A Closer Look at Diffusion Models

Michael Elad

CS - Technion & NVIDIA

May 12th, 2025

2014 Was a Wonderful Year ... for generative AI

- VAE: Kingma & Welling introduced Variational Auto-Encoders
- GAN: Goodfellow, ... Courville & Bengio, presented Generative Adversarial Networks
- NF: Dinh, Krueger & Bengio brought Normalizing Flow
- RNN: Alex Graves presented Recurrent Neural Networks (a.k.a. Auto-Regressive models)
- EBM: Rezende et Al. harnessed successfully energy-machines for challenging tasks
- Diffusion: Sohl-Dickstein et Al. offered the very first version of Diffusion Models

Common to all is the desire to learn (through many examples) a synthesis machine

$$\hat{x} = G_\theta(z)\ ,\ \ z \sim \mathbb{N}(0, \mathbf{I})$$

that can sample fairly from complex distributions $P(\underline{x})$

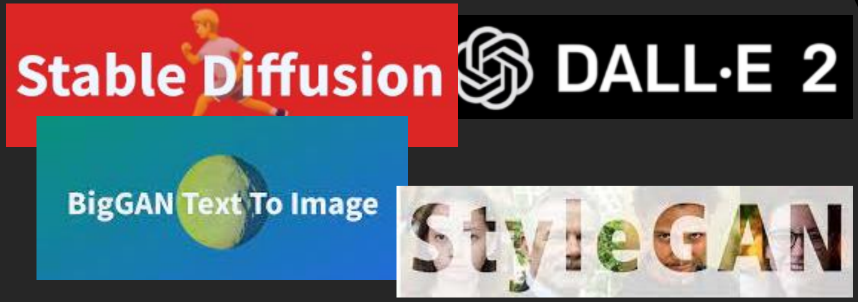
Generative AI

Wikipedia: Generative artificial intelligence (generative AI) is artificial intelligence capable of generating text, images, or other media, using generative models

Gen-AI has two main and separate branches:

Text/Code Synthesis

Image/Video/Audio Synthesis



At the lead: Autoregressive Models

A very recent trend: Using Diffusion Models for LLM's

At the lead: Diffusion Models

Michael Elad
The Computer-Science Department
The Technion

This Talk is all About

Our recent work on …

Diffusion Models

Agenda:

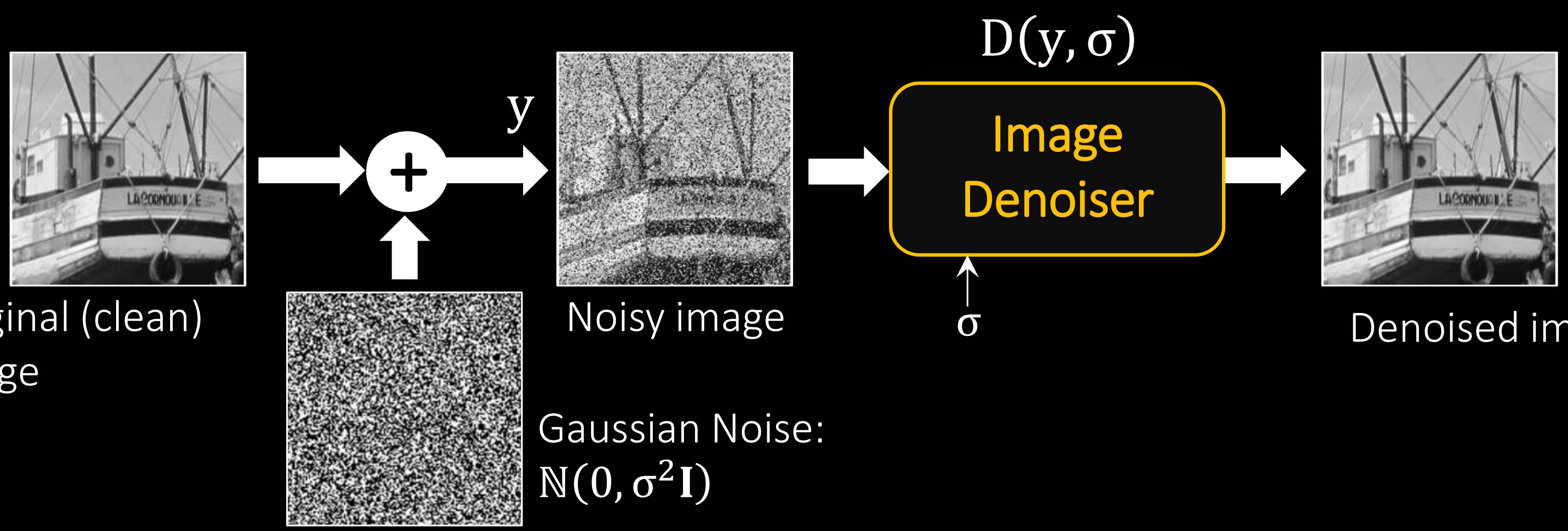
1. Introduction to Diffusion Models
2. Diffusion Models as Posterior Samplers
3. Posterior-Sampling based Compression (PSC)
4. Conclusion

1. Introduction to Diffusion Models
2. Diffusion Models as Posterior Samplers
3. Posterior-Sampling based Coding (PSC)
4. Conclusion

At the Center of Our Story …

Image Denoisers

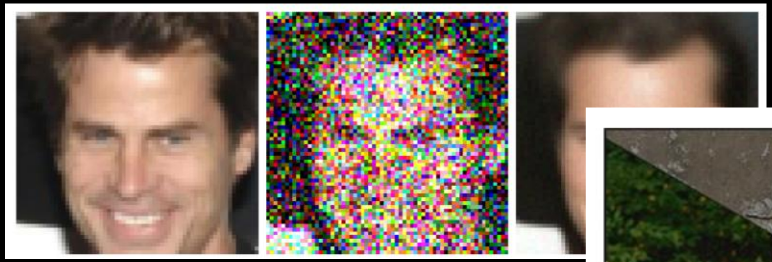
For cleaning White Additive Gaussian Noise from an Image



Michael Elad
The Computer-Science Department
The Technion

At the Center of Our Story …

Original Noisy Denoised



Original Noisy Denoised

MMSE Image Denoising: a Solved Problem !!

