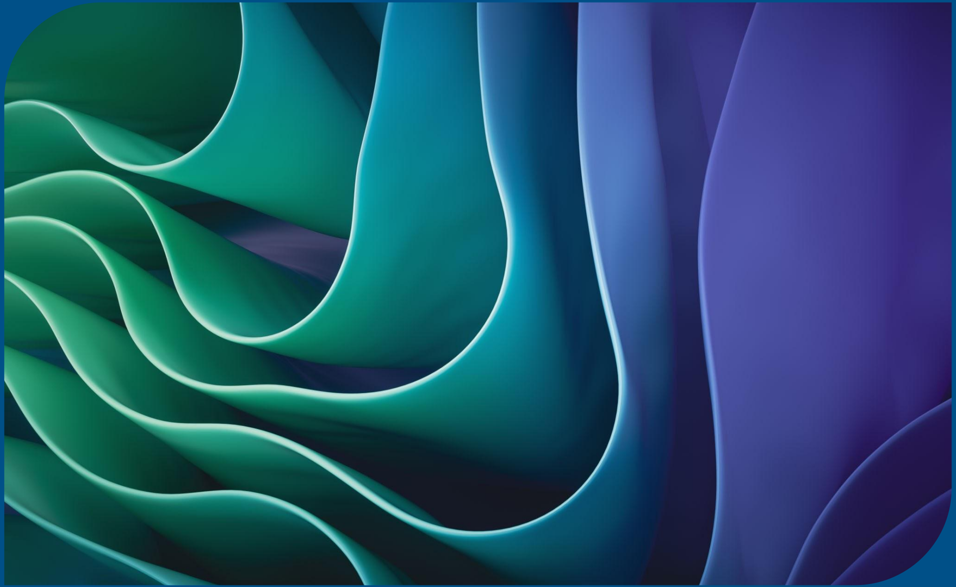
Jan'30, 2026

From Pixels to Cells: Advancing Diffusion Models for Biomedical Data

Jakub M. Tomczak, PhD



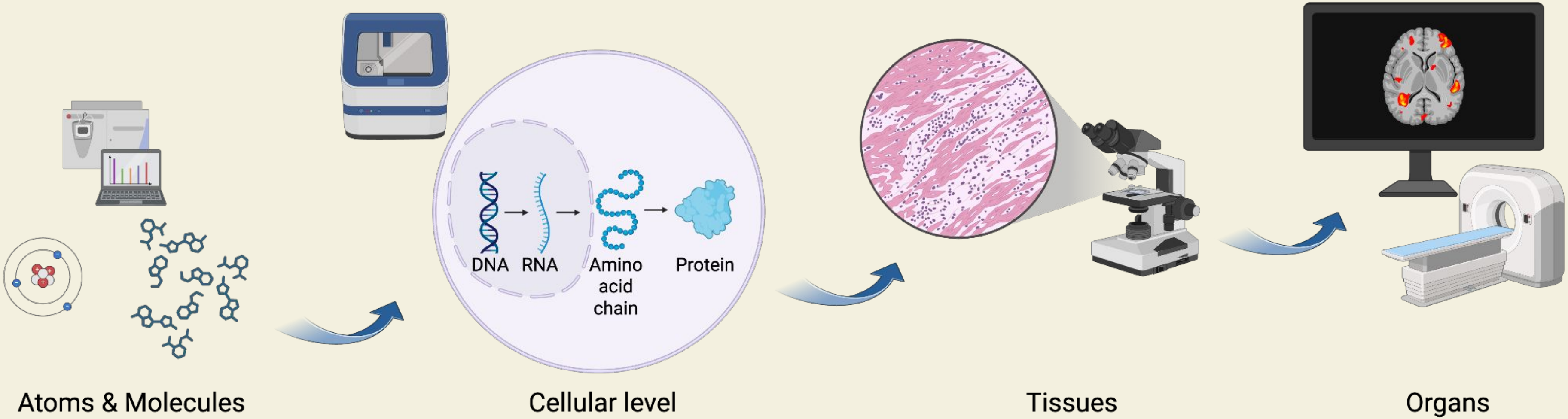
The Inaugural AI in Science Symposium 2026
UC Irvine AI in Science Institute

1. **The complexity of Biomedical Data**
2. **Diffusion models lead the *new* AI**
3. **Understanding diffusion models**
4. **Latent diffusion models are the way to go?**
5. **Conclusion**

The complexity of Biomedical Data

Biomedical data span molecules to organs, combining extreme scale, heterogeneity, and structure, motivating advanced AI models for sequences, graphs, images, and temporal data.

Biomedical data – atoms, molecules, cells, tissues, organs (& clinical)



Up to 10^{60} molecules

~3 x 10^{13} cells in a human body
~6 x 10^{9} nucleotides/seq
>60k genes

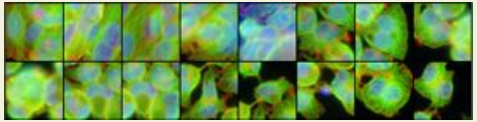
~ 10^{10} pixels per a histopathology slide

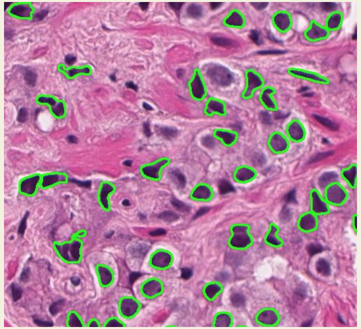
~ 10^{6} pixels per a single slide

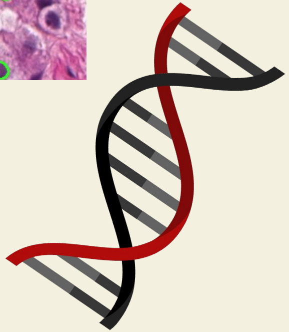
Biomedical data – atoms, molecules, cells, tissues, organs (& clinical)

As a result, we work with:

- 3D structures (e.g., molecules, proteins)
- Graph-based structured (e.g., molecules)
- Very long sequences (e.g., DNA sequences, amino acid chains)
- **Sparse matrices** (e.g., gene expression data)
- **Highly structured images** (e.g., images of cells, tissues)
- Temporal data



