Applications

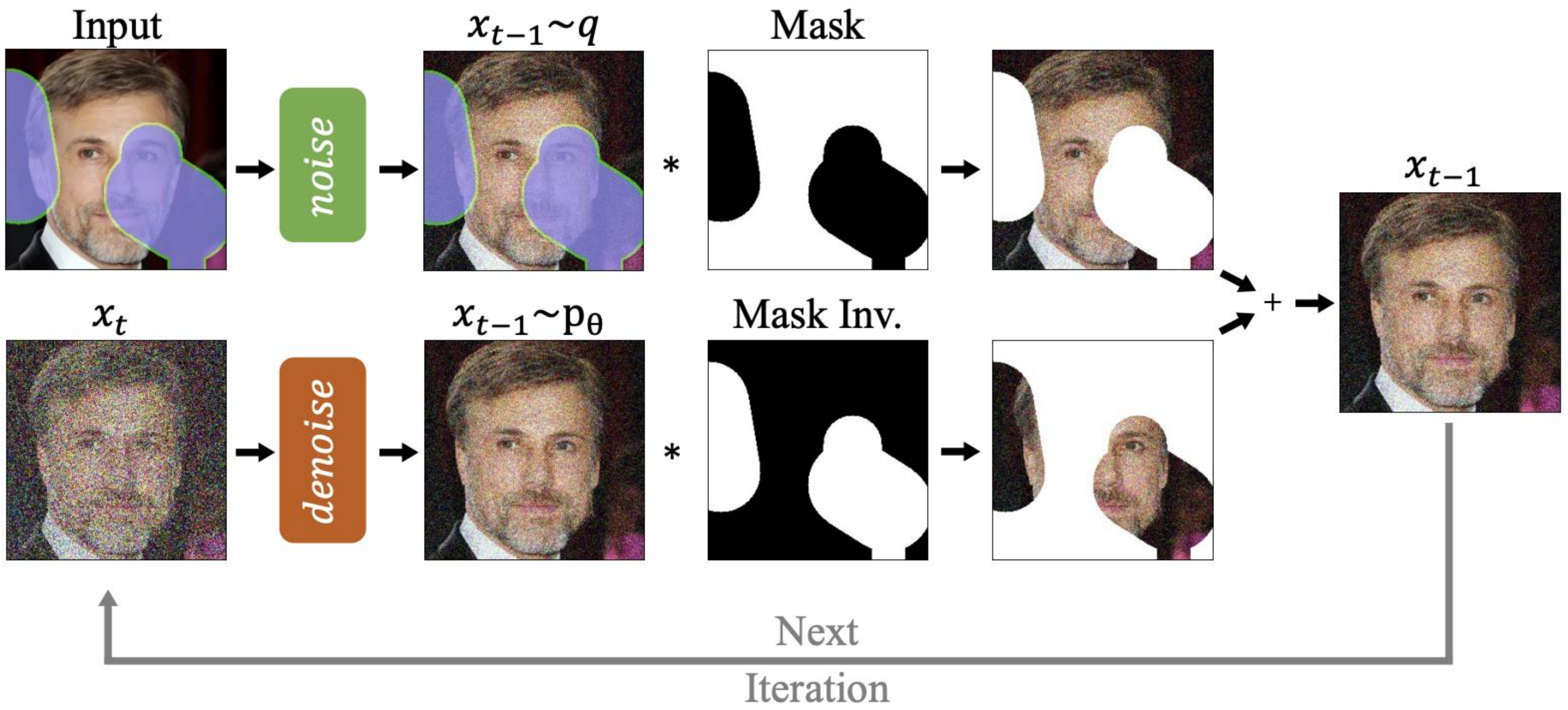
Applications

Diffusion Models: A Comprehensive Survey of Methods and Applications

LING YANG, Peking University, China
ZHILONG ZHANG*, Peking University, China
YANG SONG, OpenAI, USA
SHENDA HONG, Peking University, China
RUNSHENG XU, University of California, Los Angeles, USA
YUE ZHAO, Carnegie Mellon University, USA
WENTAO ZHANG, Peking University, China
BIN CUI, Peking University, China
MING-HSUAN YANG†, University of California at Merced, USA

