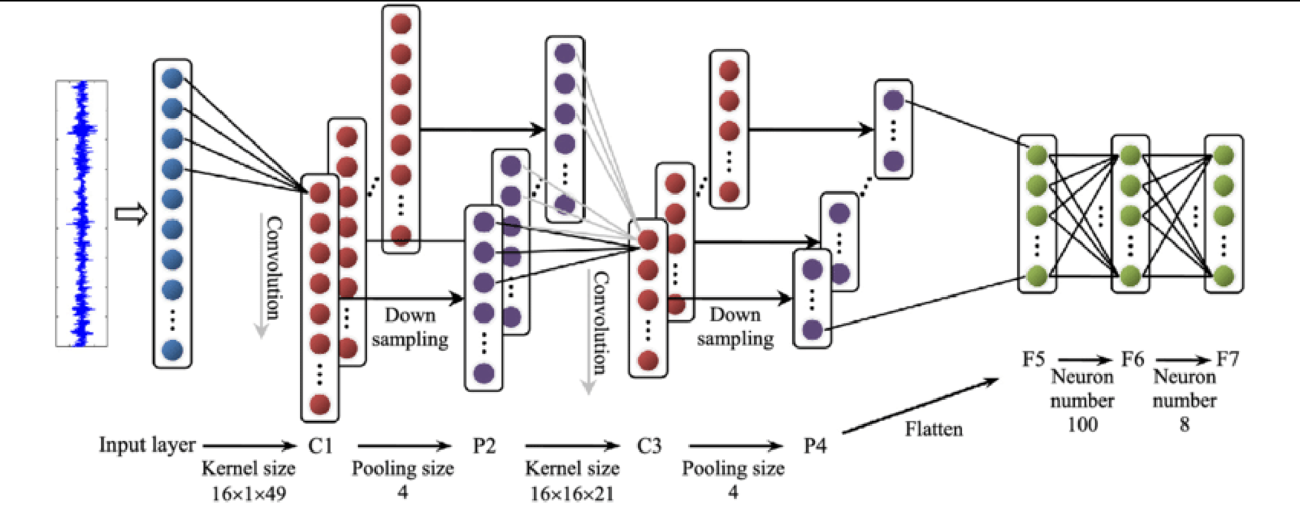
Michael Elad
The Computer-Science Department
The Technion

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

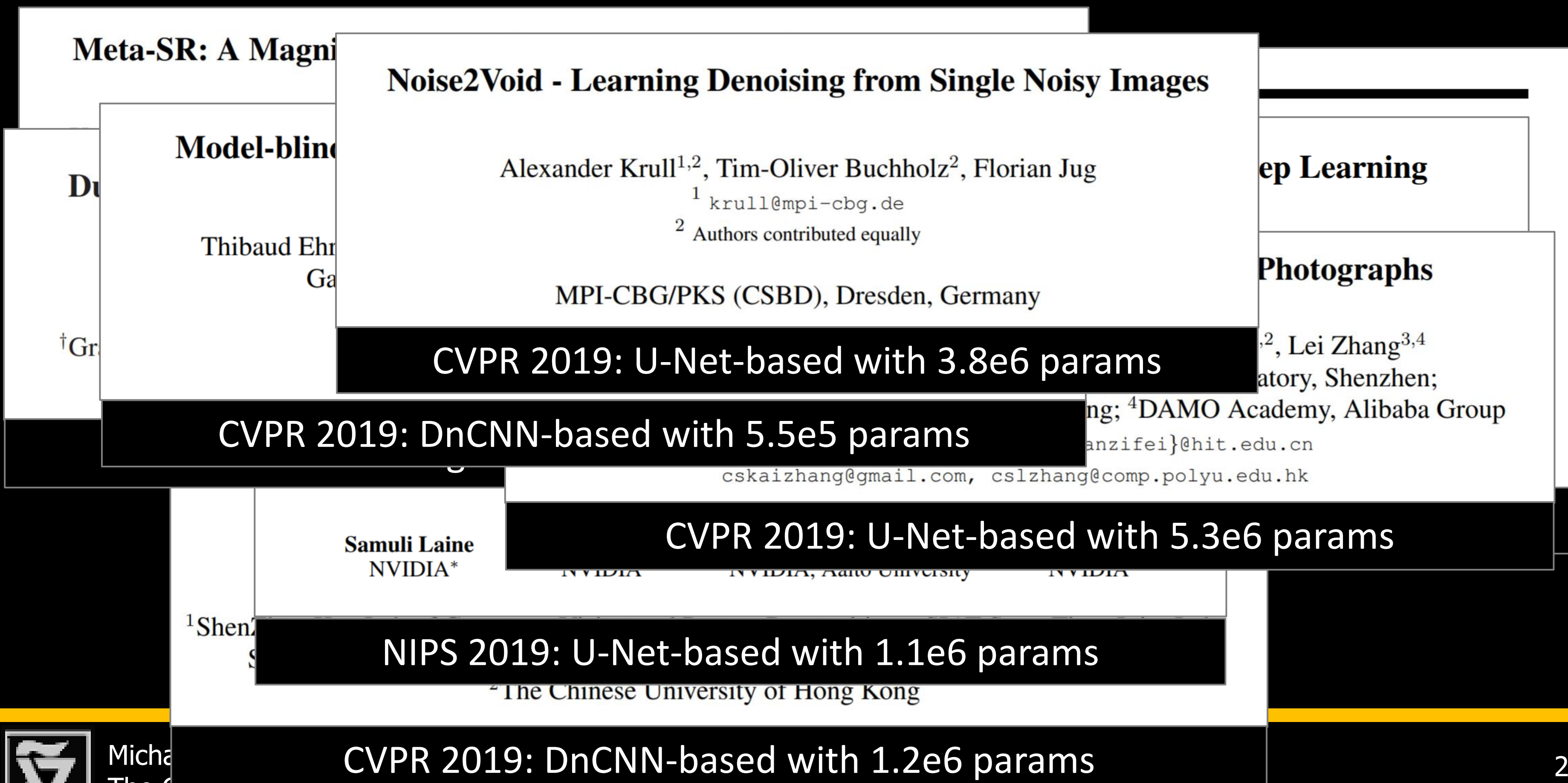
Option 2 – Pile and Guess a series of steps that mix known pieces such as convolutions, fully connected layer, batch-norm, ReLU, pooling, stride, skips, upscale/downscale, connections, ...
and add new "tricks"



Option 3 – Neural Architecture Search

Image Denoising Architectures

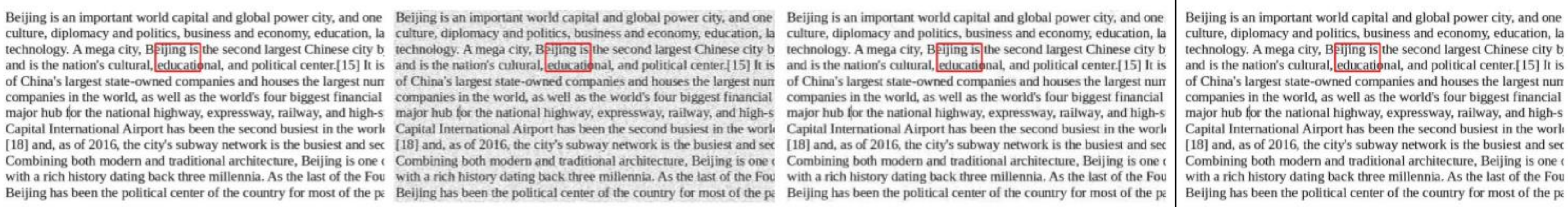
Here are several paper examples from CVPR/NIPS 2019 that illustrate these architectures