Michael Elad
The Computer-Science Department
The Technion

Image Synthesis via Denoisers

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	4155	2019
Improved techniques for training score-based generative models Y Song, S Ermon Advances in neural information processing systems 33, 12438-12448	1203	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	150	2021

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

Langevin Dynamics

Here is the core idea in a nutshell:

- Our goal: draw a sample from the distribution of images $P(x)$
  - Start with a random noise image x_0
  - Climb to a more probable image by the iterative equation:

$$x_{k+1} = x_k + a \cdot \underbrace{\boxed{\nabla_{x_k} \log P(x_k)}} + b \cdot z_k$$ (Langevin Dynamics [1908])

This is known [Miyasawa `61] as the Score Function and it is approximately proportional to $[\hat{x}_k - D(\hat{x}_k, \sigma)]$ for a small value of σ

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

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

Annealed Langevin Dynamics

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

Text-2-Image via Diffusion Models

Here is a taste from Google's Imagen, in the context of Text-2-Image:

A cute sloth holding a small treasure chest. A bright golden glow is coming from the chest	A cute corgi lives in a house made out of sushi	Teddy bears in a swimming pool

Michael Elad
The Computer-Science Department
The Technion

1. Introduction to Diffusion Models
2. Diffusion Models as Posterior Samplers
3. Posterior-Sampling based Coding (PSC)
4. Conclusion

Moving to Posterior Sampling

- Goal: Recovery from corrupted measurements

  De-Noising De-Blurring

  In-Painting De-Mosaicing

  Tomography Image Scale-Up & super-resolution

- Can we suggest a "sampler" from the Posterior $P(x|y)$ for handling these problems?

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

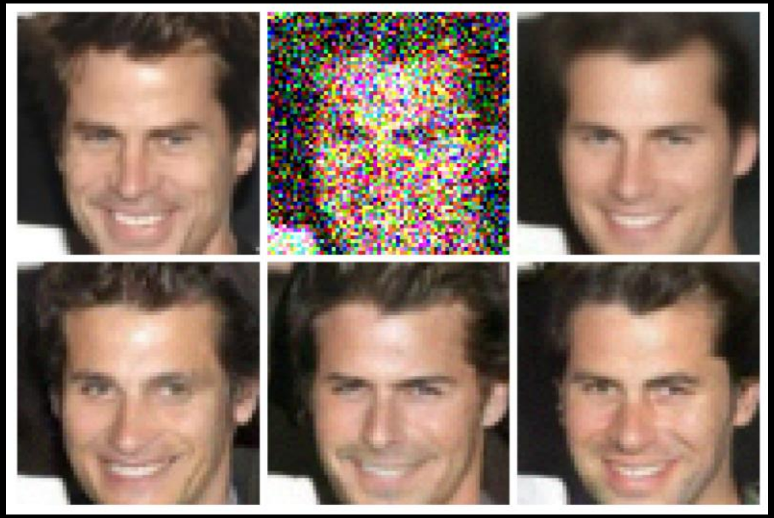
$$y = Hx + n$$

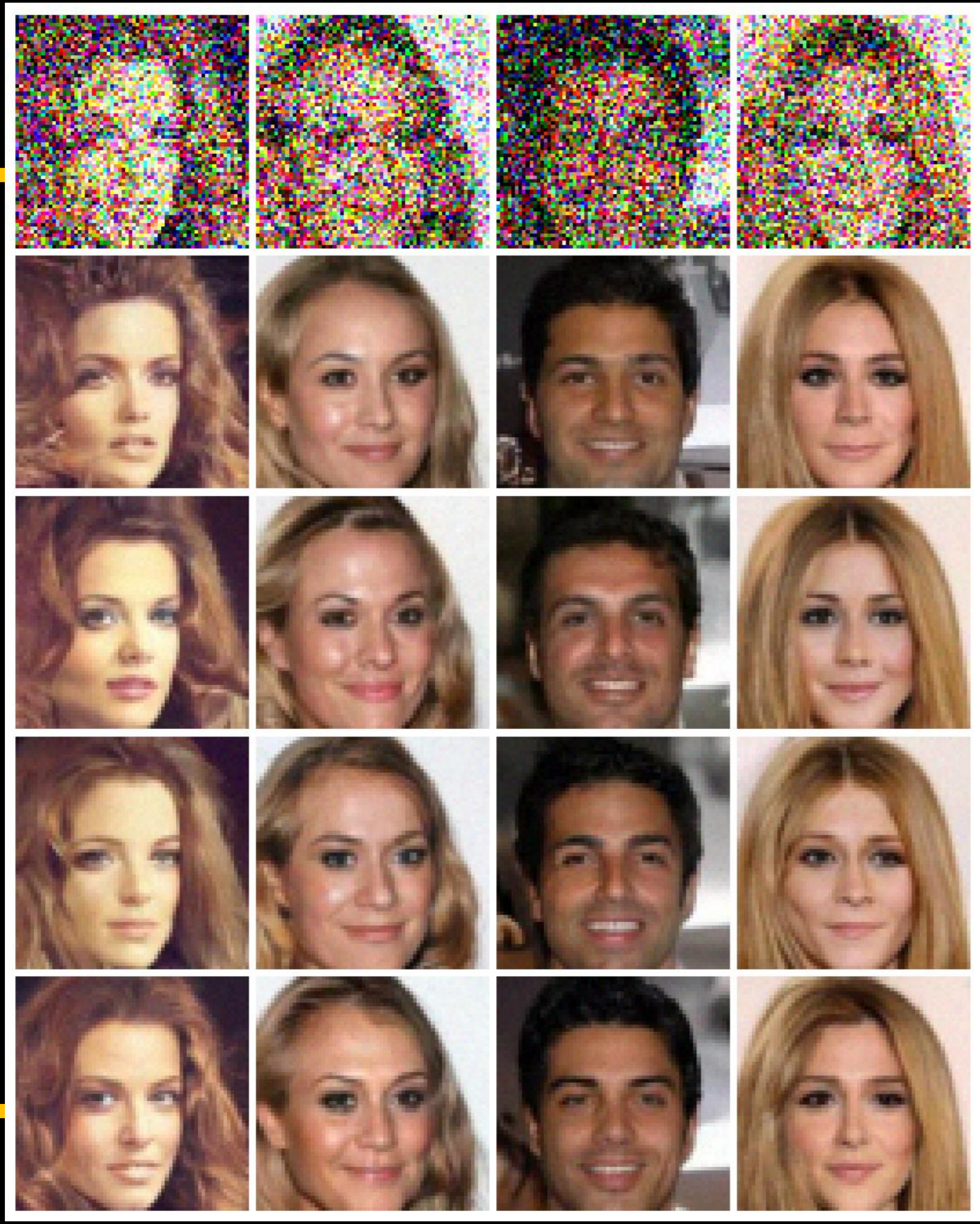
Snips: Solving noisy inverse problems stochastically | 179 | 2021
B Kawar, G Vaksman, M Elad
Advances in Neural Information Processing Systems 34