Applications

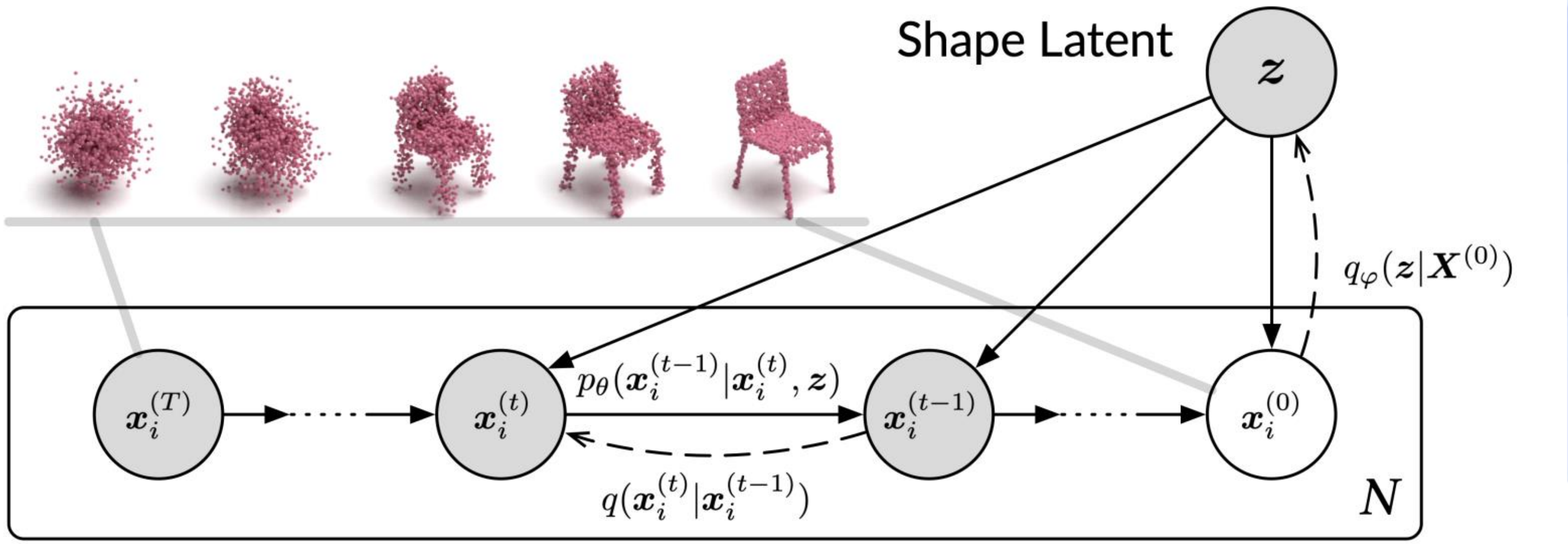
Computer Vision – inpainting



Applications

Computer vision – point cloud generation

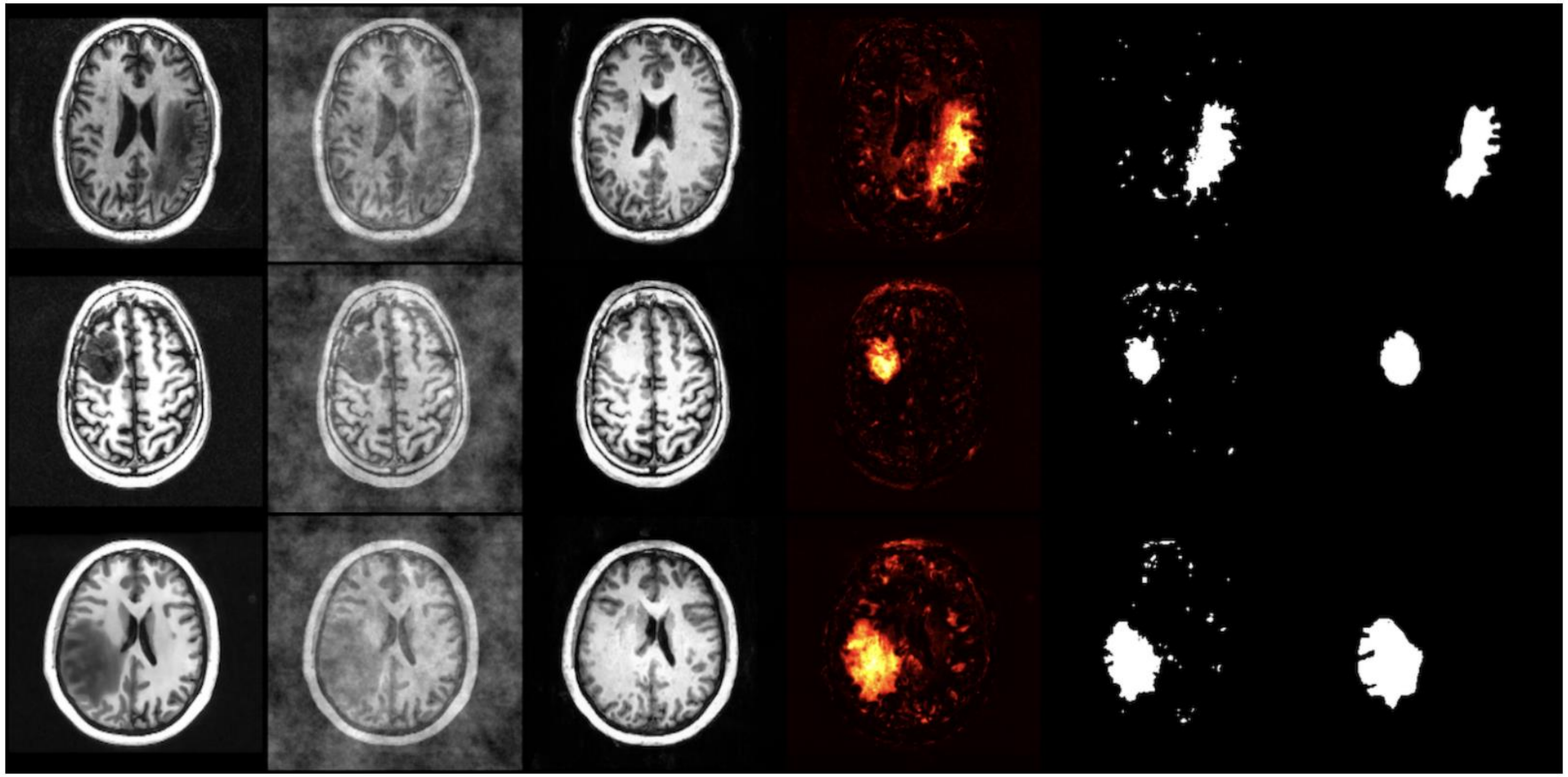
- Scans often miss information due to partial observation or occlusion.
- Use diffusion models to infer missing parts.
- Treat the points in a cloud as a set of particles in an evolving thermodynamic system.