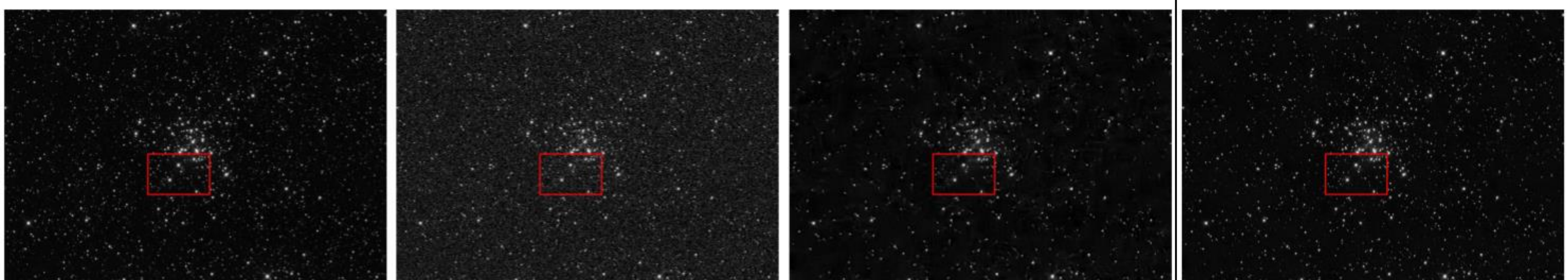
Alternative Architecture Design

- Message: 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]
  - 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]

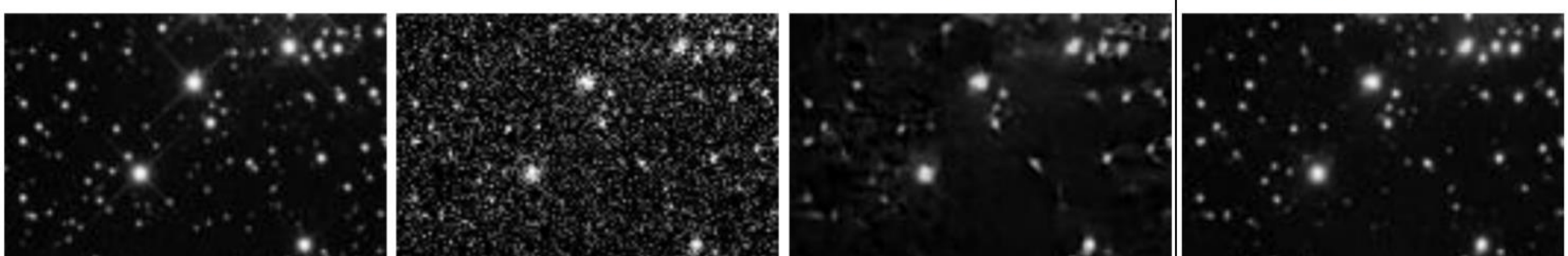
A Closer Look at Adaptation



(a) Clean text (704 × 356) (b) Noisy with $\sigma = 50$ (c) Denoised (d) Denoised



(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)

Our Focus Today: Recent Discoveries

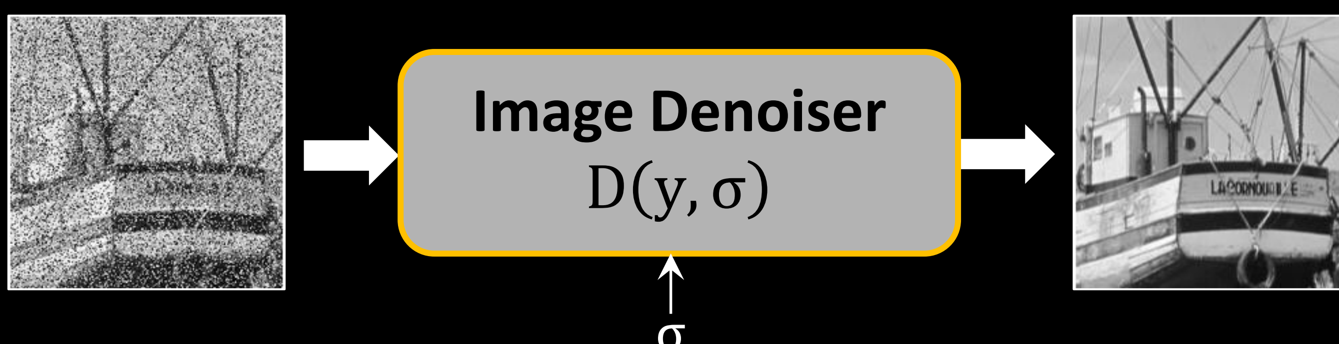
Our Focus Today

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



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

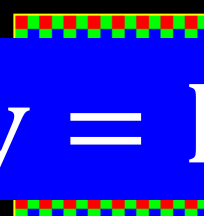
- ❑ De-Noising
- ❑ De-Blurring
- ❑ In-Painting
- ❑ De-Mosaicing
- ❑ Tomography
- ❑ Image Scale-Up & super-resolution

… and more …

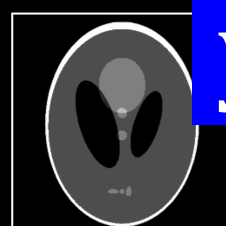
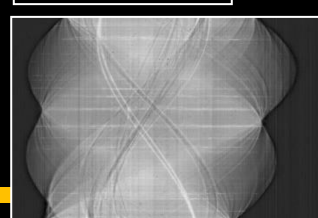








$$y = Hx + n$$

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 $\rho(x)$ (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 path for 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!

Michael Elad
The Computer-Science Department
The Technion

Discovery 1: Solving Inverse Problems

Question: Given a denoiser $\mathrm{D}(\mathrm{y},\sigma)$
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	922	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	670	2017

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

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

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

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:

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

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

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

Let's use the Steepest Descent

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

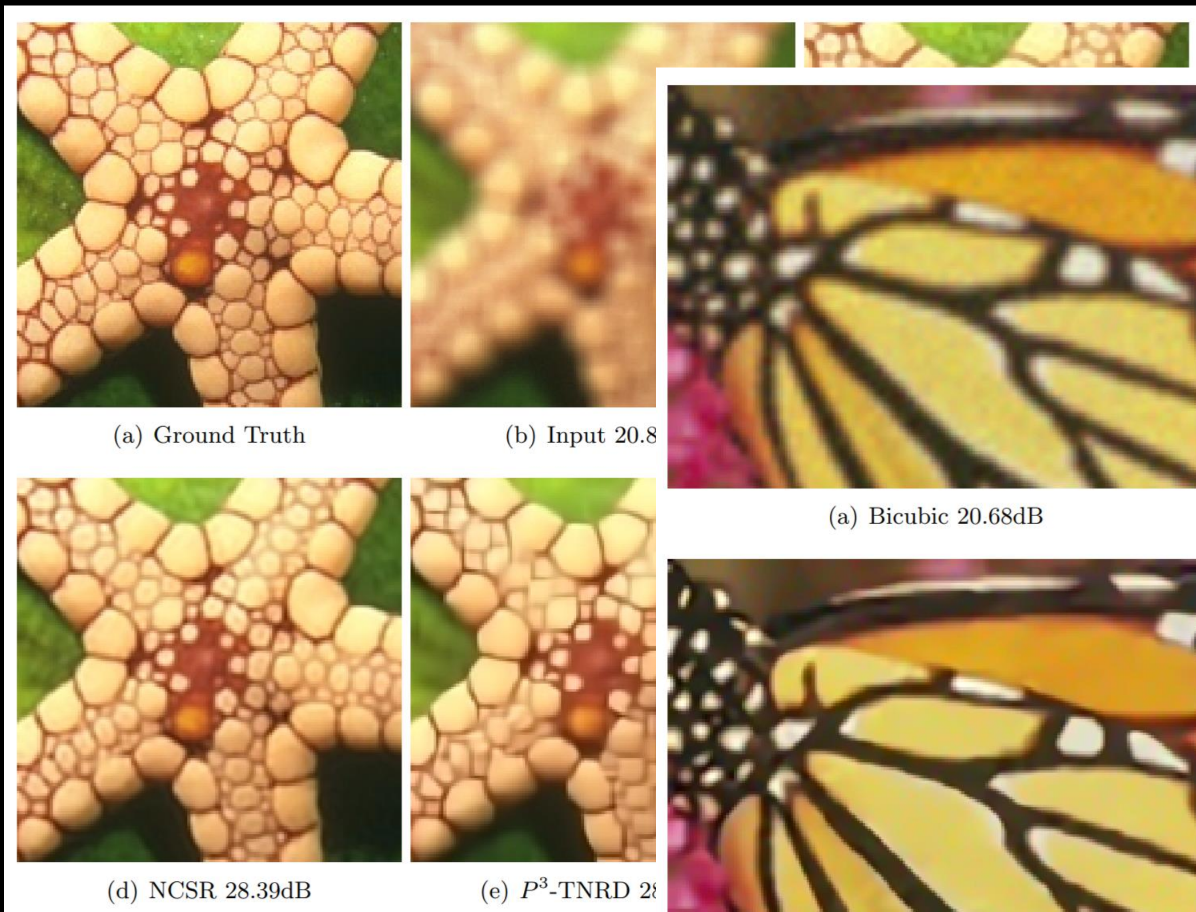
… and this way we got an iterated algorithm that keeps calling to a denoiser, and is guaranteed to achieve the global minimum

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

Michael Elad
The Computer-Science Department
The Technion

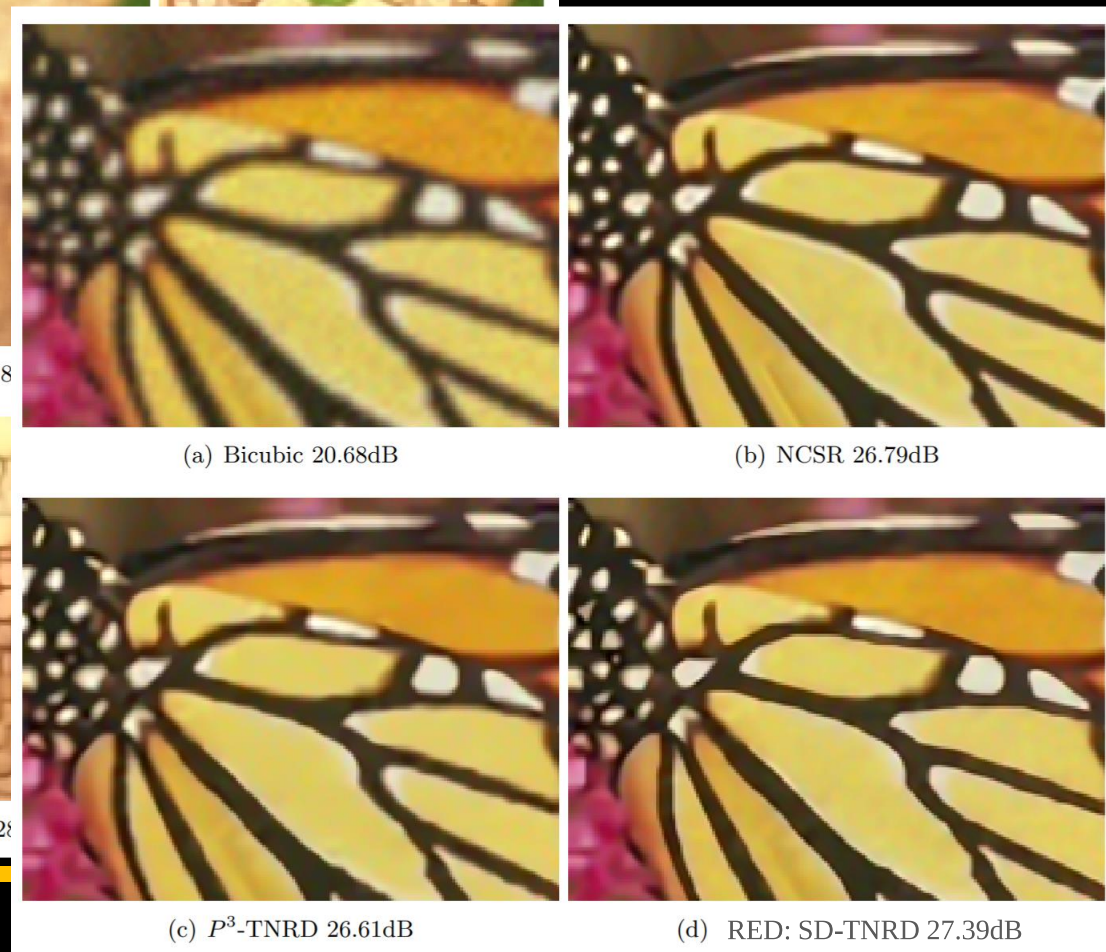
Discovery 1: Solving Inverse Problems

Here are some results for Deblurring and Super-Resolution



(a) Ground Truth (b) Input 20.8

(d) NCSR 28.39dB (e) P^3-TNRD 2



(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

- PnP and RED are heavily cited and extensively studied, owing to their generality and elegance

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	922	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	670	2017

- 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 …
- 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

Discovery 2: Image Synthesis

Question: Given a denoiser $\mathrm{D}(\mathrm{y}, \sigma)$ 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	1548	2019
Improved techniques for training score-based generative models Y Song, S Ermon Advances in neural information processing systems 33, 12438-12448	528	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	56	2021

Answer: Use $\mathrm{D}(\mathrm{y}, \sigma)$ 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 → $\mathrm{D}(\cdot, \sigma_2)$ → Simpl[e] Operati[on]

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:

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

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:

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

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

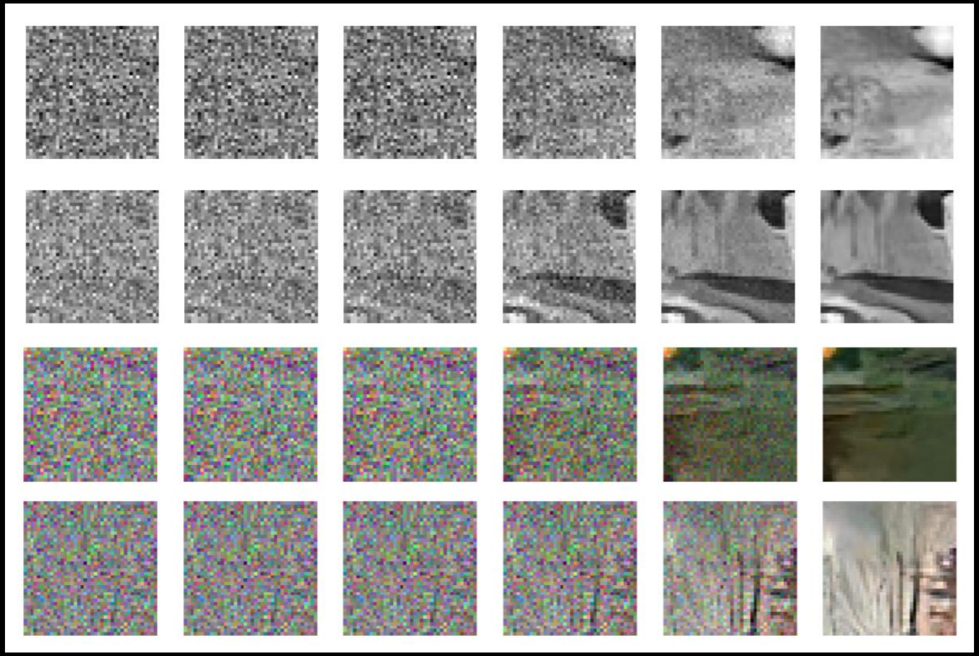
$$\text{with} \quad \sigma_0 > \sigma_1 > \sigma_2 \quad \cdots \quad > \sigma_N > 0$$