Denoising Diffusion Restoration Models | 842 | 2022
B Kawar, M Elad, S Ermon, J Song
Advances in Neural Information Processing Systems (NeurIPS)

Michael Elad
The Computer-Science Department
The Technion

Few Results: Image Denoising



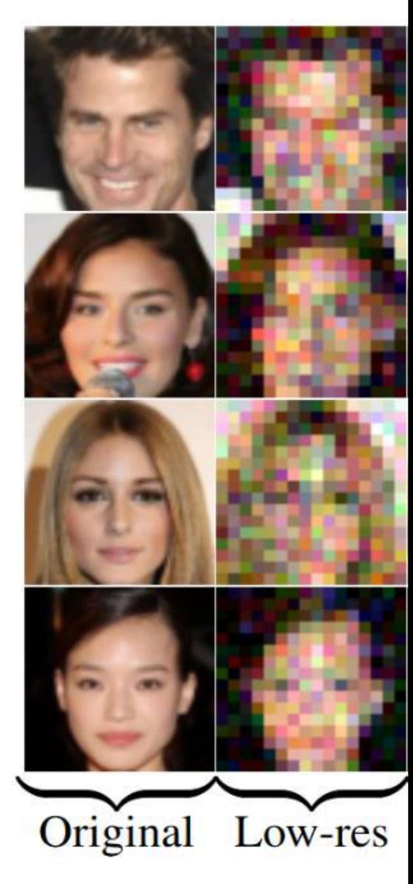


Few Results: Image Inpainting

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

Few Results: Super-Resolution

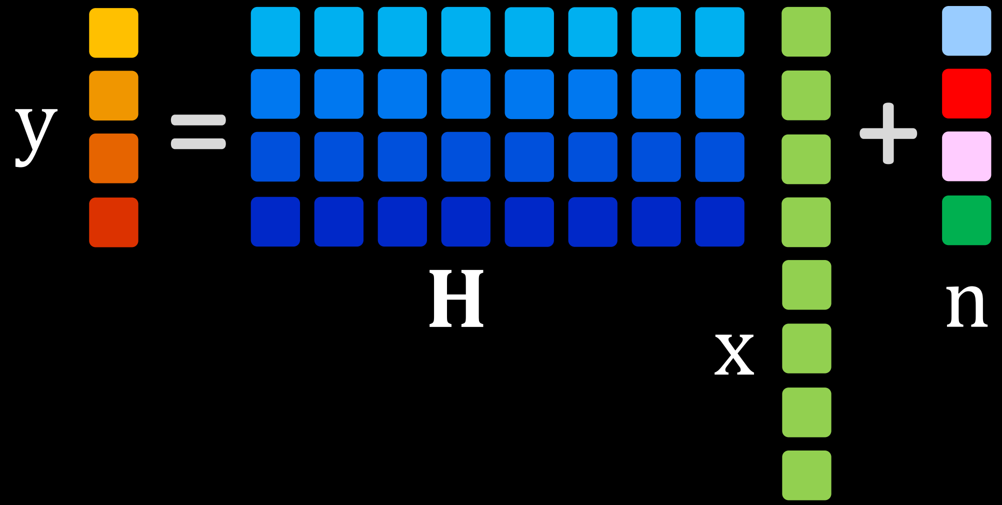
Downscaling by 4 with additive noise of $\sigma_0 \approx 25$



1. Introduction to Diffusion Models
2. Diffusion Models as Posterior Samplers
3. Posterior-Sampling based Coding (PSC)
4. Conclusion

Here is What We Know Now

From the previous discussion (and the vast work on these topics by other teams) it is now clear that



Sampling images from the Posterior $\hat{\mathrm{x}} \sim \mathrm{P}(\mathrm{x}|\mathrm{y} = \mathbf{H}\mathrm{x} + \mathrm{n})$ is within reach

Here is something quite interesting that we came up with just recently ….