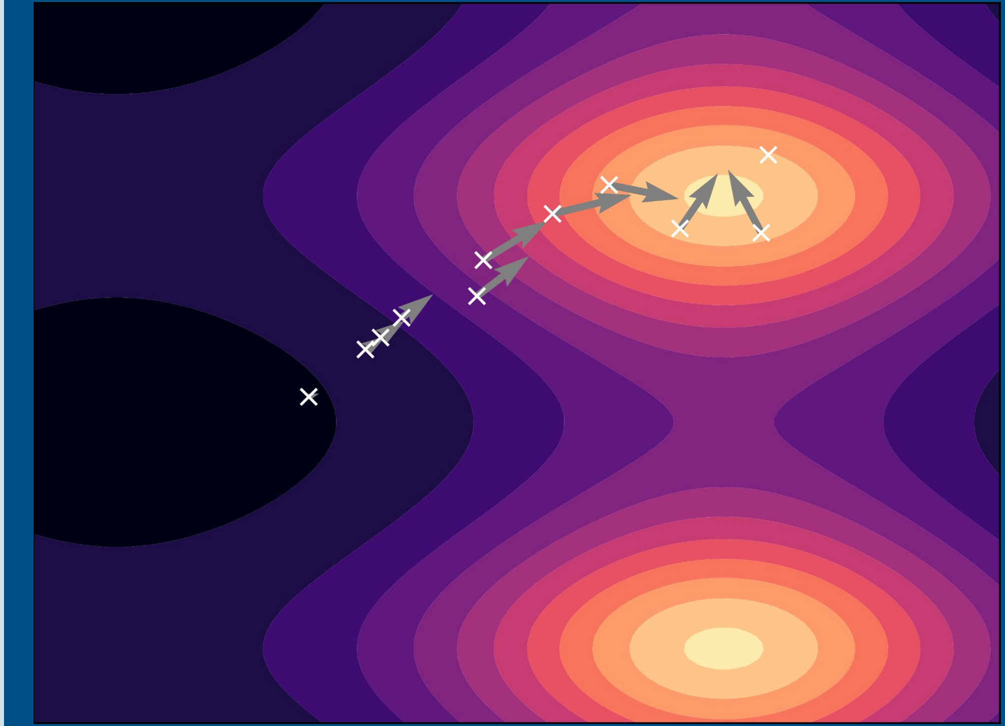




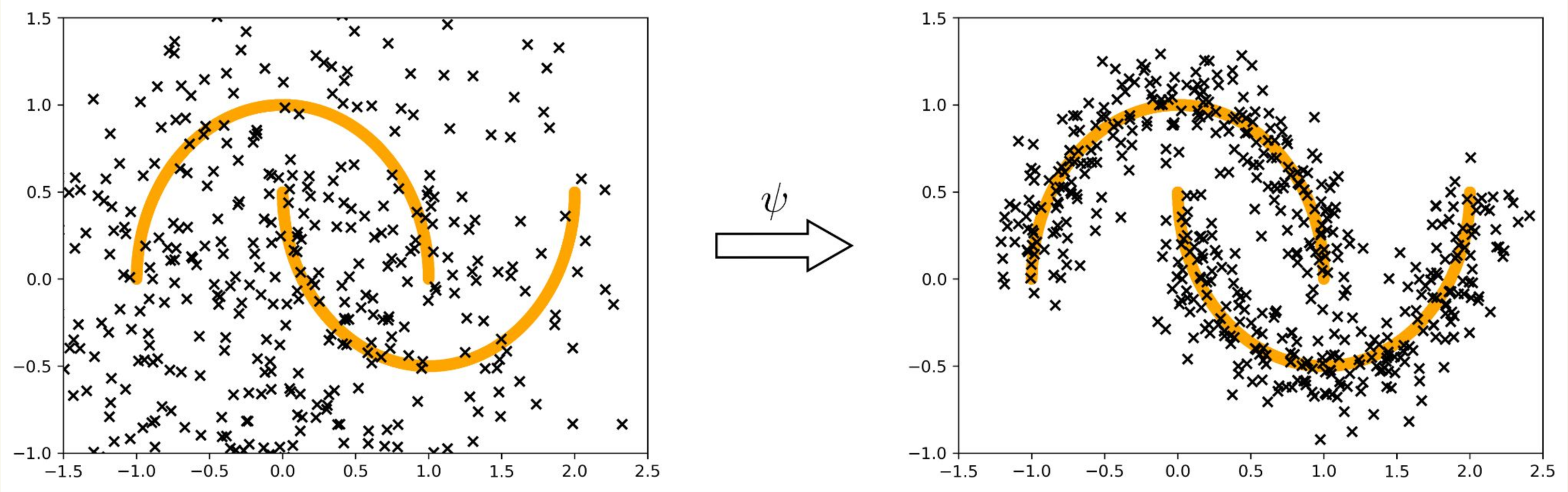
Science and medical data are **fascinating** for applications of AI!

Diffusion models lead the *new* AI

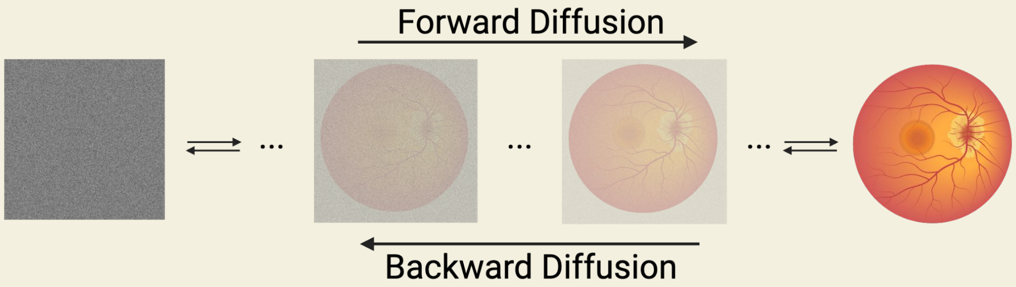
Diffusion models redefine generative modeling via noise-driven learning, enabling rich research directions and applications from image generation and NLP to life sciences.

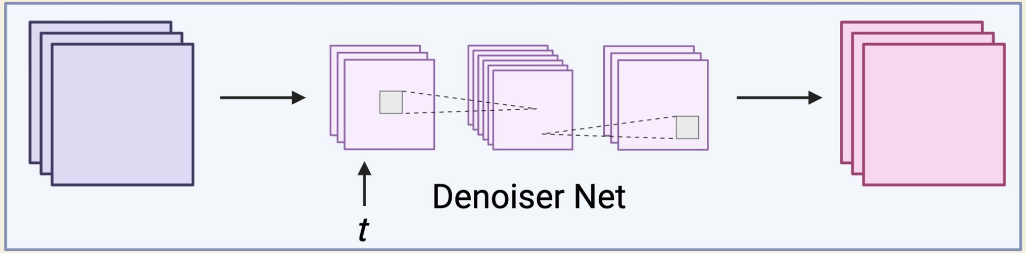


“Creating noise from data is easy; creating data from noise is generative modeling” (Song et al., 2020)



Diffusion-based models





Forward diffusion:

- adds Gaussian noise

Backward diffusion:

- removes noise

Denoiser Net:

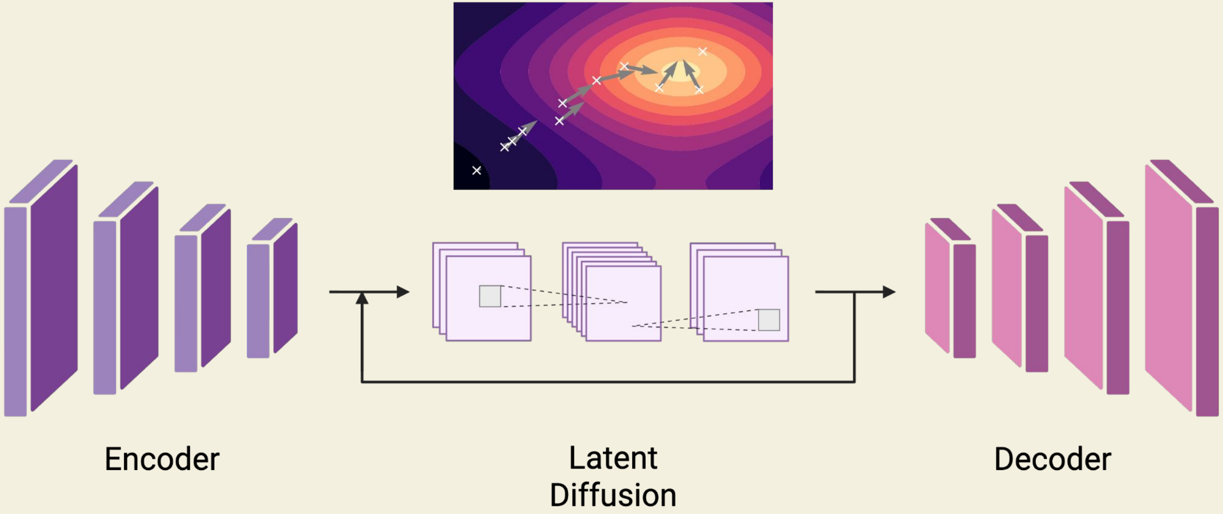
- learns how to denoise
- keeps the dimensionality

Connections to VAEs:

- infinitely many latents
- variational posterior is forward diffusion

Training objective: $\mathcal{L}(\mathbf{x}_0) = \sum_t \lambda_t \underbrace{\|\epsilon_t - \epsilon_{NN}(\mathbf{x}_t;t)\|^2}_{L_t}$

Latent Diffusion Models (LDMs)



Auto-encoder:

- compresses objects
- ideally: <u>no</u> distortion

Diffusion:

- in the latent space

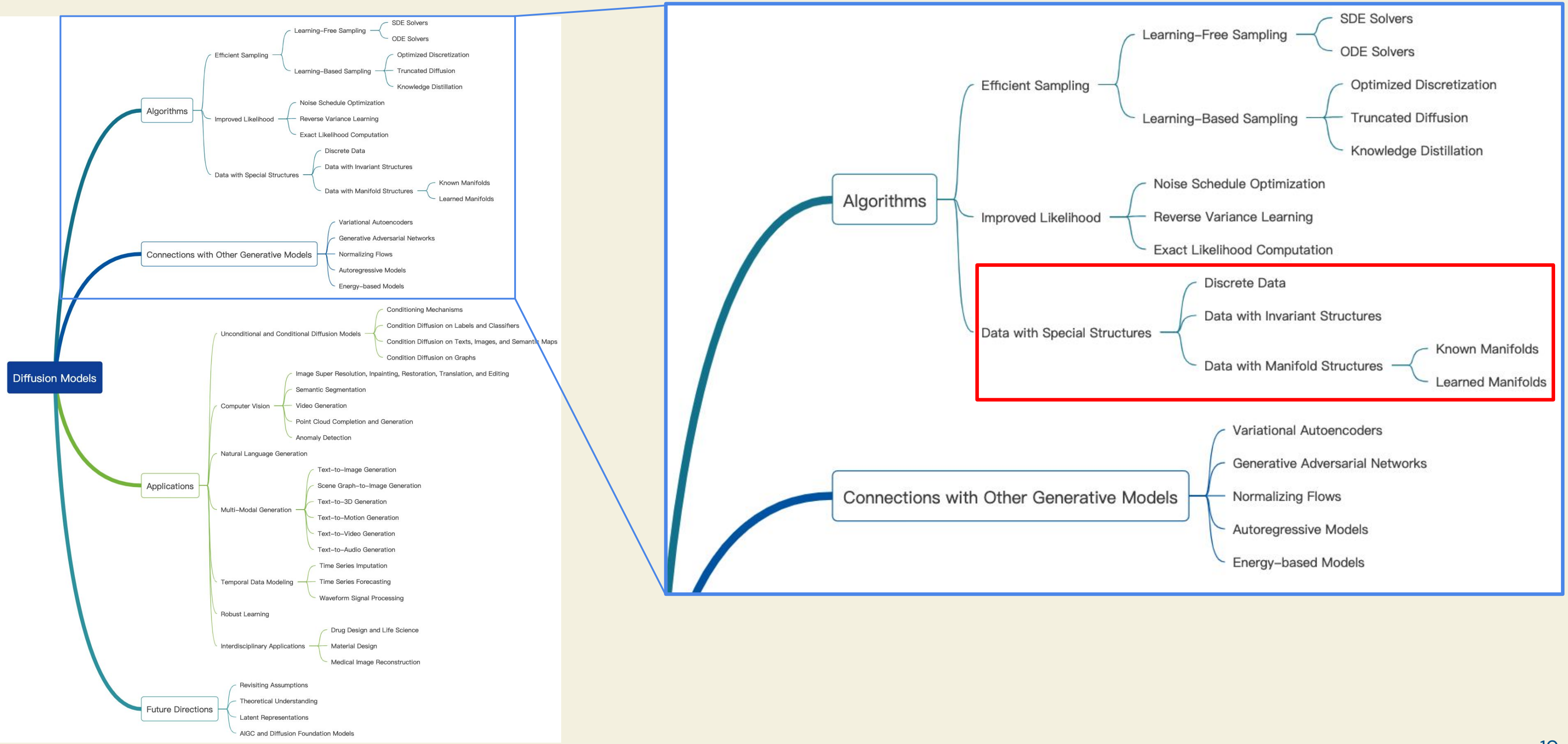
Training:

- first AE
- AE – fixed, then Diffusion

Connections to VAEs:

- diffusion is a prior

There is a lot of ***diffusion*** out there

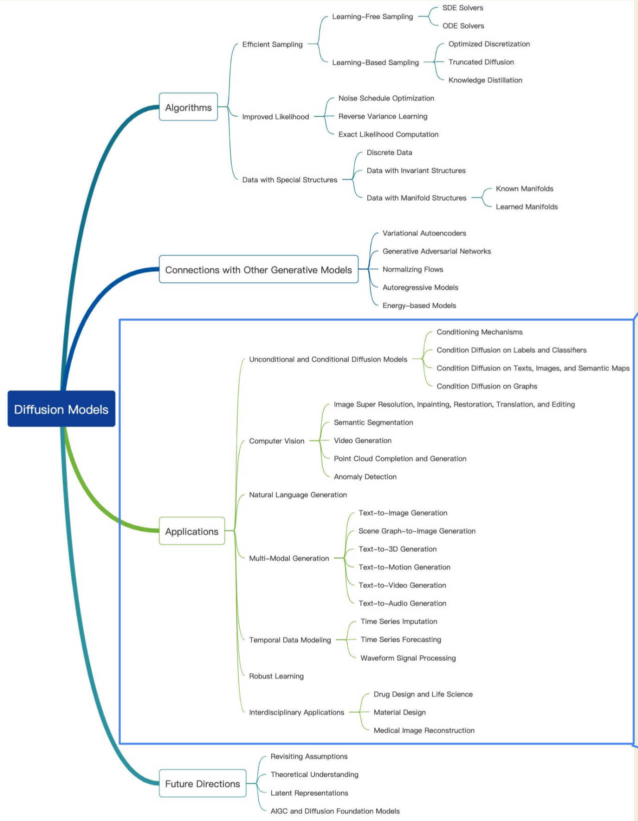


- Diffusion Models
  - Algorithms
    - Efficient Sampling
      - Learning-Free Sampling
        - SDE Solvers
        - ODE Solvers
      - Learning-Based Sampling
        - Optimized Discretization
        - Truncated Diffusion
        - Knowledge Distillation
    - Improved Likelihood
      - Noise Schedule Optimization
      - Reverse Variance Learning
      - Exact Likelihood Computation
    - Data with Special Structures
      - Discrete Data
      - Data with Invariant Structures
      - Data with Manifold Structures
        - Known Manifolds
        - Learned Manifolds
  - Connections with Other Generative Models
    - Variational Autoencoders
    - Generative Adversarial Networks
    - Normalizing Flows
    - Autoregressive Models
    - Energy-based Models
  - Applications
    - Unconditional and Conditional Diffusion Models
      - Conditioning Mechanisms
      - Condition Diffusion on Labels and Classifiers
      - Condition Diffusion on Texts, Images, and Semantic Maps
      - Condition Diffusion on Graphs
    - Computer Vision
      - Image Super Resolution, Inpainting, Restoration, Translation, and Editing
      - Semantic Segmentation
      - Video Generation
      - Point Cloud Completion and Generation
      - Anomaly Detection
    - Natural Language Generation
    - Multi-Modal Generation
      - Text-to-Image Generation
      - Scene Graph-to-Image Generation
      - Text-to-3D Generation
      - Text-to-Motion Generation
      - Text-to-Video Generation
      - Text-to-Audio Generation
    - Temporal Data Modeling
      - Time Series Imputation
      - Time Series Forecasting
      - Waveform Signal Processing
    - Robust Learning
    - Interdisciplinary Applications
      - Drug Design and Life Science
      - Material Design
      - Medical Image Reconstruction
  - Future Directions
    - Revisiting Assumptions
    - Theoretical Understanding
    - Latent Representations
    - AIGC and Diffusion Foundation Models

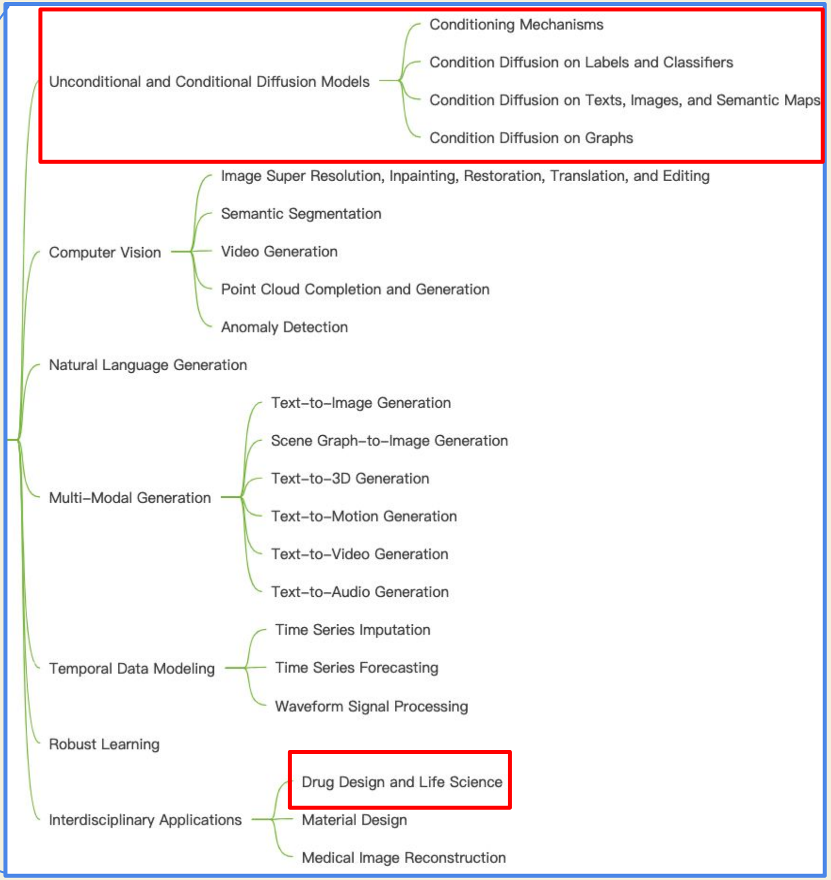
- Algorithms
  - Efficient Sampling
    - Learning-Free Sampling
      - SDE Solvers
      - ODE Solvers
    - Learning-Based Sampling
      - Optimized Discretization
      - Truncated Diffusion
      - Knowledge Distillation
  - Improved Likelihood
    - Noise Schedule Optimization
    - Reverse Variance Learning
    - Exact Likelihood Computation
  - Data with Special Structures
    - Discrete Data
    - Data with Invariant Structures
    - Data with Manifold Structures
      - Known Manifolds
      - Learned Manifolds
- Connections with Other Generative Models
  - Variational Autoencoders
  - Generative Adversarial Networks
  - Normalizing Flows
  - Autoregressive Models
  - Energy-based Models

Yang, Ling, et al. "Diffusion models: A comprehensive survey of methods and applications." ACM computing surveys 56.4 (2023): 1-39.

There is a lot of ***diffusion*** out there



- Diffusion Models
  - Algorithms
    - Efficient Sampling
      - Learning–Free Sampling
        - SDE Solvers
        - ODE Solvers
      - Learning–Based Sampling
        - Optimized Discretization
        - Truncated Diffusion
        - Knowledge Distillation
    - Improved Likelihood
      - Noise Schedule Optimization
      - Reverse Variance Learning
      - Exact Likelihood Computation
    - Data with Special Structures
      - Discrete Data
      - Data with Invariant Structures
      - Data with Manifold Structures
        - Known Manifolds
        - Learned Manifolds
  - Connections with Other Generative Models
    - Variational Autoencoders
    - Generative Adversarial Networks
    - Normalizing Flows
    - Autoregressive Models
    - Energy–based Models
  - Applications
    - Unconditional and Conditional Diffusion Models
      - Conditioning Mechanisms
      - Condition Diffusion on Labels and Classifiers
      - Condition Diffusion on Texts, Images, and Semantic Maps
      - Condition Diffusion on Graphs
    - Computer Vision
      - Image Super Resolution, Inpainting, Restoration, Translation, and Editing
      - Semantic Segmentation
      - Video Generation
      - Point Cloud Completion and Generation
      - Anomaly Detection
    - Natural Language Generation
    - Multi–Modal Generation
      - Text–to–Image Generation
      - Scene Graph–to–Image Generation
      - Text–to–3D Generation
      - Text–to–Motion Generation
      - Text–to–Video Generation
      - Text–to–Audio Generation
    - Temporal Data Modeling
      - Time Series Imputation
      - Time Series Forecasting
      - Waveform Signal Processing
    - Robust Learning
    - Interdisciplinary Applications
      - Drug Design and Life Science
      - Material Design
      - Medical Image Reconstruction
  - Future Directions
    - Revisiting Assumptions
    - Theoretical Understanding
    - Latent Representations
    - AIGC and Diffusion Foundation Models

