Diffusion Models

How AI Learns to Create Images from Noise

Presented by: *Anindya, Tamal and Din*

February 19, 2026

CSE 200: Technical Writing and Presentation
Bangladesh University of Engineering and Technology

What Problem Are We Solving?

Generative Modeling

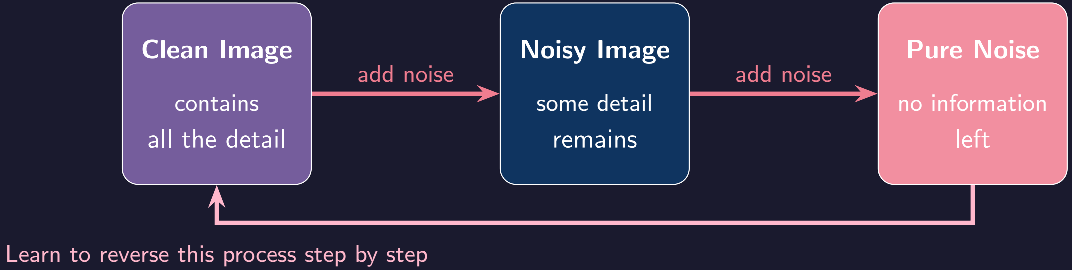
Given many examples of real data (e.g. photos of faces), can a machine learn to **create new, realistic samples** that look just as real?

Insert image:
Real photos vs. AI-generated
(e.g. faces, landscapes)

Why is this hard?

- Real data is incredibly complex and high-dimensional
- The model must capture fine details *and* global structure
- Previous approaches each had trade-offs:

The Core Idea: Learning by Destroying



Forward process (fixed, no learning):

- Gradually add small amounts of random noise
- Eventually all structure is destroyed

Reverse process (learned):

- Train a neural network to undo each tiny noise step
- Starting from pure noise, the network reconstructs an image

The Forward Process: Adding Noise

Insert image: progressive noising sequence

Clean image $\rightarrow$ slightly noisy $\rightarrow$ more noisy $\rightarrow \ldots \rightarrow$ pure noise

What Happens

Take a real image and add a *tiny* bit of random noise. Repeat this hundreds of times until only noise remains.

The Reverse Process: Removing Noise

Insert image: progressive denoising sequence

Pure noise $\rightarrow$ rough shape $\rightarrow$ clearer structure $\rightarrow$ final image

How Generation Works

- Start from **pure noise**
- Predict and remove a little noise at each step
- Repeat until a realistic image appears

How Do We Train It?

1. Take a real image from the training set
2. Pick a random step and add the corresponding noise
3. Ask the neural network: *"what noise was added?"*
4. Compare the prediction to the **actual noise**, minimize the error

Key Insight & Why It Works

Key Insight

- We add **known** noise
- So we have exact labels
- Training target is clear and reliable

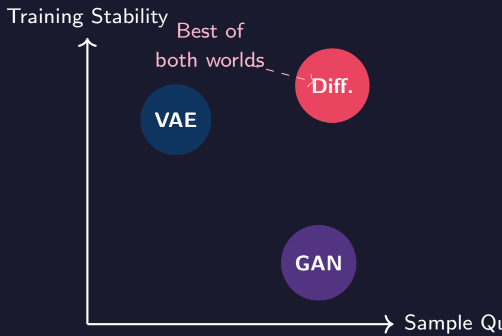
Why It Works

- Model learns many small denoising steps
- Small steps are easier to learn
- Chaining them builds a full image

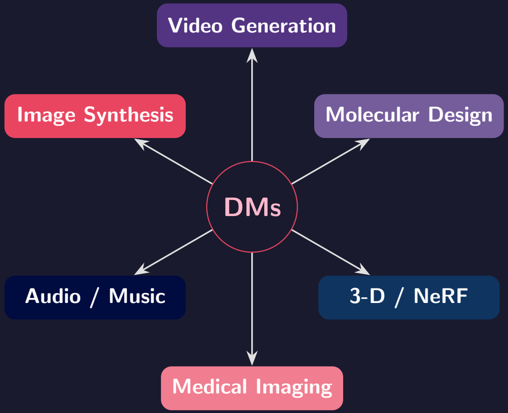
Why Diffusion Models Work So Well

Advantages

- **Stable training** — no adversarial game; just predict noise
- **Small steps** — each denoising step is a tiny, easy task
- **High quality** — state-of-the-art image fidelity
- **Flexible** — works on images, audio, video, 3-D, molecules. . .



Applications



Real-world tools powered by diffusion:

- **DALL·E 3** — text-to-image
- **Stable Diffusion** — open-source image generation
- **Midjourney** — art and design
- **Sora** — text-to-video

Limitations & Open Challenges

Current Limitations

- **Slow generation** — requires many denoising steps (hundreds to thousands)
- **High compute cost** — training requires significant GPU resources
- **Large model size** — not easy to run on consumer hardware

Open Questions

- Can we make generation **faster**? (Active research area)
- How do we ensure generated content is **safe and ethical**?
- Can we extend to **longer videos** and **interactive content**?

Thank You!

Questions?

References

[1] J. Ho et al. **“Denoising diffusion probabilistic models”.** In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. NIPS '20. Vancouver, BC, Canada: Curran Associates Inc., 2020.

[2] Y. Song and S. Ermon. **“Generative modeling by estimating gradients of the data distribution”.** In: *Advances in neural information processing systems* 32 (2019).

[3] J. Song et al. **“Denoising diffusion implicit models”.** In: *arXiv preprint arXiv:2010.02502* (2020).

[4] A. Q. Nichol and P. Dhariwal. **“Improved denoising diffusion probabilistic models”.** In: *International conference on machine learning*. PMLR. 2021, pp. 8162–8171.