Adaptive Compressed Sensing with Diffusion-Based Posterior Sampling | 4 | 2024
N Elata, T Michaeli, M Elad
European Conference on Computer Vision (ECCV)

Zero-Shot Image Compression with Diffusion-Based Posterior Sampling | 2024
N Elata, T Michaeli, M Elad
arXiv preprint arXiv:2407.09896


Noam Elata


Tomer Michaeli

Compression Result: a Teaser

Original

Compress

$\times 128$

JPEG

Michael Elad
The Computer-Science Department
The Technion

Compression Result: a Teaser

Original

Compress

$\times 128$

PSC (Ours)

Michael Elad
The Computer-Science Department
The Technion

Recall: Transform Coding

$\mathbf{x}$

Quantization

Encoder $\mathbf{Q\{Hx\}}$

Transform

Entropy Coding

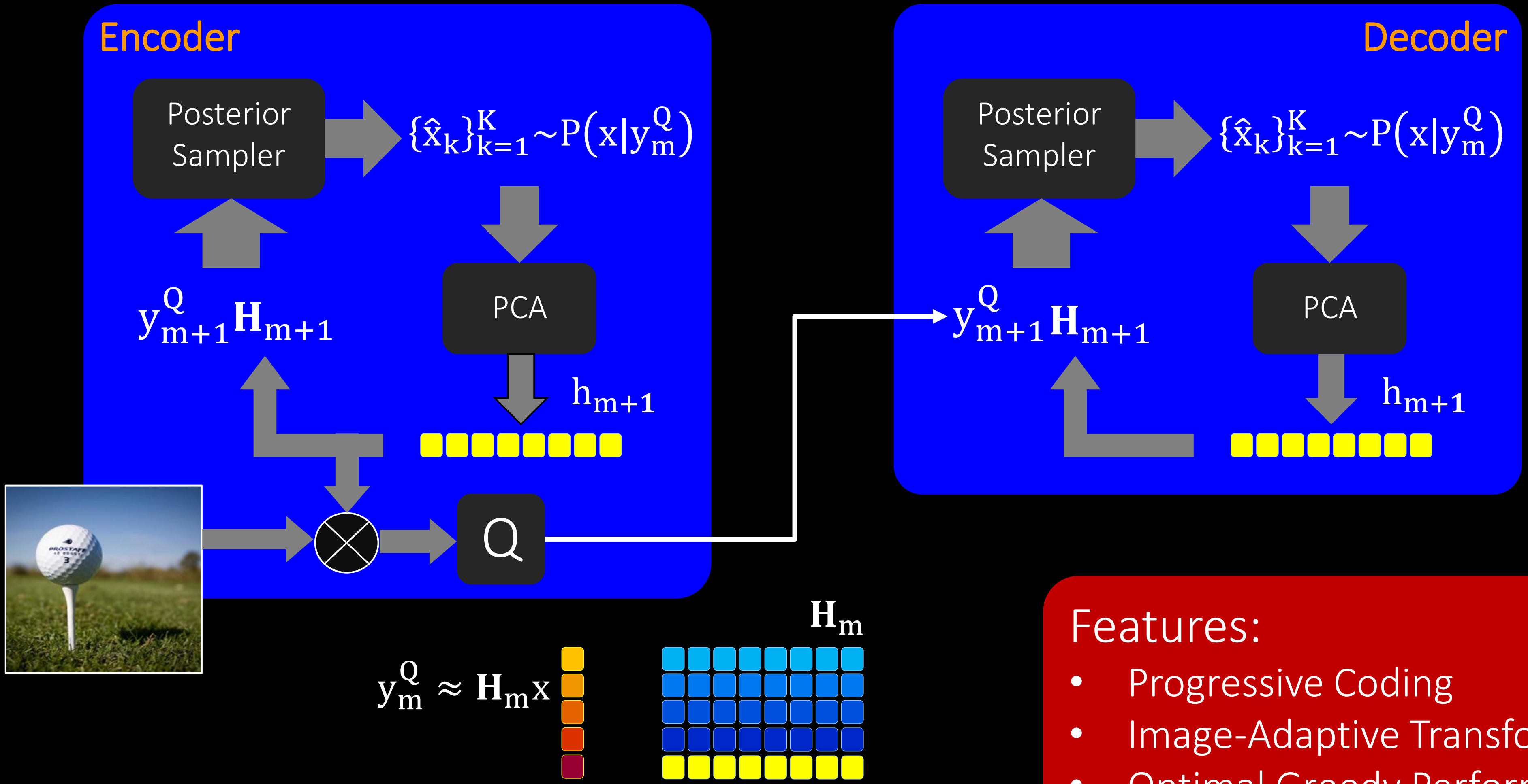
Decoder

$\hat{\mathbf{x}}$

Could we construct an effective coding algorithm using an IMAGE-ADAPTIVE transform? → NO! → Image-dependent transform should be transmitted too, and this will ruin the coding performance → So … → Common coding schemes rely on a fixed transform → … → Is there a way around it ?