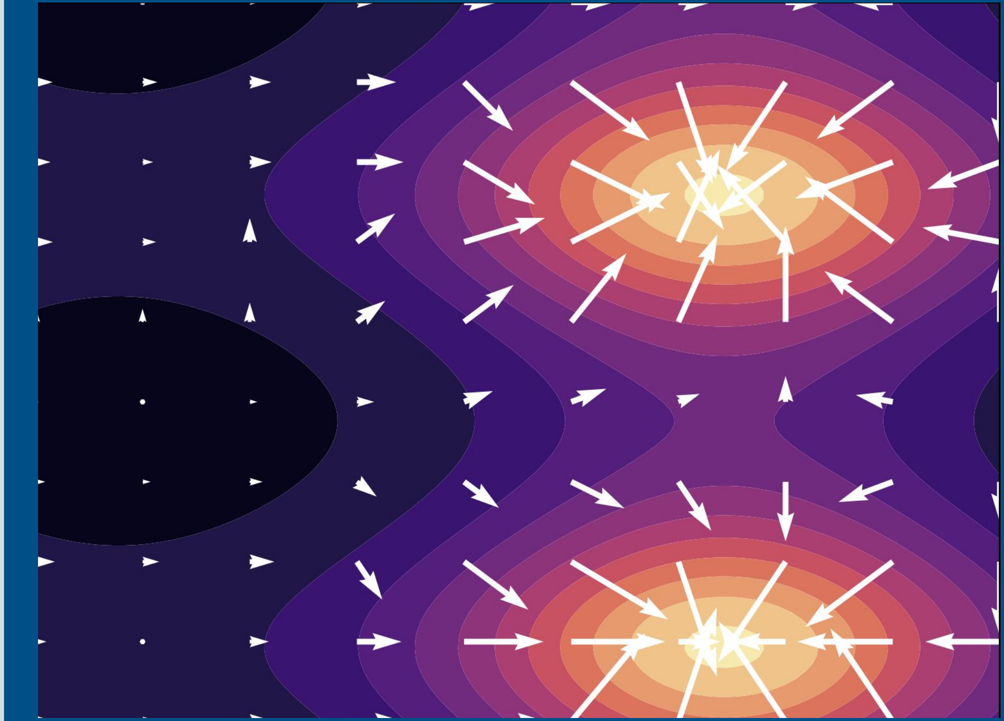


- Unconditional and Conditional Diffusion Models
  - Conditioning Mechanisms
  - Condition Diffusion on Labels and Classifiers
  - Condition Diffusion on Texts, Images, and Semantic Maps
  - Condition Diffusion on Graphs
- Computer Vision
  - Image Super Resolution, Inpainting, Restoration, Translation, and Editing
  - Semantic Segmentation
  - Video Generation
  - Point Cloud Completion and Generation
  - Anomaly Detection
- Natural Language Generation
- Multi–Modal Generation
  - Text–to–Image Generation
  - Scene Graph–to–Image Generation
  - Text–to–3D Generation
  - Text–to–Motion Generation
  - Text–to–Video Generation
  - Text–to–Audio Generation
- Temporal Data Modeling
  - Time Series Imputation
  - Time Series Forecasting
  - Waveform Signal Processing
- Robust Learning
- Interdisciplinary Applications
  - Drug Design and Life Science
  - Material Design
  - Medical Image Reconstruction

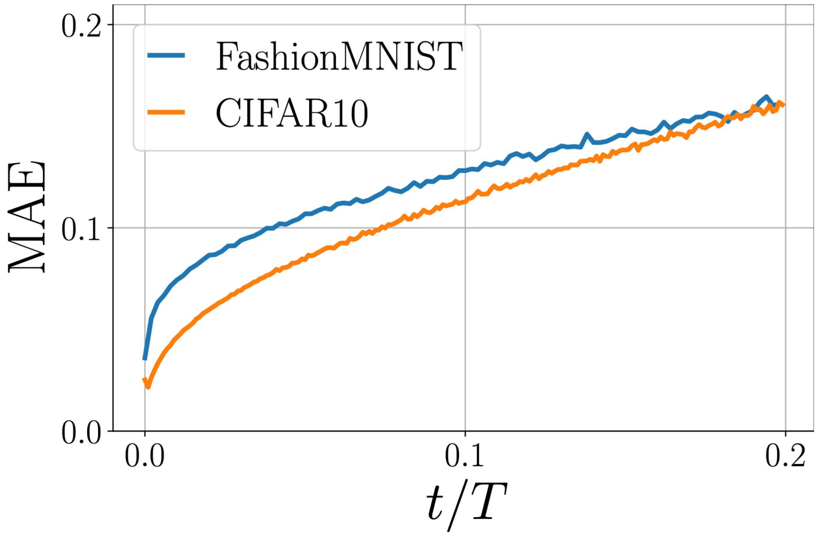
Yang, Ling, et al. "Diffusion models: A comprehensive survey of methods and applications." ACM computing surveys 56.4 (2023): 1-39.

Understanding diffusion models

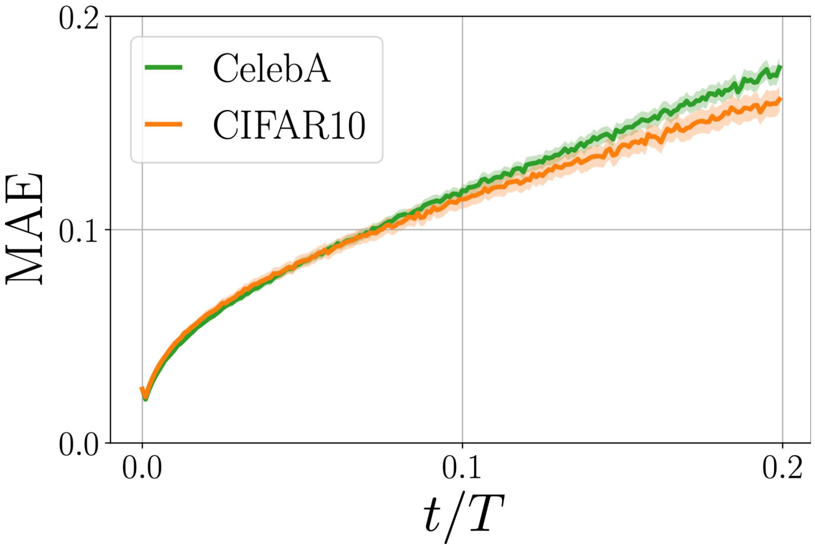
Early diffusion steps behave similarly across datasets and that denoisers learn rich, reusable representations. Building on this insight enables efficient generation, improved generalization, and interpretable visual counterfactuals



Are all steps in diffusion models born equal?



After ~10% of the steps, the reconstruction error starts growing, and the MAE increases linearly above 0.1 (i.e., about 6% of error per pixel).