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

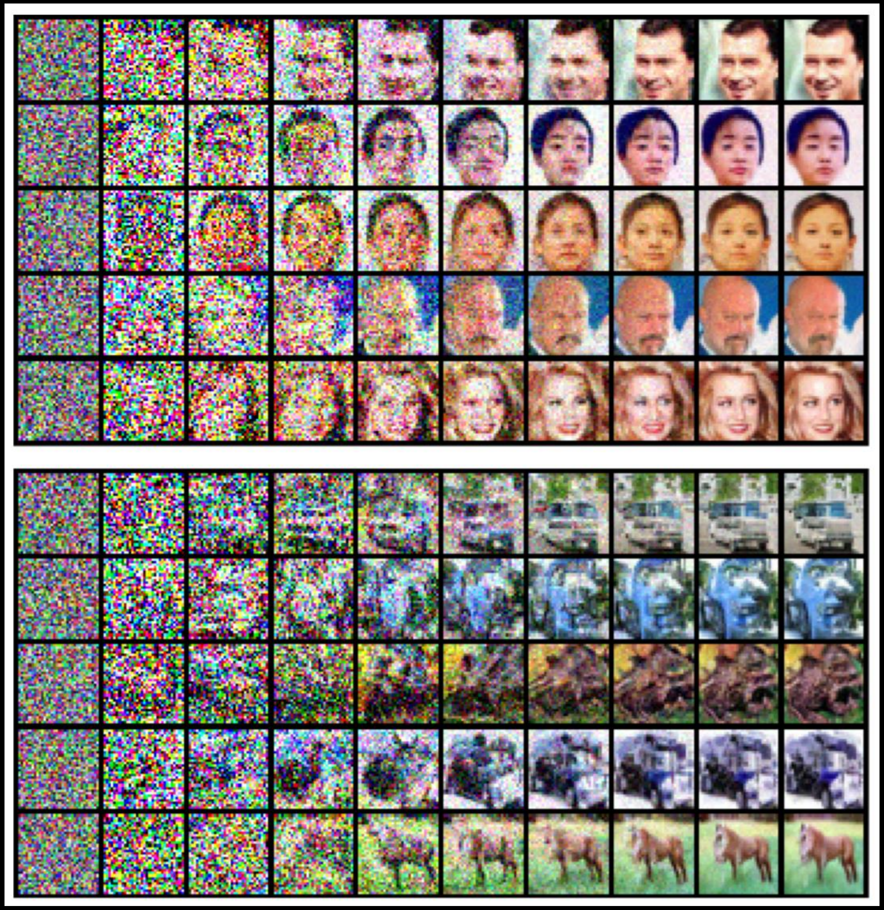
Michael Elad
The Computer-Science Department
The Technion