Michael Elad
The Computer-Science Department
The Technion

A Novel Compression Scheme: PSC

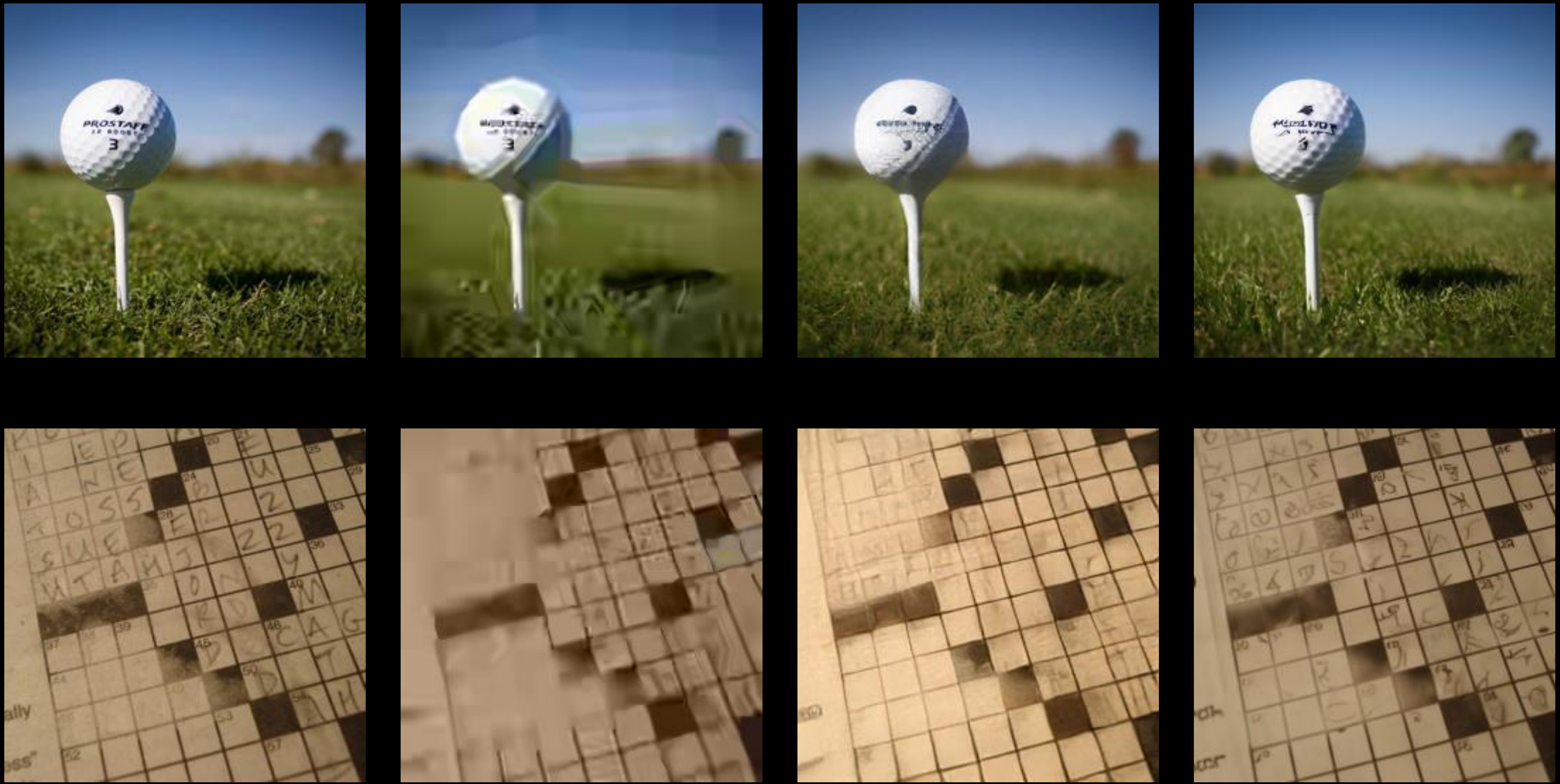


Features:

- Progressive Coding
- Image-Adaptive Transform
- Optimal Greedy Performance
- Zero-Shot (no Training)

Does it Work?

Original | BPG | HiFiC | PSC (Ours)