Applications

Computer vision – anomaly detection

- AnoDDPM – attempt to "repair" patient data:
- Instead of adding Gaussian noise, add power law noise



Applications

Natural language generation

- Due to the success of ChatGPT, we tend to forget that Stable Diffusion 2 was released BEFORE GPT-4.
- It is reasonable to think that there might be some good diffusion-based language models…

…but LLMs are still king in this arena

- For a detailed exploration of why see:

https://sander.ai/2023/01/09/diffusion-language.html

Applications

Multimodal generation – text-to-image

- We already covered SD, but there are numerous other models:
  - DALL-E 1,2 and 3
  - Imagen
  - GLIDE
  - VQ-Diffusion
- There are also models which build upon latent diffusion models:
  - DreamBooth – a method of fine tuning pretrained models

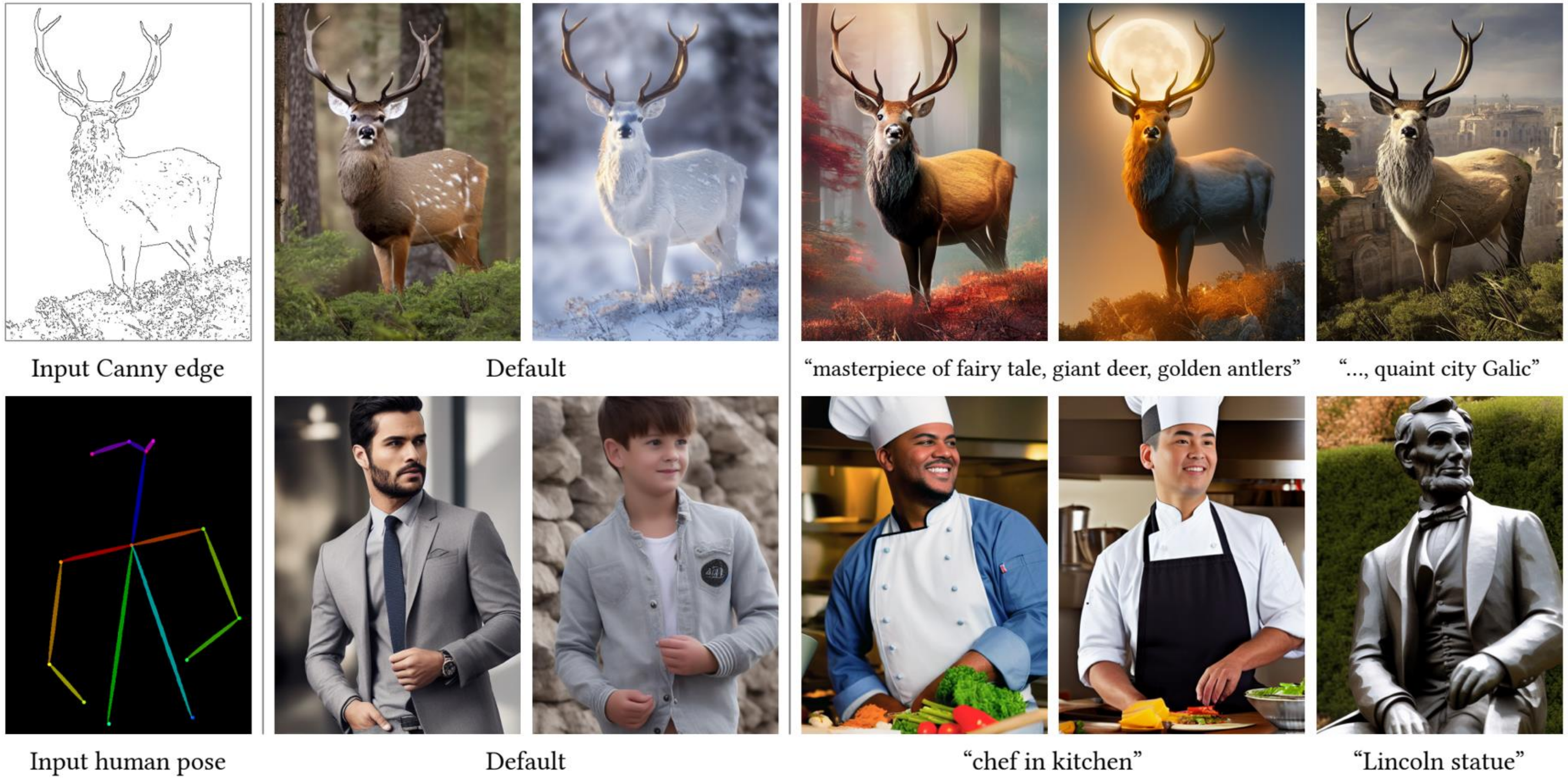
ACCELERATE PROGRAMME FOR SCIENTIFIC DISCOVERY