Discovery 2: Image Synthesis

Does it work? Here are some results



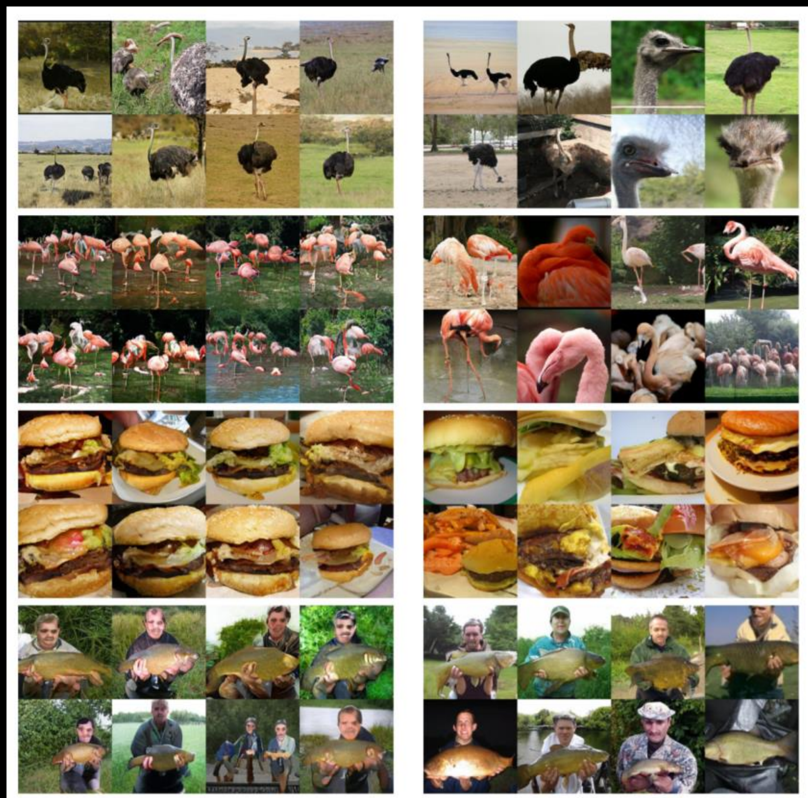
Kadkhodaie & Simoncelli



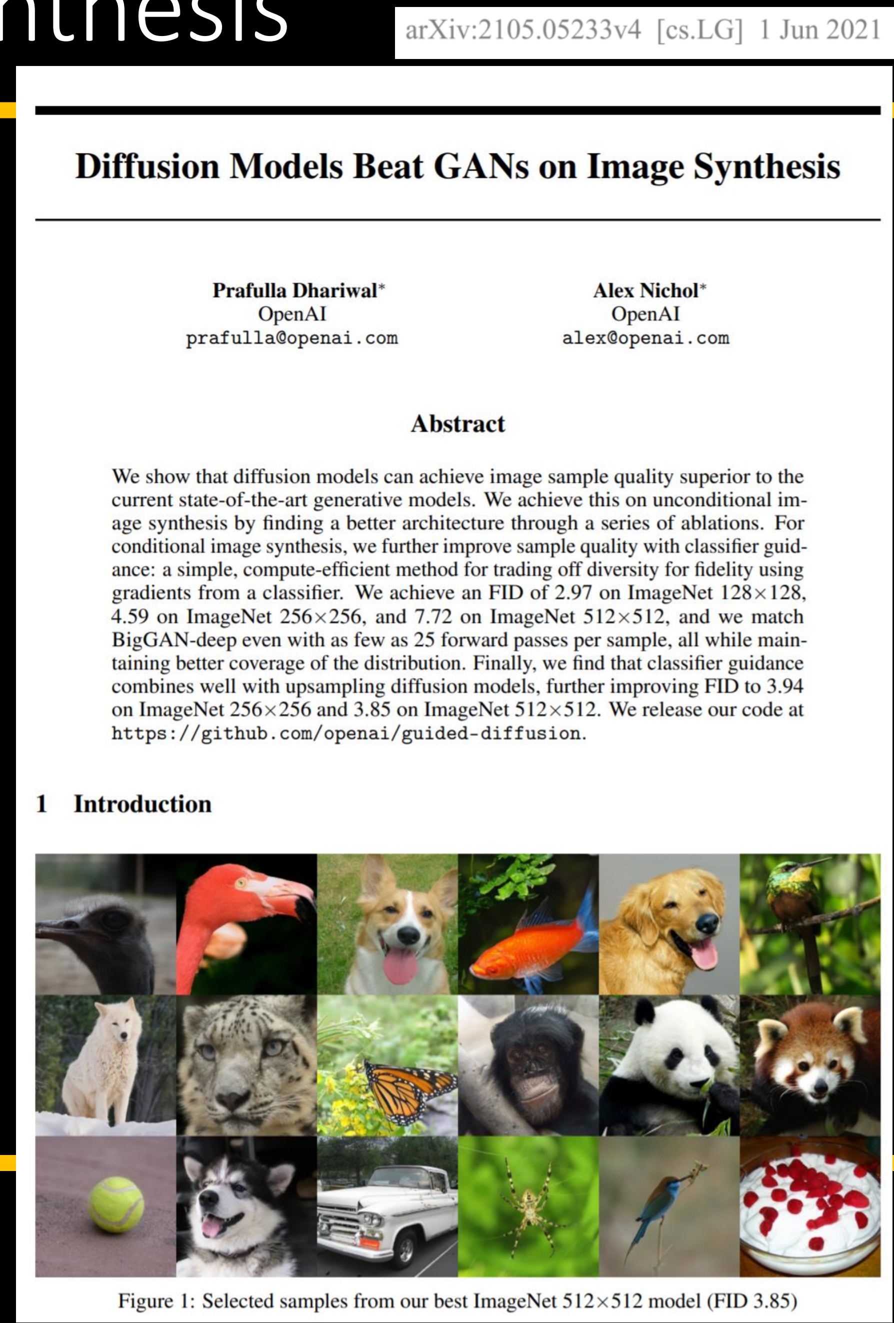
Song & Ermon

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)

Michael Elad
The Computer-Science Department
The Technion

Discovery 2: Image Synthesis

Surely, you have heard of …

Imagen

unprecedented photorealism × deep level of language understanding

Google

DALL·E 2

OpenAI

Michael Elad
The Computer-Science Department
The Technion

Discovery 2: Image Synthesis

Surely, you have heard of …

Stable Diffusion

stability.ai

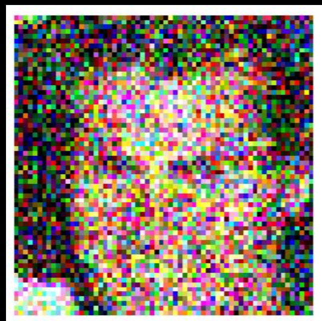
WAKEUP THE ARTIST IN YOU

Midjourney

ARTIFICIAL INTELLIGENCE ART GENERATOR

Discovery 3: Targeting Perceptual Quality

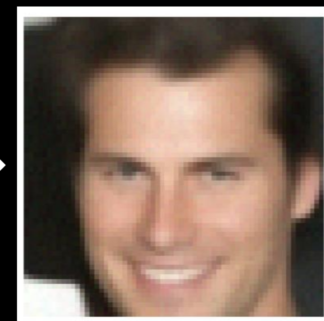
Suppose that we need to denoise the following image:



Noisy ($\sigma = 100$)

Image Denoiser

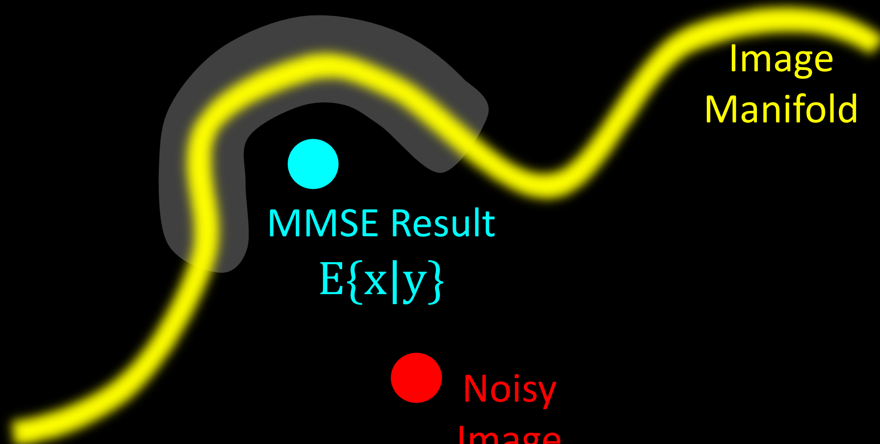
σ



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 …	23	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 …	41	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 $\mathrm{P(x|y)}$ where $[\mathrm{y = x + n, n \sim \mathbb{N}(0, \sigma_0^2 \mathbf{I})}]$

- 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

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

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

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: $v \sim \mathbb{N}\left(0, \sigma_k^2 \mathbf{I}\right)$ 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

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

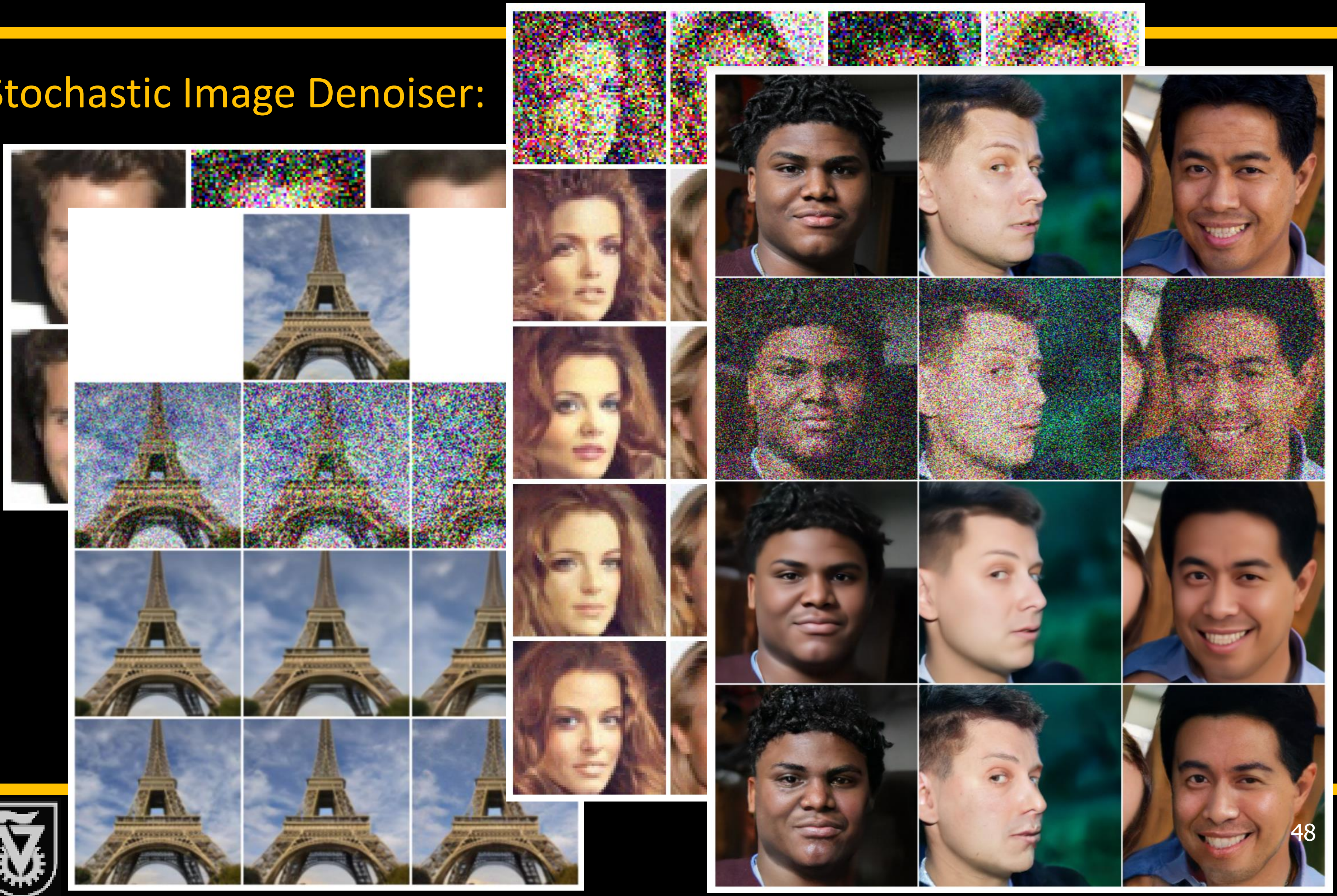
Discovery 3: Targeting Perceptual Quality