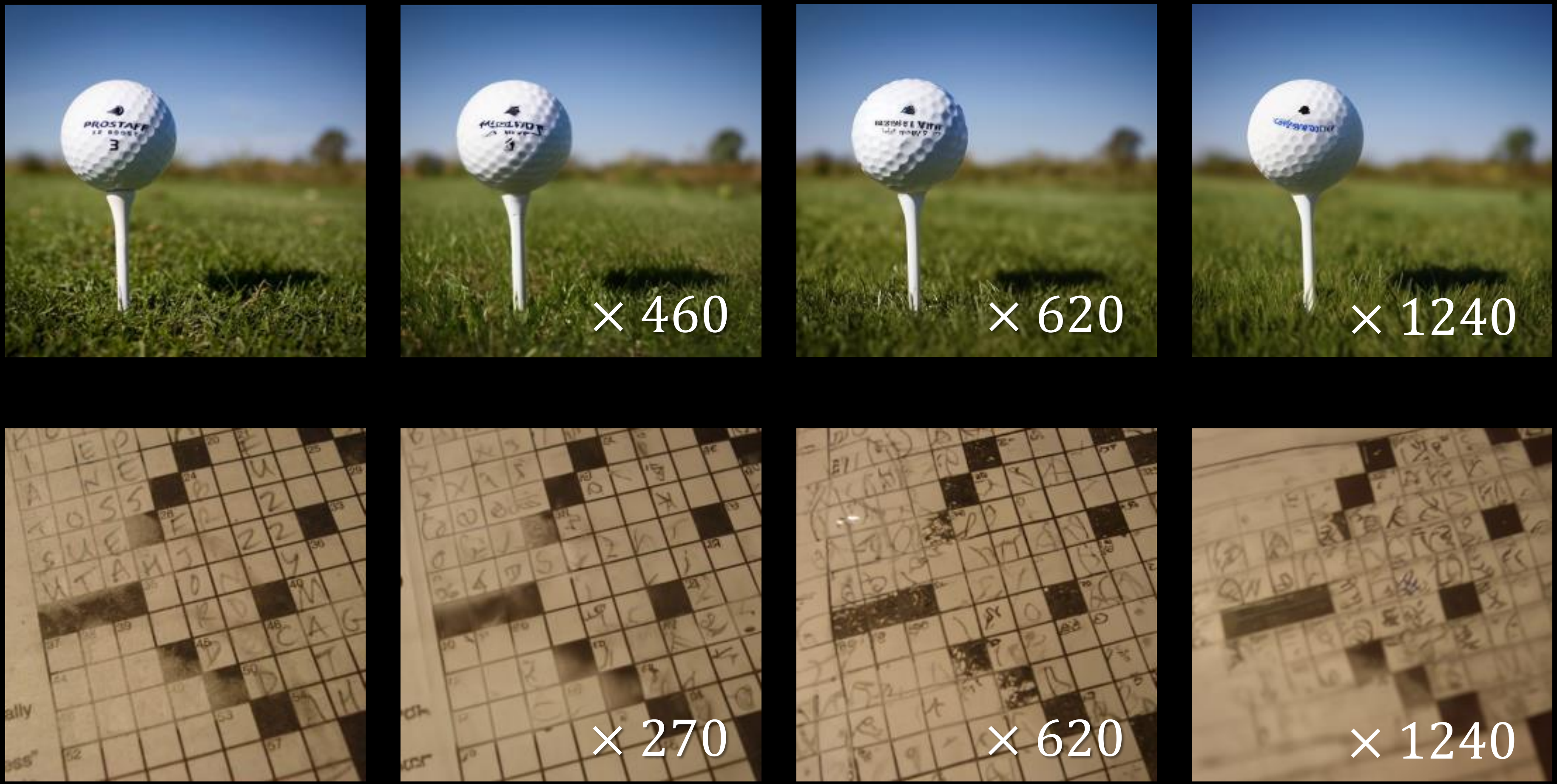
$\times 460$

$\times 270$

Michael Elad
The Computer-Science Department
The Technion

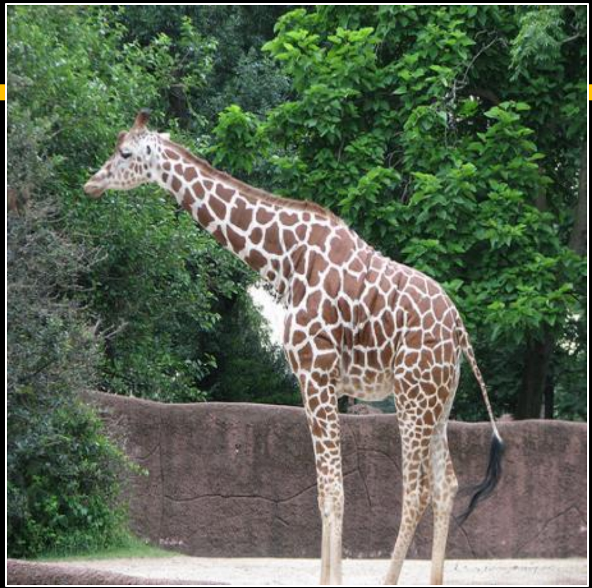
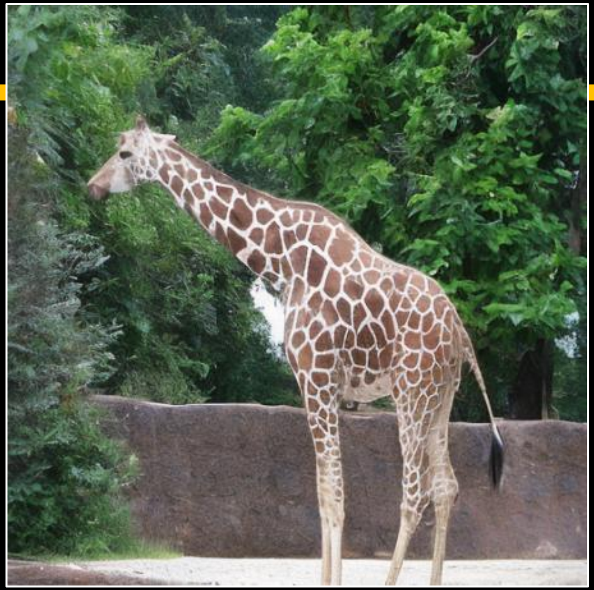
Does it Work?

Original | PSC (Ours) | PSC (Ours) | PSC (Ours)



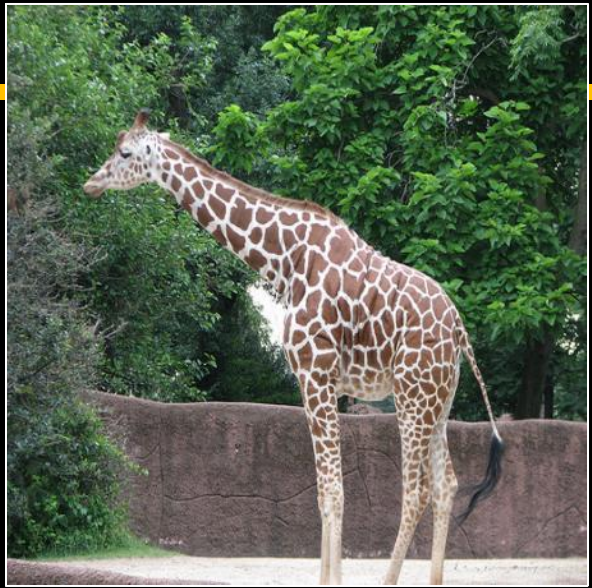
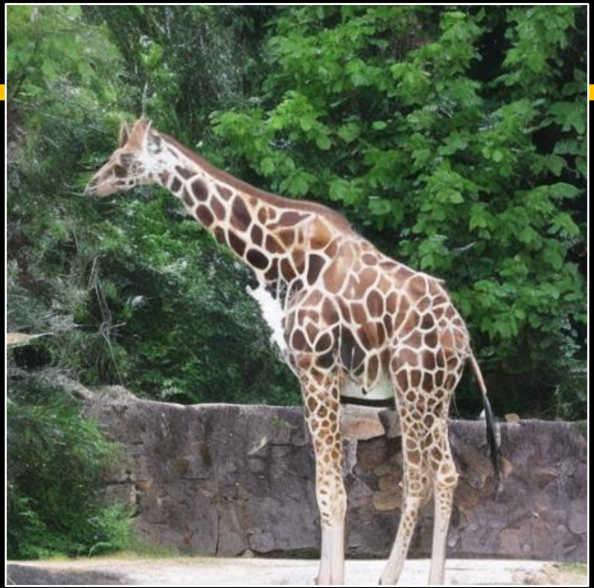
Michael Elad
The Computer-Science Department
The Technion

Does it Work?