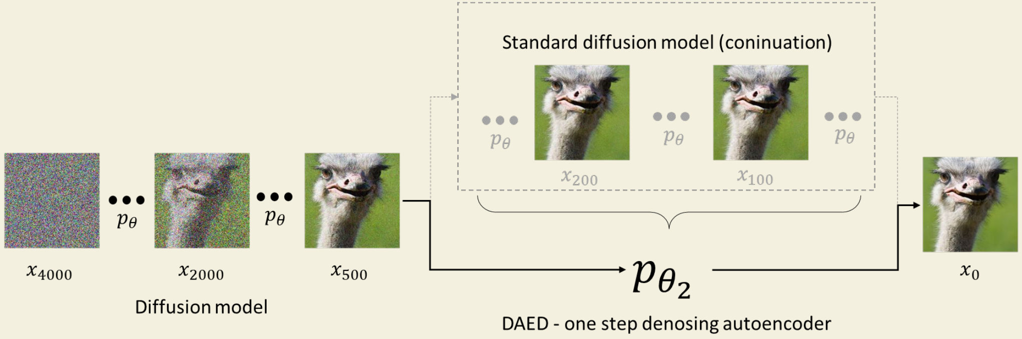


The MAE for a DDGM trained on CIFAR10 and evaluated on CIFAR10 & CelebA: For the first ~10% of steps MAE is the same! ***Can we reuse?***

Deja, (...), **Tomczak**, "On Analyzing Generative and Denoising Capabilities of Diffusion-based Deep Generative Models", NeurIPS 2022

We proposed **DAED**: Denoising Auto-Encoder with Diffusion

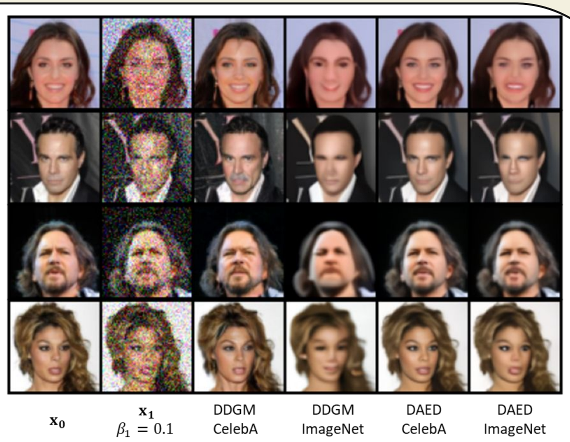
Idea: Take a **denoising auto-encoder** and add a **diffusion-based prior**.



We have:

- two denoising nets;
- ***denoising*** is done in 1 step;
- ***generation*** is done in multiple steps;
- The objective:

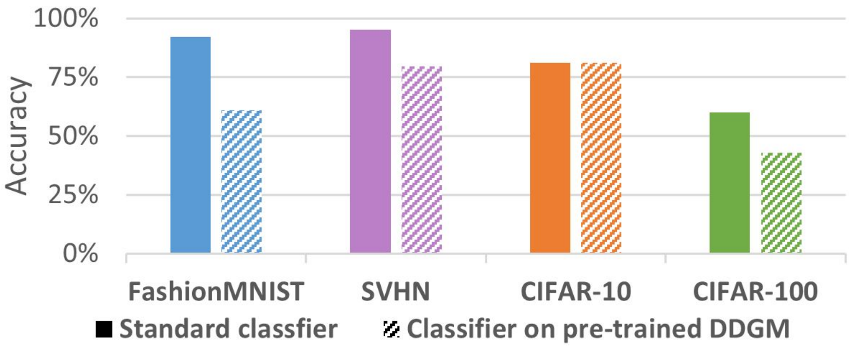
$$\bar{\ell}(\mathbf{x}_0;\varphi,\theta) = \mathbb{E}_{\mathbf{x}_1\sim q(\mathbf{x}_1|\mathbf{x}_0)}\left[\ln p\left(\mathbf{x}_0|f_\varphi\left(\mathbf{x}_1\right)\right) + \ln p_\theta(\mathbf{x}_1)\right]$$
$$\geq \underbrace{\mathbb{E}_{\mathbf{x}_1\sim q(\mathbf{x}_1|\mathbf{x}_0)}\left[\ln p\left(\mathbf{x}_0|f_\varphi\left(\mathbf{x}_1\right)\right)\right]}_{\ell_{\mathrm{DAE}}(\mathbf{x}_0;\varphi)} + \underbrace{\mathbb{E}_{q(\mathbf{x}_2,\ldots,\mathbf{x}_T|\mathbf{x}_1)}\left[\frac{\ln p_\theta(\mathbf{x}_1,\ldots,\mathbf{x}_T)}{q(\mathbf{x}_1,\ldots,\mathbf{x}_T|\mathbf{x}_0)}\right]}_{\ell_{\mathrm{D}}(\mathbf{x}_0;\theta)}$$



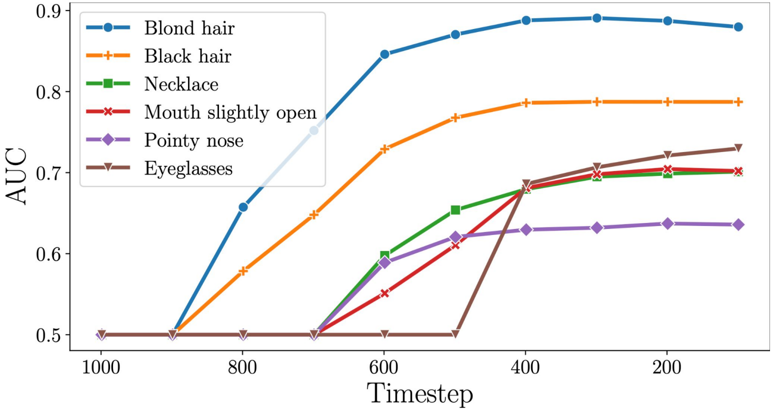
Denoising an image from 10% noise: For DDGM, if a denoiser's trained on another data, it fails; but it <u>works</u> for DAED!

Deja, (...), **Tomczak**, "On Analyzing Generative and Denoising Capabilities of Diffusion-based Deep Generative Models", NeurIPS 2022

What is *inside* denoiser nets? Useful representations!



Averaged representations of an image given by a denoiser net (e.g., UNet) are *useful* (here: classification accuracy) for an MLP-based classifier trained on them.

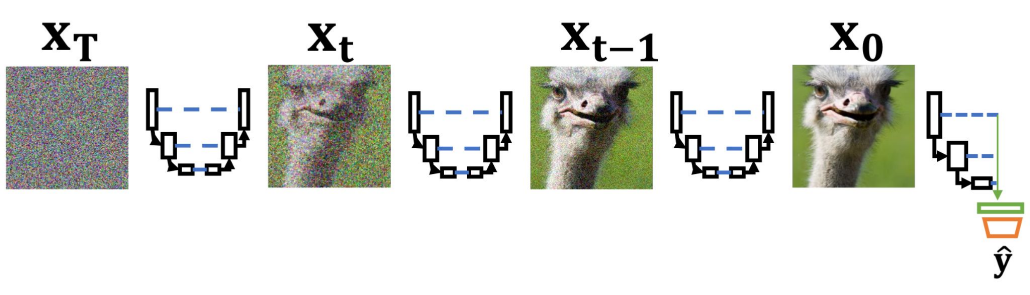


Training binary logistic regressors over attributes from CelebA based on averaged representations from a denoiser net results in non-random performance over time (and sometimes quite quickly)!