SCHMIDT FUTURES

Applications

Multimodal generation – text-to-image

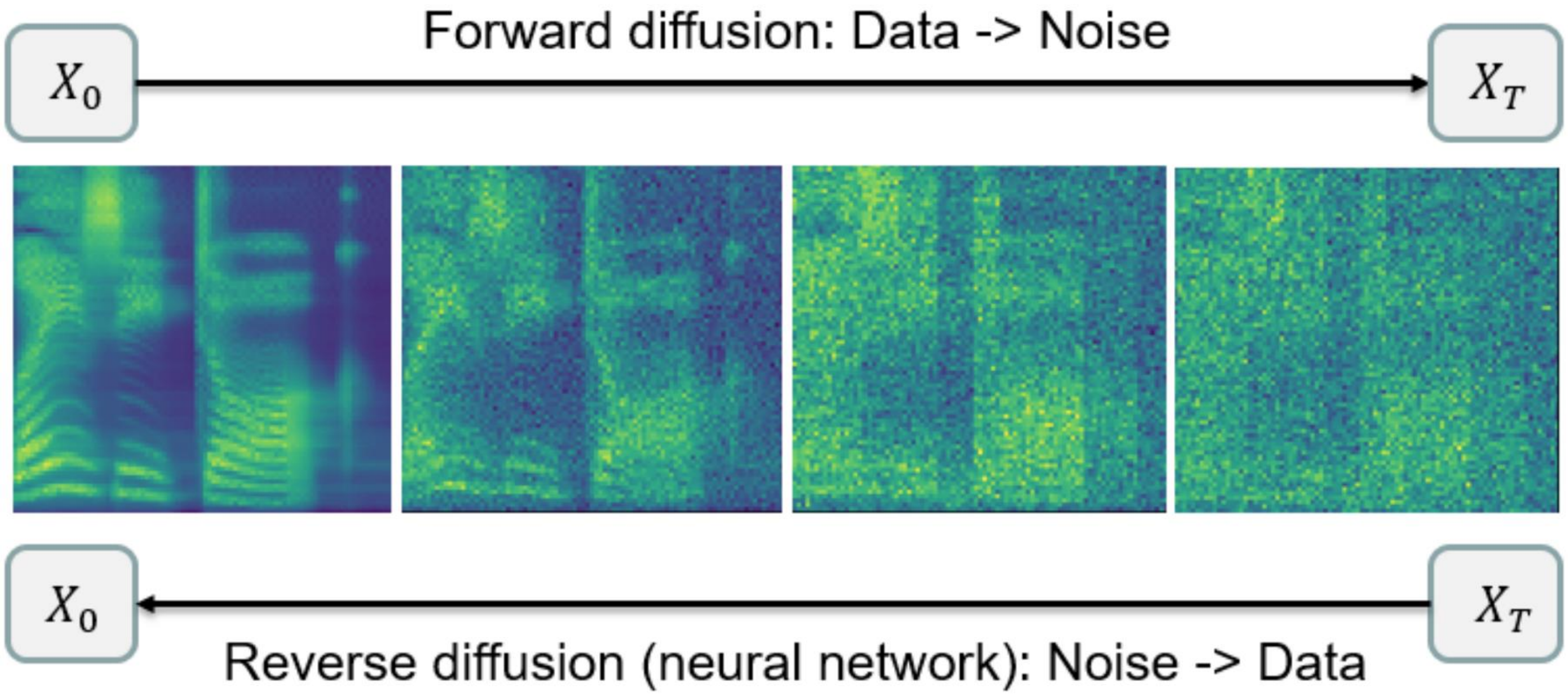
- ControlNet – additional spatial conditioning controls:



Applications

Multimodal generation – text-to-audio

- Grad-TTS – text-to-speech:



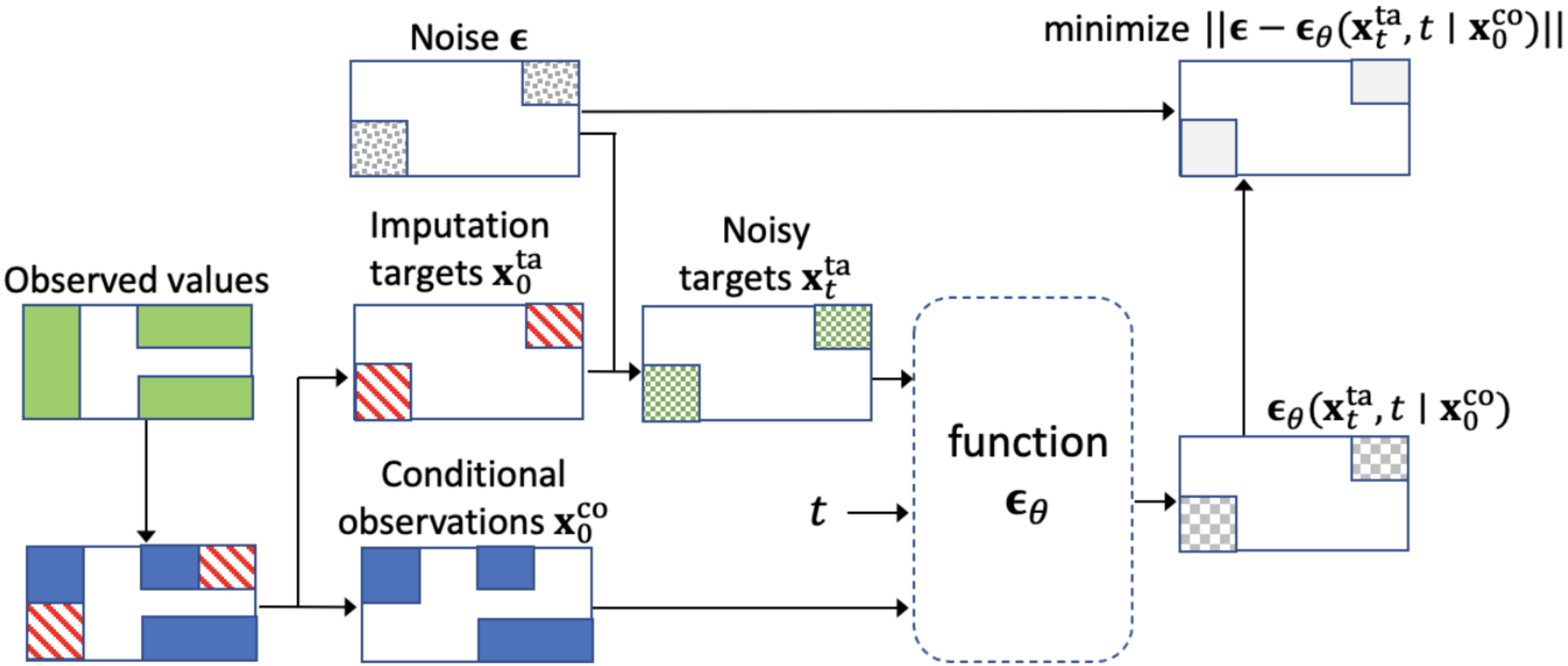
ACCELERATE PROGRAMME FOR SCIENTIFIC DISCOVERY

SCHMIDT FUTURES

Applications

Temporal data modeling – imputation

- Real world time series data often contains missing information
- Imputation is the process of filling in that missing data:



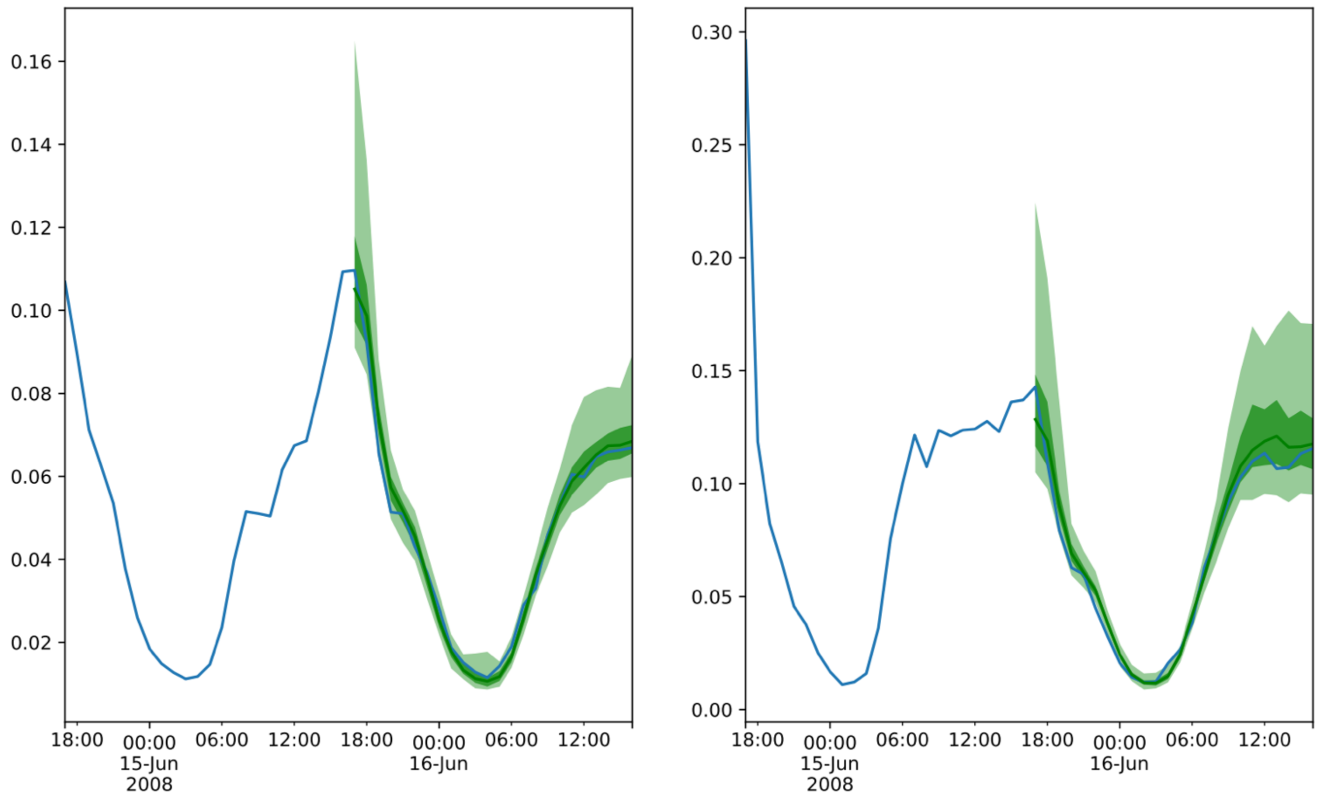
ACCELERATED PROGRAM FOR SCIENTIFIC DISCOVERY