	Original	LPSC (Ours)	
A giraffe grazing from a tree with rock wall in background			$\times 156$
A big bowl of different kinds of fruit inside			$\times 156$

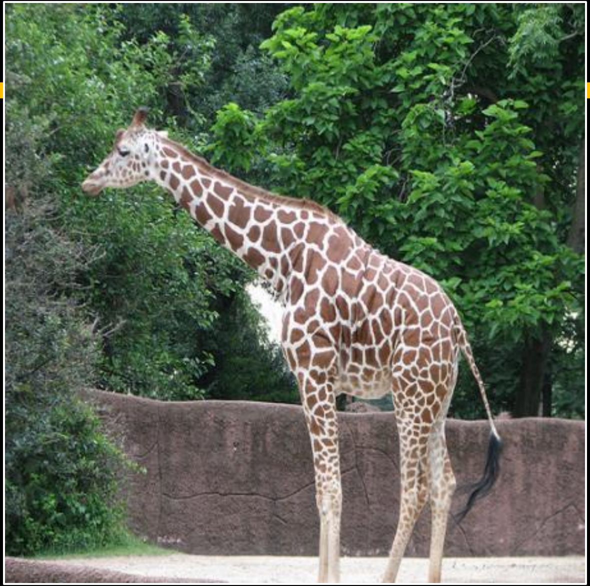
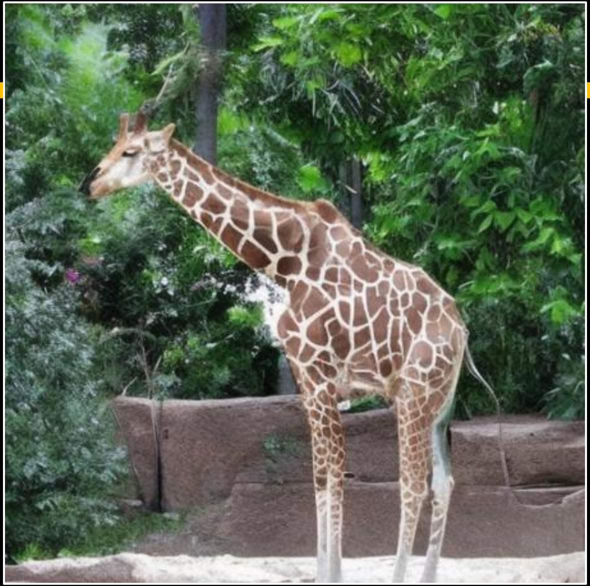
Michael Elad
The Computer-Science Department
The Technion

Does it Work?

	Original	LPSC (Ours)	
A giraffe grazing from a tree with rock wall in background			$\times 312$
A big bowl of different kinds of fruit inside			$\times 312$

Michael Elad
The Computer-Science Department
The Technion

Does it Work?

	Original	LPSC (Ours)	
A giraffe grazing from a tree with rock wall in background			$\times 624$
A big bowl of different kinds of fruit inside			$\times 624$

Michael Elad
The Computer-Science Department
The Technion

1. Introduction to Diffusion Models
2. Diffusion Models as Posterior Samplers
3. Posterior-Sampling based Coding (PSC)
4. Conclusion

Summary

Diffusion Models (DM) have taken the lead in handling various Generative AI tasks

DM can be "easily" adapted to become reliable posterior samplers for linear sensing

As such, posterior samplers can be harnessed to address adaptive CS – AdaSense is such a method

AdaSense is shown to be the foundation for PSC: a novel and competitive lossy compression scheme

More broadly, Diffusion Models expose new opportunities for revisiting the topic of Lossy compression schemes

Thank You

David Malah

… and Team (Past and Present):

Yoram Or-Chen, Nimrod Peleg, Ziva Avni, Avi Rozen, Yair Moshe, Ori Bryt …

Image Synthesis

- In the past decade, the AI revolution brought a growing interest in synthesizing images "out of thin air"
- Popular tools: VAE, GAN, NF, AR, EBM, Diffusion
- The essence of this synthesis task: Sampling from $P(x)$
- Why synthesize? Because
  - We can, and it is fascinating
  - This can be leveraged for practical needs (compression, restoration)
- Key question: Could we sample images from $P(x)$ by using an image denoiser?