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

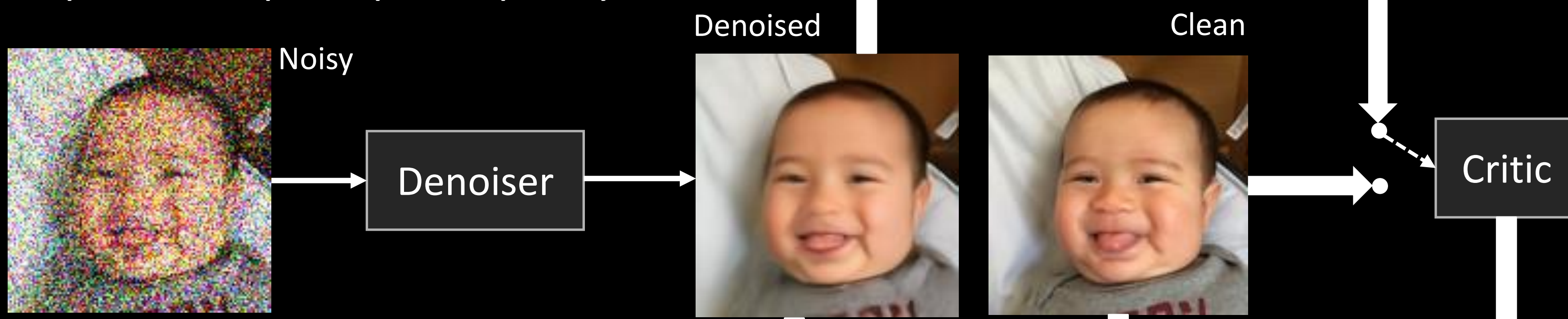
Stochastic Image Denoiser:



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



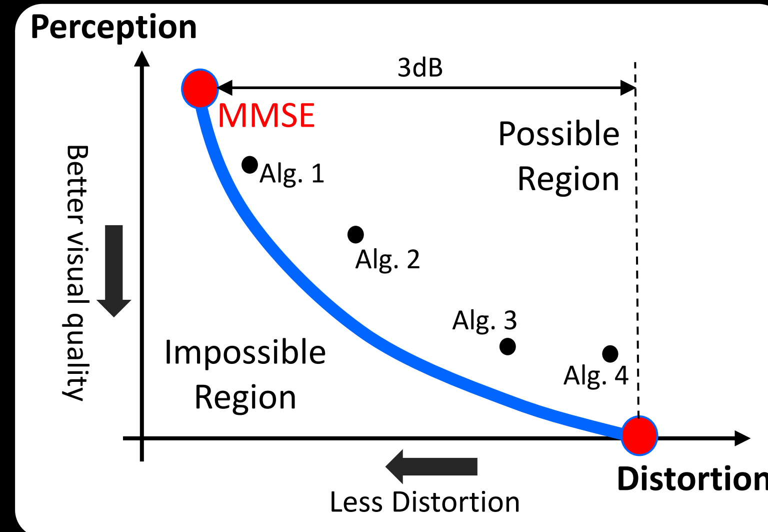
$$\mathrm{MSE}(\;,\;) + \lambda\, \mathrm{L}_{\mathrm{adv}}(\;)$$

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

Michael Elad
The Computer-Science Department
The Technion

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 [Blau & Michaeli 2017]
- We propose to sample from the posterior via a Conditional GAN mechanism (PSCGAN)

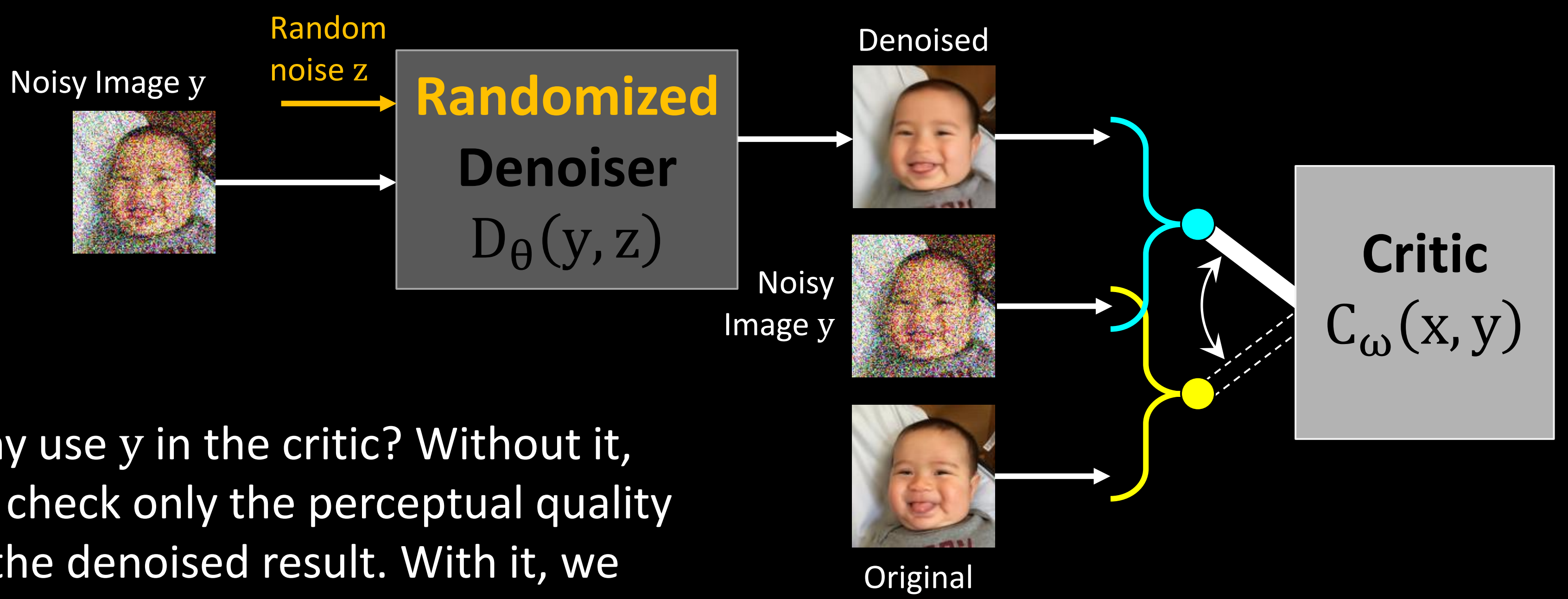


$x \sim P_X$ $\quad$ $y \sim P_{Y|X=x}$ $\quad$ $\underbrace{\qquad\qquad}_{}$ Samples from $P_{X|Y=y}$