SCHMIDT FUTURES

Applications

Temporal data modeling – forecasting

- TimeGrad – multivariate probabilistic time series forecasting
- Uses a RNN to predict traffic data…



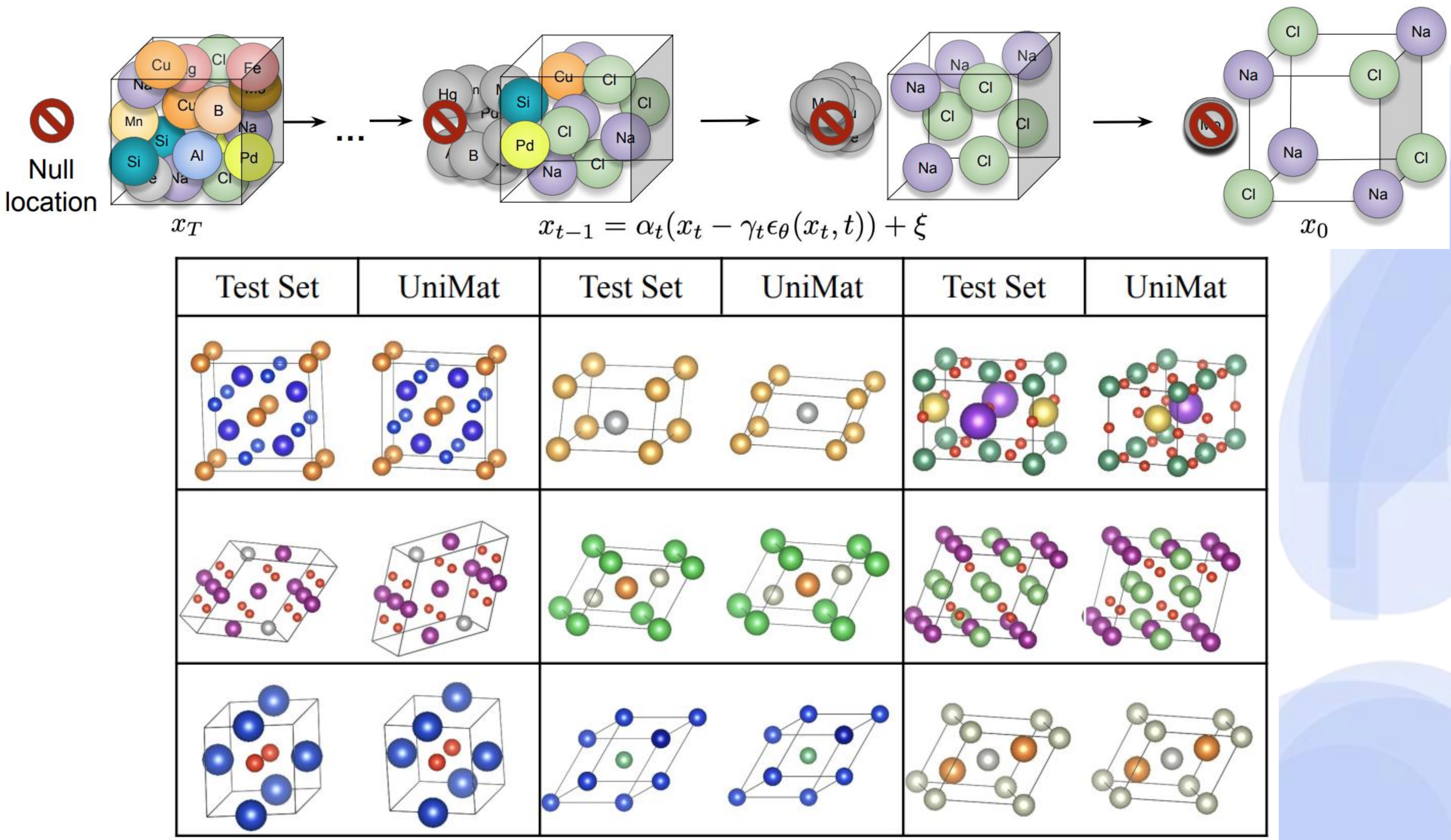
ACCELERATE PROGRAMME FOR SCIENTIFIC DISCOVERY