Deja, Trzcinski, **Tomczak**, "Learning Data Representations with Joint Diffusion Models", ECML 2023

We proposed **Joint Diffusion:** DDGM + classifier trained together

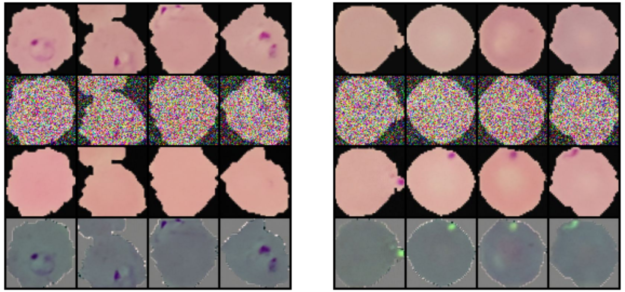
<u>Idea</u>: Take a **diffusion model** and add **a classifier to the denoiser net.**



$$\ln p_{\nu,\psi,\omega}(\mathbf{x}_{0:T}, y) = \ln p_{\nu,\omega}(y|\mathbf{x}_0) + \ln p_{\nu,\psi}(\mathbf{x}_{0:T})$$

We have:

- a single denoising net;
- an extra ***classification head***;
- we can use classifier for guidance: use 1-step SGD during sampling through ln p(*y*|**x**)



Malaria ->No Malaria

No Malaria ->Malaria

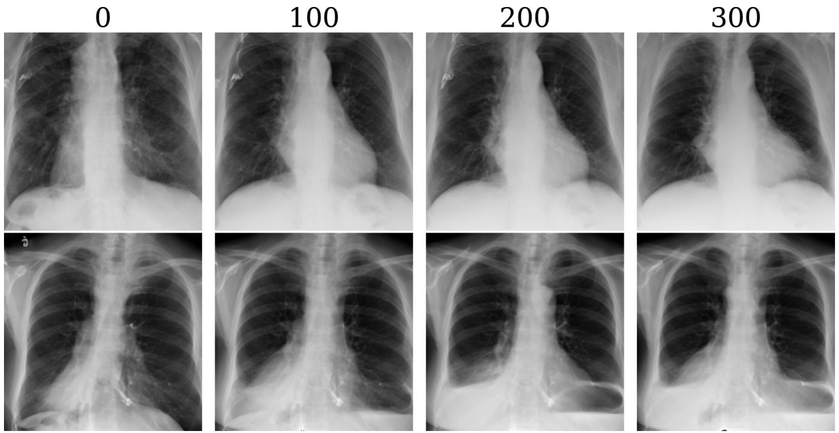
Visual counterfactuals
Take an image, add ~10% noise & flip the class label, and *reconstruct*.
The model <u>removes</u>/<u>adds</u> info!

Deja, Trzcinski, **Tomczak**, "Learning Data Representations with Joint Diffusion Models", ECML 2023

Latent diffusion models are the way to go?

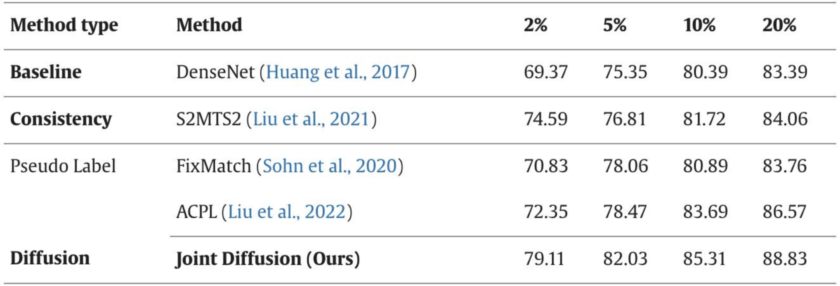
Latent diffusion models scale diffusion to high-resolution images and discrete biomedical data by combining auto-encoding, diffusion, and prediction in latent space. Latent diffusion provides a unifying framework from pixels to cells.

How to deal with high-res data and have a joint model? **Joint LDMs**!



Examples of generations with increasing classifier guidance strength.

Using previous ideas works just fine!



Method type	Method	2%	5%	10%	20%
Baseline	DenseNet (Huang et al., 2017)	69.37	75.35	80.39	83.39
Consistency	S2MTS2 (Liu et al., 2021)	74.59	76.81	81.72	84.06
Pseudo Label	FixMatch (Sohn et al., 2020)	70.83	78.06	80.89	83.76
	ACPL (Liu et al., 2022)	72.35	78.47	83.69	86.57
Diffusion	**Joint Diffusion (Ours)**	79.11	82.03	85.31	88.83

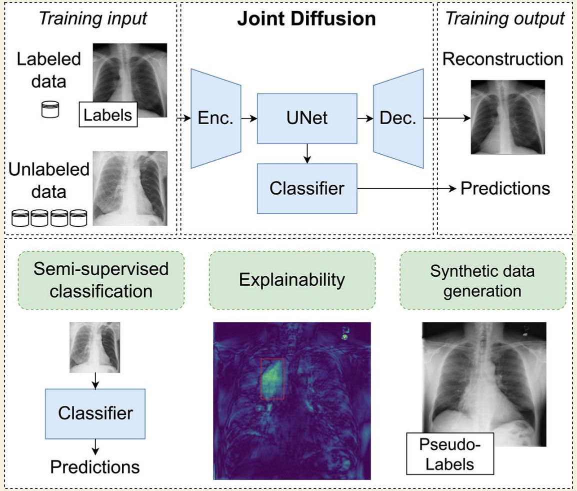
Taking advantage of semi-supervised learning. Performance comparison of different methods at various label percentages on the ISIC 2019 dataset.

We can significantly improve the classification accuracy without using any additional tricks!

Kaleta, (...), **Tomczak**, Deja "JointDiffusion: Joint representation learning for generative, predictive, and self-explainable AI in healthcare." *Computerized Medical Imaging and Graphics*, 2025

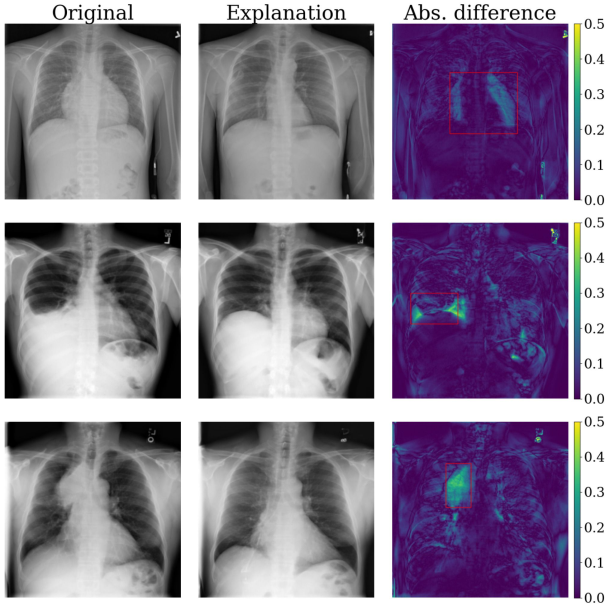
We proposed **Joint LDM**: Formulate a joint model in the latent space

Idea: Take a **joint diffusion** to the **latent space.**



We have:

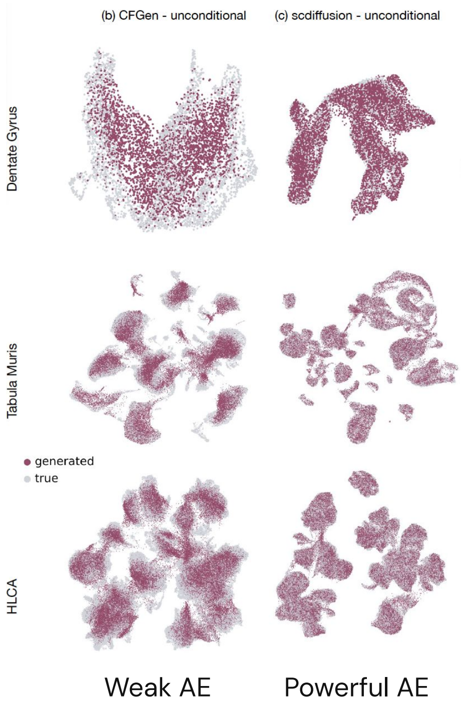
- AE + joint diffusion;
- semi-supervised learning for free;
- a way of dealing with high-res images like medical scans.