Michael Elad
The Computer-Science Department
The Technion

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 (many x's for each y and many y's for each x)
- On the other hand, we would like to avoid a penalty of the form

$$\mathbb{E}_{X,Y,Z}\left[\|x - D_{\theta}(y,z)\|_2^2\right]$$

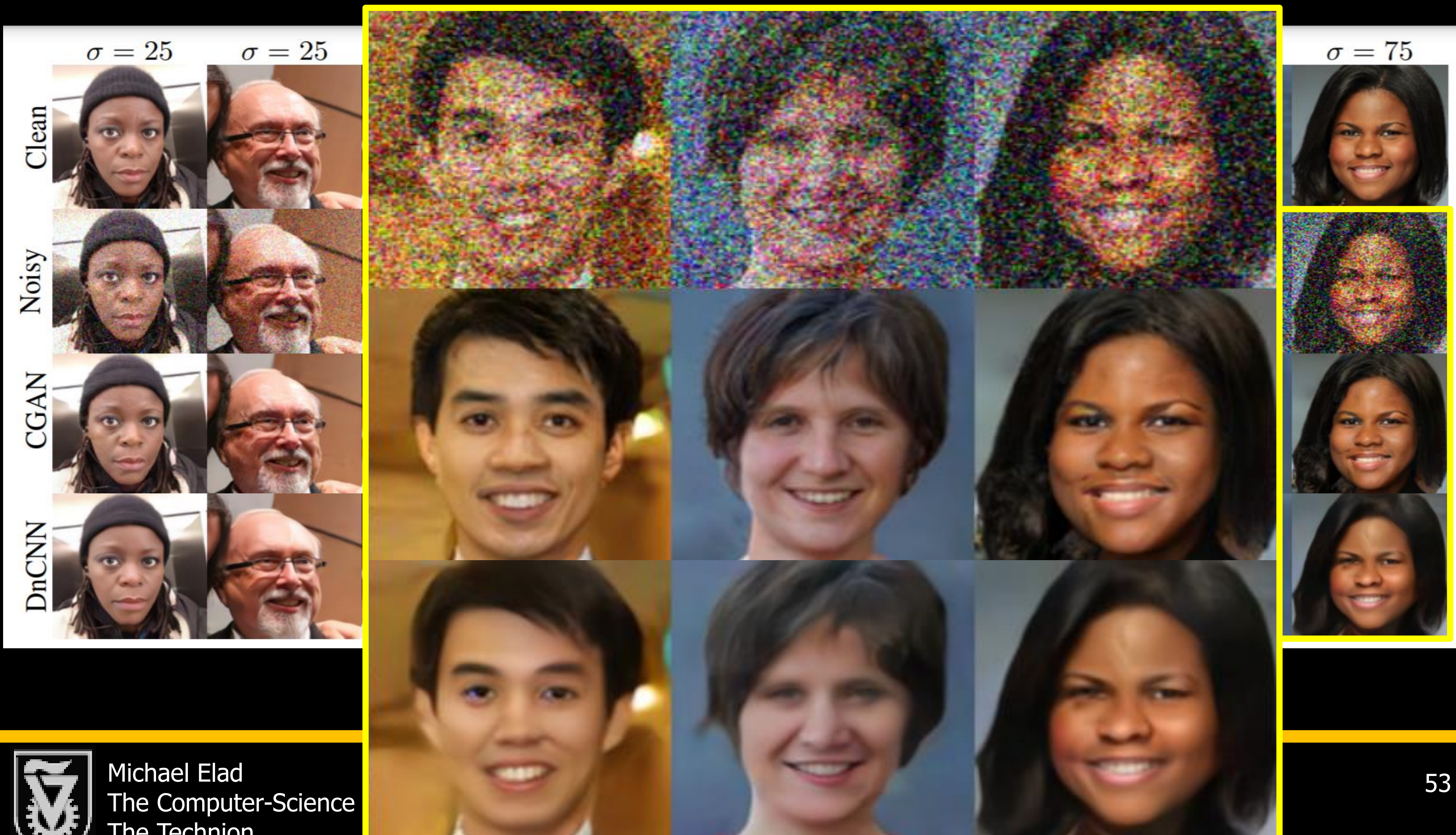
- Our remedy: adding an MMSE oriented penalty term:

$$\mathbb{E}_{X,Y}\left[\|x - \mathbb{E}_Z[D_{\theta}|y]\|_2^2\right]$$

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

Discovery 3: Targeting Perceptual Quality

CGAN:



Michael Elad
The Computer-Science
The Technion

Oh … and One Last Thing

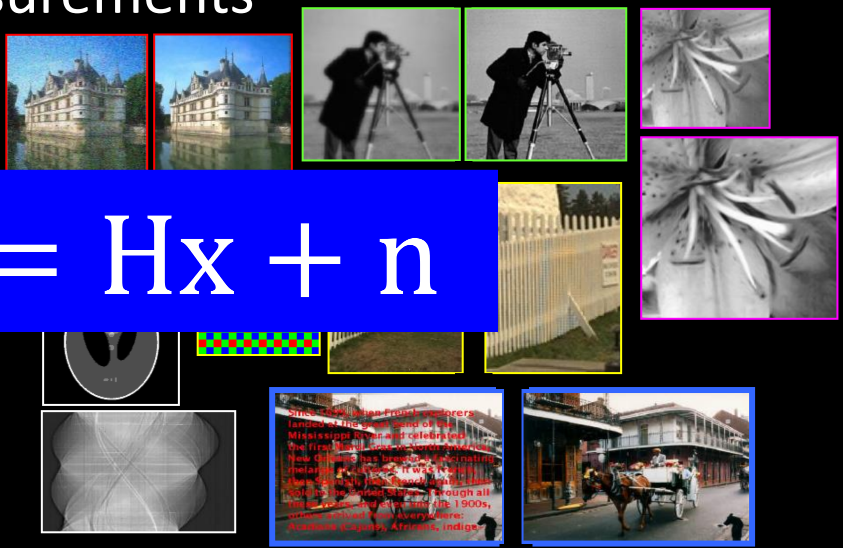
Back to Inverse Problems

- 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, in order to obtain “perfect looking” results?

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

Snips: Solving noisy inverse problems stochastically B Kawar, G Vaksman, M Elad Advances in Neural Information Processing Systems 34	77	2021
Denoising Diffusion Restoration Models B Kawar, M Elad, S Ermon, J Song Advances in Neural Information Processing Systems (NeurIPS) 2022	215	2022

Back to Inverse Problems

A&A 672, A51 (2023)
https://doi.org/10.1051/0004-6361/202243054

Astronomy & Astrophysics

Probabilistic mass-mapping with neural score estimation★

B. Remy[1], F. Lanusse[1], N. Jeffrey[2,3], J. Liu[4,5,6], J.-L. Starck[1], K. Osato[7,2], and T. Schrabback[8]

Context. Weak lensing mass-mapping is a useful tool for accessing the full distribution of dark matter on the sky, but because of intrinsic galaxy ellipticies, finite fields, and missing data, the recovery of dark matter maps constitutes a challenging, ill-posed inverse problem

Aims. We introduce a novel methodology that enables the efficient sampling of the high-dimensional Bayesian posterior of the weak lensing mass-mapping problem, relying on simulations to define a fully non-Gaussian prior. We aim to demonstrate the accuracy of the method to simulated fields, and then proceed to apply it to the mass reconstruction of the HST/ACS COSMOS field.