SCHMIDT FUTURES

Applications

Materials

- Scalable diffusion for Materials Generation, DeepMind



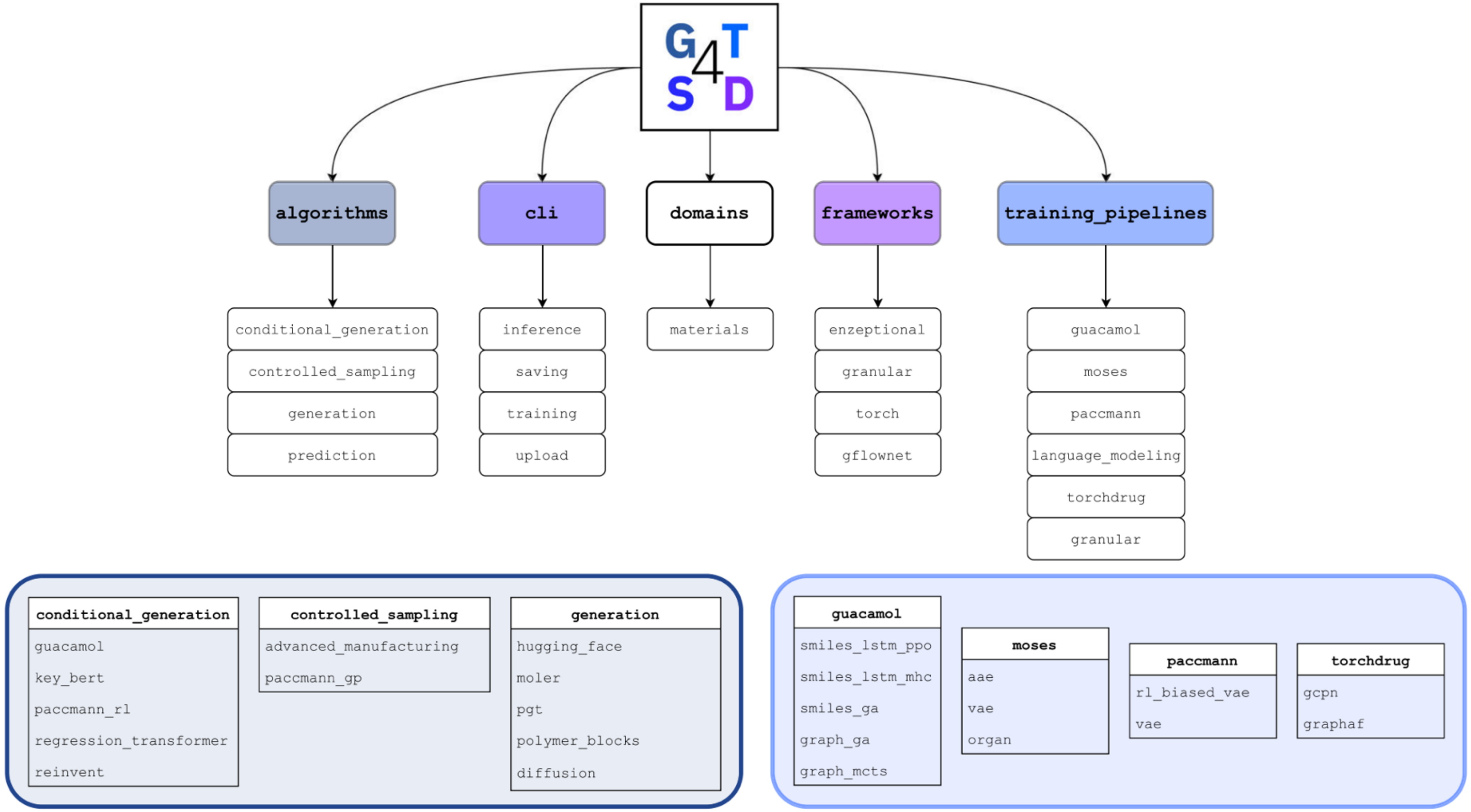
ACCELERATE PROGRAMME FOR SCIENTIFIC DISCOVERY

SCHMIDT FUTURES

Applications

Materials

- Generative toolkit for scientific discovery



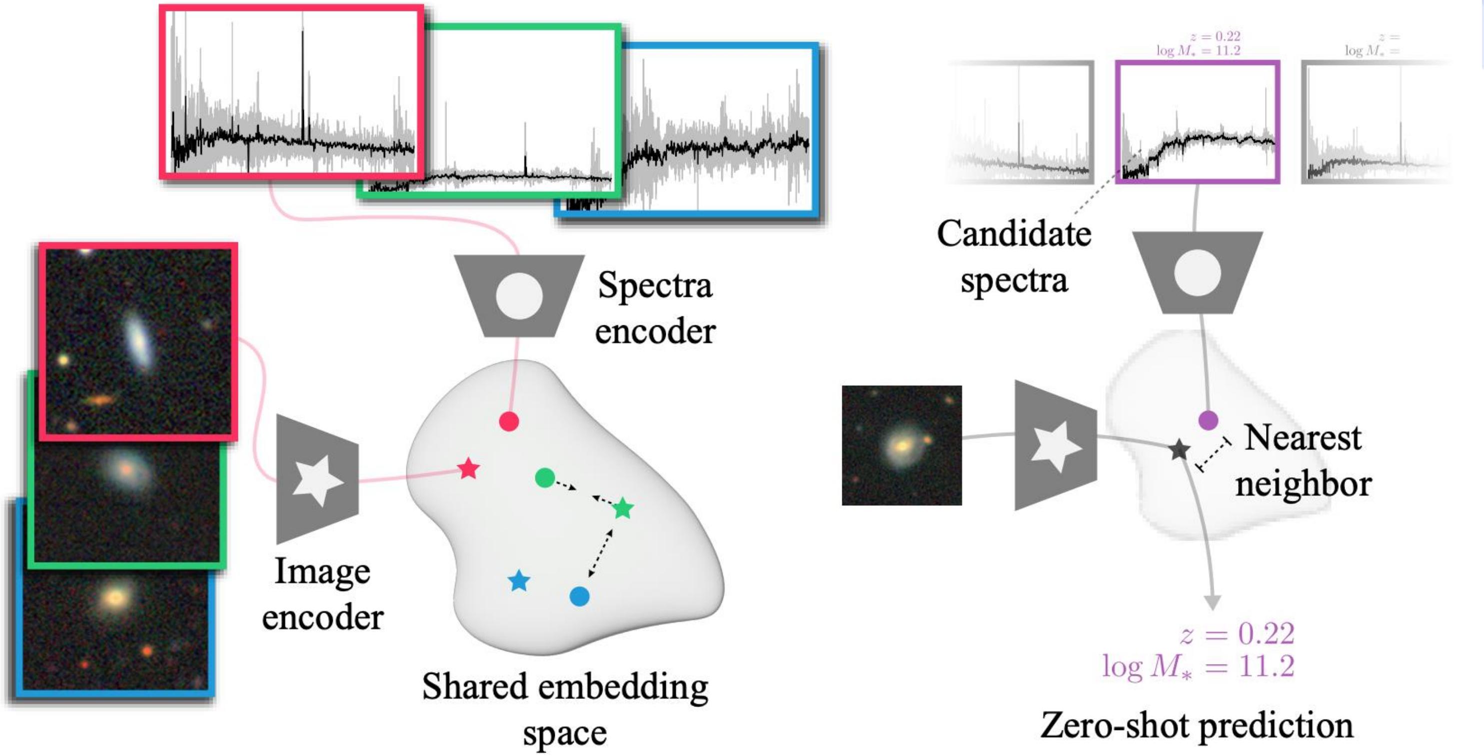
ACCELERATE PROGRAMME FOR SCIENTIFIC DISCOVERY