Helping with diagnosis: Images are modified to reduce disease prediction probability across different classes. Expert-annotated bounding boxes highlight ROIs.

Kaleta, (...), **Tomczak**, Deja "JointDiffusion: Joint representation learning for generative, predictive, and self-explainable AI in healthcare." *Computerized Medical Imaging and Graphics*, 2025

How to deal with discrete data like single-cell transcriptomics?



Working directly in the discrete space is troublesome.

VAEs are *the way to go* since they allow working in the latent space.

However, **weak AEs can cripple** the performance.

Now, questions are the following:

(1) How to deal with exchangeable data?

(2) How to formulate adaptive and powerful autoencoders?

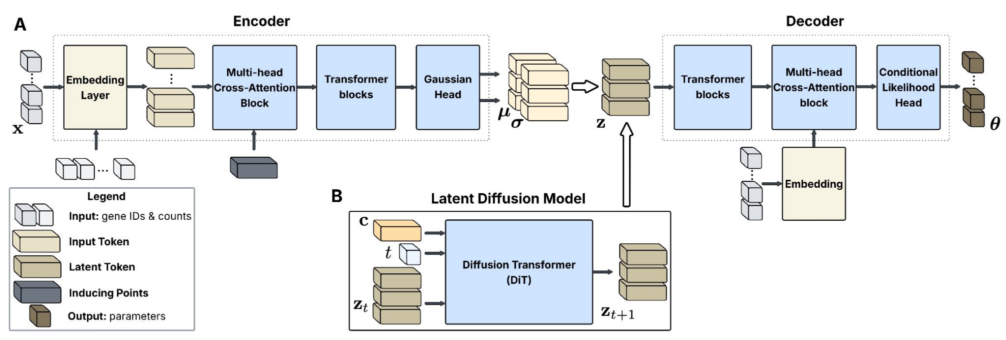
(3) How to ensure *rich* and *flexible* latent spaces (embeddings)?

(4) How to make the whole approach scalable, i.e., more data = better performance?

Palla, (...), **Tomczak**, "Scalable Single-Cell Gene Expression Generation with Latent Diffusion Models", arXiv 2025

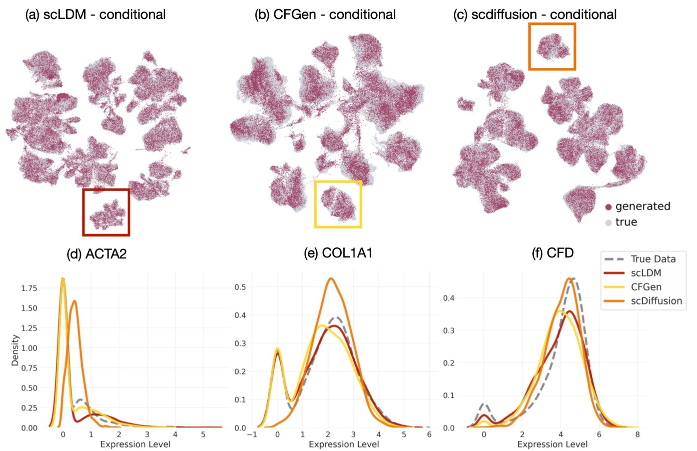
We proposed **scLDM**: LDM + classifier-free guidance for perturbations

Idea: Train a **fully transformer-based LDM**



We have:

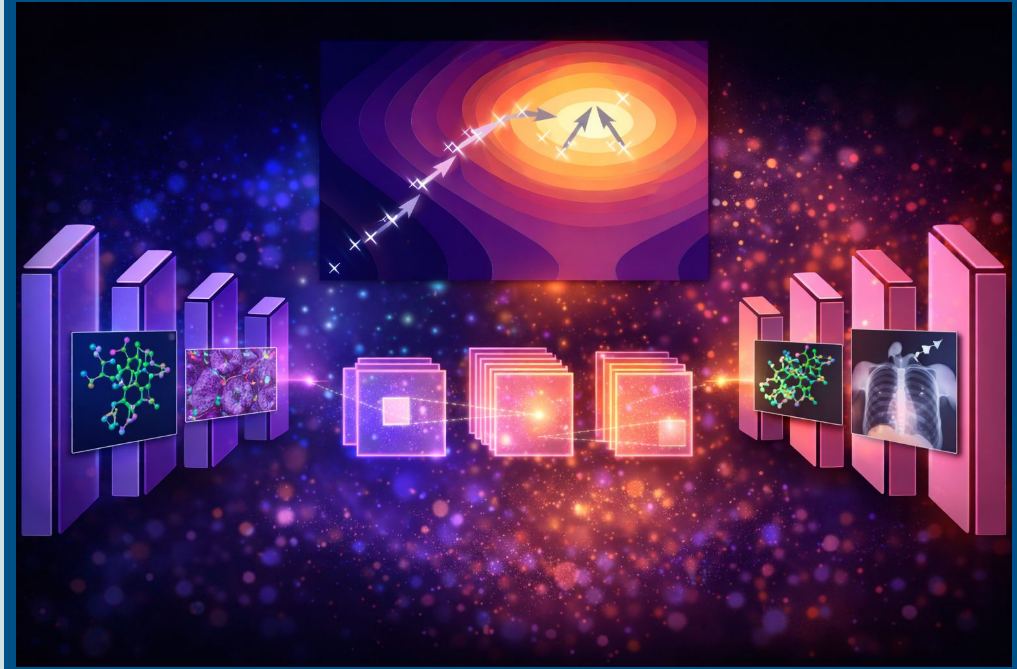
- a transformer-based AE, permutation-invariant encoder, permutation-equivariant decoder ⇒ exchanchable model;
- tokenized latent space (important!);
- out-of-the-box Diffusion Transformers in the latent space;
- classifier-free guidance for perturbations.



scLDM fits data well locally: Comparing distributions locally reveal our approach achieves better fit to data distribution, covering modes better.

Palla, (...), **Tomczak**, "Scalable Single-Cell Gene Expression Generation with Latent Diffusion Models", arXiv 2025

Conclusion



AI4Science has evolved from expert systems, through data mining, to deep learning era. Now, the key is how to get **data**, utilize **prior structures**, and blend them in **generative models**.

And we are just starting!