Methods. The proposed methodology combines elements of Bayesian statistics, analytic theory, and a recent class of deep generative models based on neural score matching. This approach allows us to make full use of analytic cosmological theory to constrain the 2pt statistics of the solution, to understand any differences between this analytic prior and full simulations from cosmological simulations, and to obtain samples from the full Bayesian posterior of the problem for robust uncertainty quantification.

Results. We demonstrate the method in the κTNG simulations and find that the posterior mean significantly outperfoms previous methods (Kaiser–Squires, Wiener filter, Sparsity priors) both for the root-mean-square error and in terms of the Pearson correlation. We further illustrate the interpretability of the recovered posterior by establishing a close correlation between posterior convergence values and the S/N of the clusters artificially introduced into a field. Finally, we apply the method to the reconstruction of the HST/ACS COSMOS field, which yields the highest-quality convergence map of this field to date.

Conclusions. We find the proposed approach to be superior to previous algorithms, scalable, providing uncertainties, and using a fully non-Gaussian prior.

Results. We demonstrate the method in the κTNG simulations and find that the posterior mean significantly outperfoms previous methods (Kaiser–Squires, Wiener filter, Sparsity priors) both for the root-mean-square error and in terms of the Pearson correlation. We further illustrate the interpretability of the recovered posterior by establishing a close correlation between posterior convergence values and the S/N of the clusters artificially introduced into a field. Finally, we apply the method to the reconstruction of the HST/ACS COSMOS field, which yields the highest-quality convergence map of this field to date.
Conclusions. We find the proposed approach to be superior to previous algorithms, scalable, providing uncertainties, and using a fully non-Gaussian prior.

Key words. cosmology: observations – methods: statistical – gravitational lensing: weak

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 $\mathrm{x_i} = \mathrm{x} + \mathrm{v_i}$ 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 $H = U\Sigma V^T$:

$$U^T y = U^T \underbrace{[U\Sigma V^T(x_i - v_i) + n]}_{y = Hx + n} = \Sigma V^T(x_i - v_i) + U^T n$$

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

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

- Thus, we can apply the Langevin dynamics algorithm on $\tilde{x}_T = V^T x_i$ given $y_T = U^T y$ 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

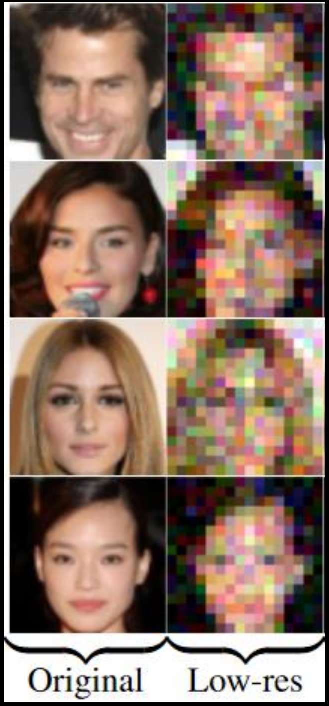
Michael Elad
The Computer-Science Department
The Technion

Back to Inverse Problems

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

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 $\sigma_0 \approx 12$

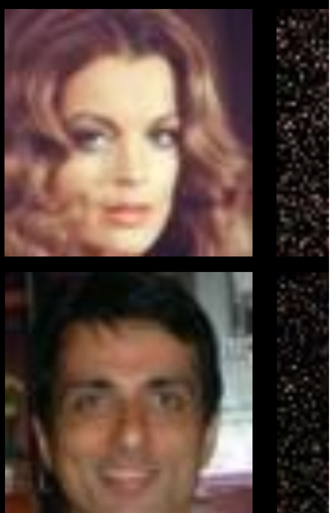
Original Degraded

Back to Inverse Problems

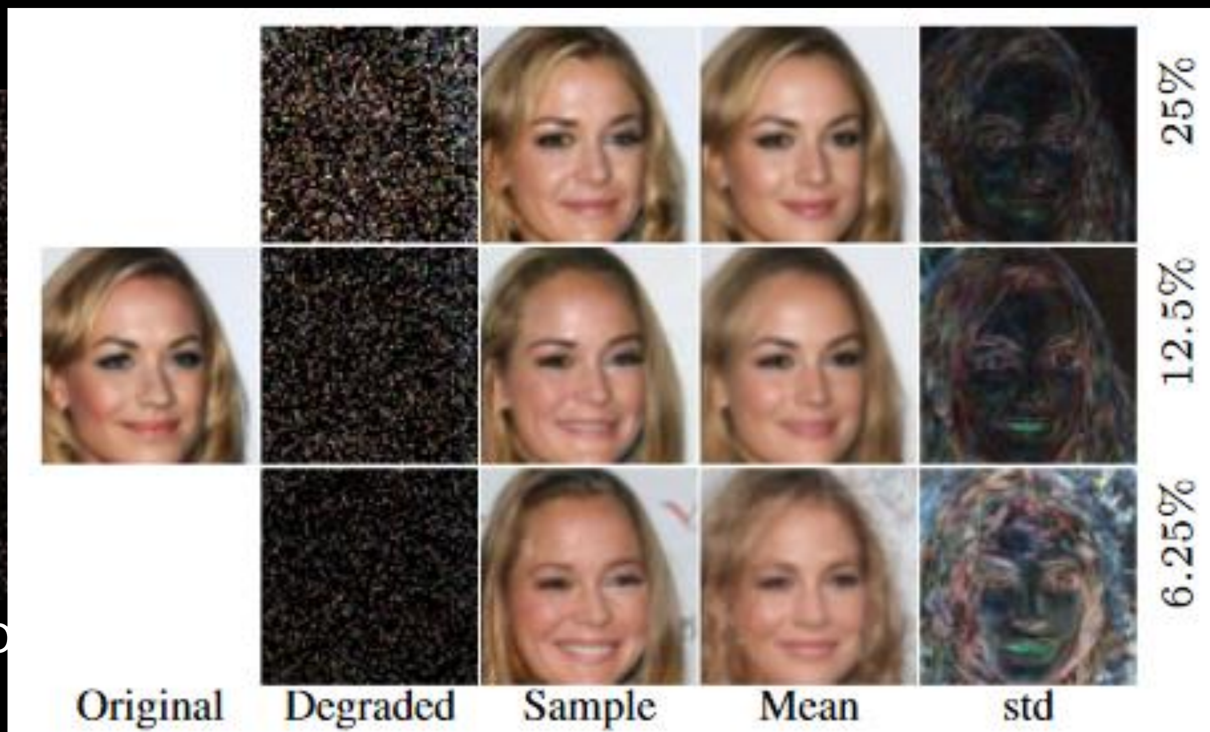
Deblurring: uniform 5×5 blur with additive noise of $\sigma_0 \approx 25$

Back to Inverse Problems

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



Original D



Time to Summarize

What Have we Seen Today?

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

WE CAN USE $\mathrm{D}(y, \sigma)$ 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 $\mathrm{D}(y, \sigma)$ iteratively

Michael Elad
The Computer-Science Department
The Technion

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:

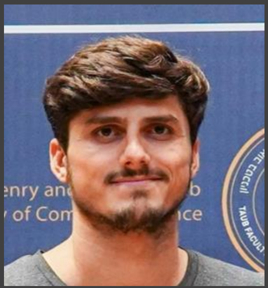

Peyman Milanfar


Yaniv Romano


Grisha Vaksman


Bahjat Kawar


Meyer Scetbon


Guy Ohayon


Theo Adrai