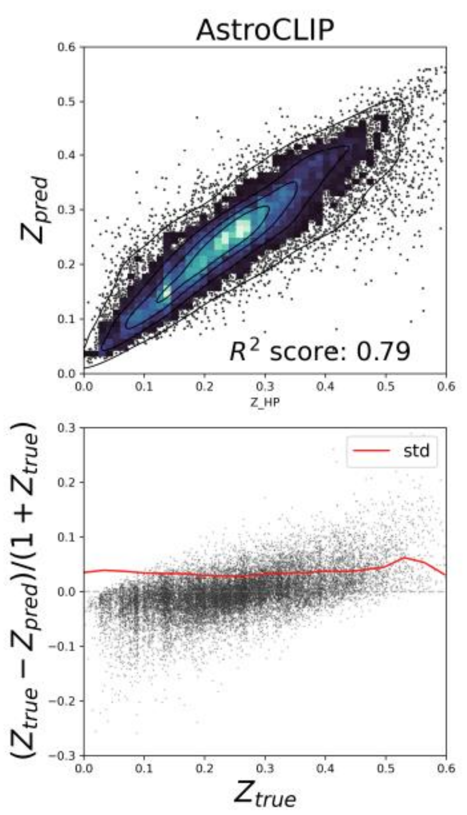
SCHMIDT FUTURES

Applications

Astrophysics

- AstroCLIP: A Cross-Modal Foundation Model for Galaxies, Parker et al, 2023





ACCELERATE PROGRAMME FOR SCIENTIFIC DISCOVERY

SCHMIDT FUTURES

Evaluating Diffusion Models

ACCELERATE PROGRAMME
FOR SCIENTIFIC DISCOVERY

SCHMIDT FUTURES

Evaluation

How can we effectively evaluate diffusion models?

- How you choose to evaluate will be heavily dependent upon the task
- The maturity of the metrics will also depend on the task
- Evaluation for stable diffusion models will be significantly more advanced than for almost all other fields.
- For many tasks, evaluation metrics will be whatever you are familiar with
  - E.g. for segmentation intersection over union
  - For imputation RMSE

Evaluation

Stable Diffusion Models – qualitative methods

- Involves human assessment of generated images
- Quality is measured across a range of aspects:
  - Compositionality
  - Image-text alignment
  - Spatial relations
- Outputs are measured for common prompts of varying degrees of difficulty:
  - DrawBench
  - PartiPrompts

Evaluation

Stable Diffusion Models – quantitative methods

- Text-guided
  - CLIP score – measures the compatibility of image-caption pairs
  - Semantic similarity between image and caption
  - Higher CLIP score is better
  - CLIP score is highly correlated with human judgement
- Image-conditioned
  - Generate an image with a prompt (e.g. “A picture of a majestic Tonkinese cat.”)
  - Feed image into the model with a prompt (e.g. “Make the cat into a samurai.”) to produce image A.
  - Feed image into the model with a second prompt (e.g. “Make the cat into a businesscat.”) to produce image B
  - We then look at the change in the CLIP score between image A and B and between the change in the two captions
  - The higher the better
  - We can also measure the similarity between the original image, and the changed image

Evaluation

Stable Diffusion Models – quantitative methods

- Class-conditioned models are pretrained on a class-labeled dataset.
  - Frechet Inception Distance (FID)
  - Kernel Inception Distance
  - Inception Score
- FID
  - Find the Frechet distance between Gaussians fitted to feature representations of Inception.
  - Use the Inception v3 model and cut off the final classification layer
  - Get the ImageNet dataset (or a subset of it)
  - Generate a bunch of images
  - Stuff both generated and ImageNet images into Inception and get the feature representations
  - Fit two Gaussians to the representations
  - Compute the Frechet Distance.