thispersondoesnotexist.com

Surely, You Have Heard of …

Stable Diffusion

stability.ai

Google

Imagen

unprecedented photorealism × deep level of language understanding

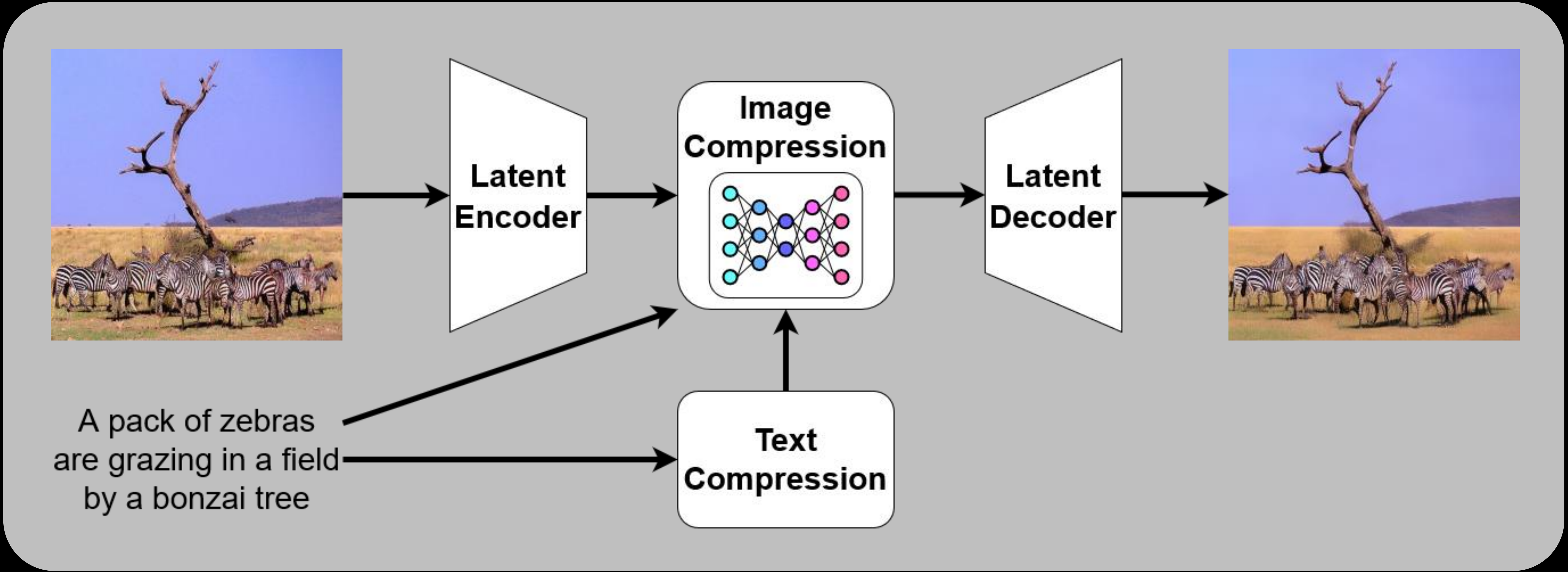
WAKEUP THE ARTIST IN YOU

Midjourney

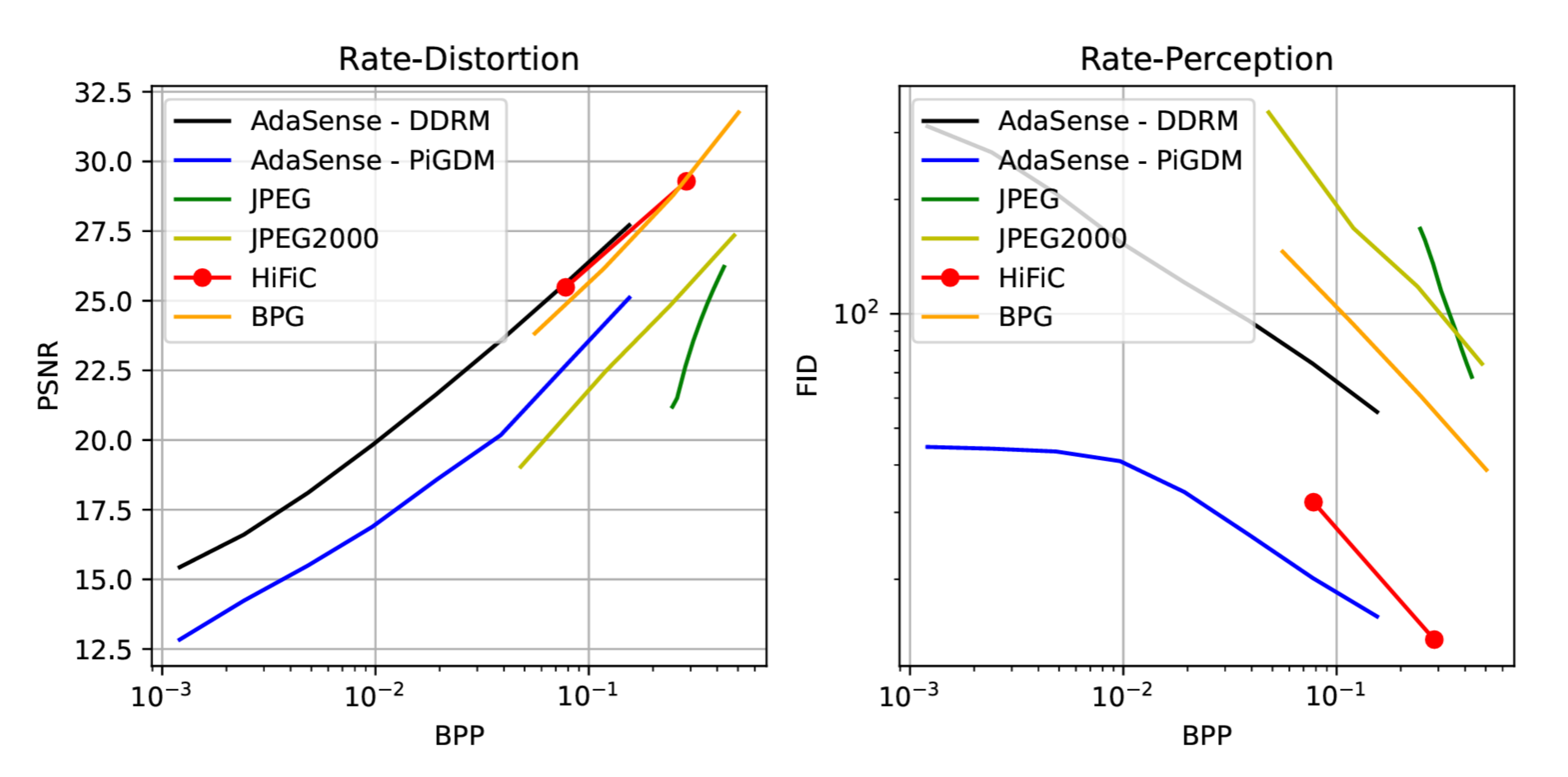
ARTIFICIAL INTELLIGENCE ART GENERATOR

Comment: OpenAI have just released an image generator within GPT-4o, relying on an Auto-Regressive model! However, if you check carefully, it uses a diffusion model to finalize the created images

Latent-Based PSC with Text-2-Image



Does it Work?



PSC has two options for the output:

- A sample – targeting perceptual quality
- An average of all samples – targeting distortion

BPG (2014): Better Portable Graphics, a high-end replacement to JPEG based on HEVC

HiFiC (2020): Google-Research based deep learning compression