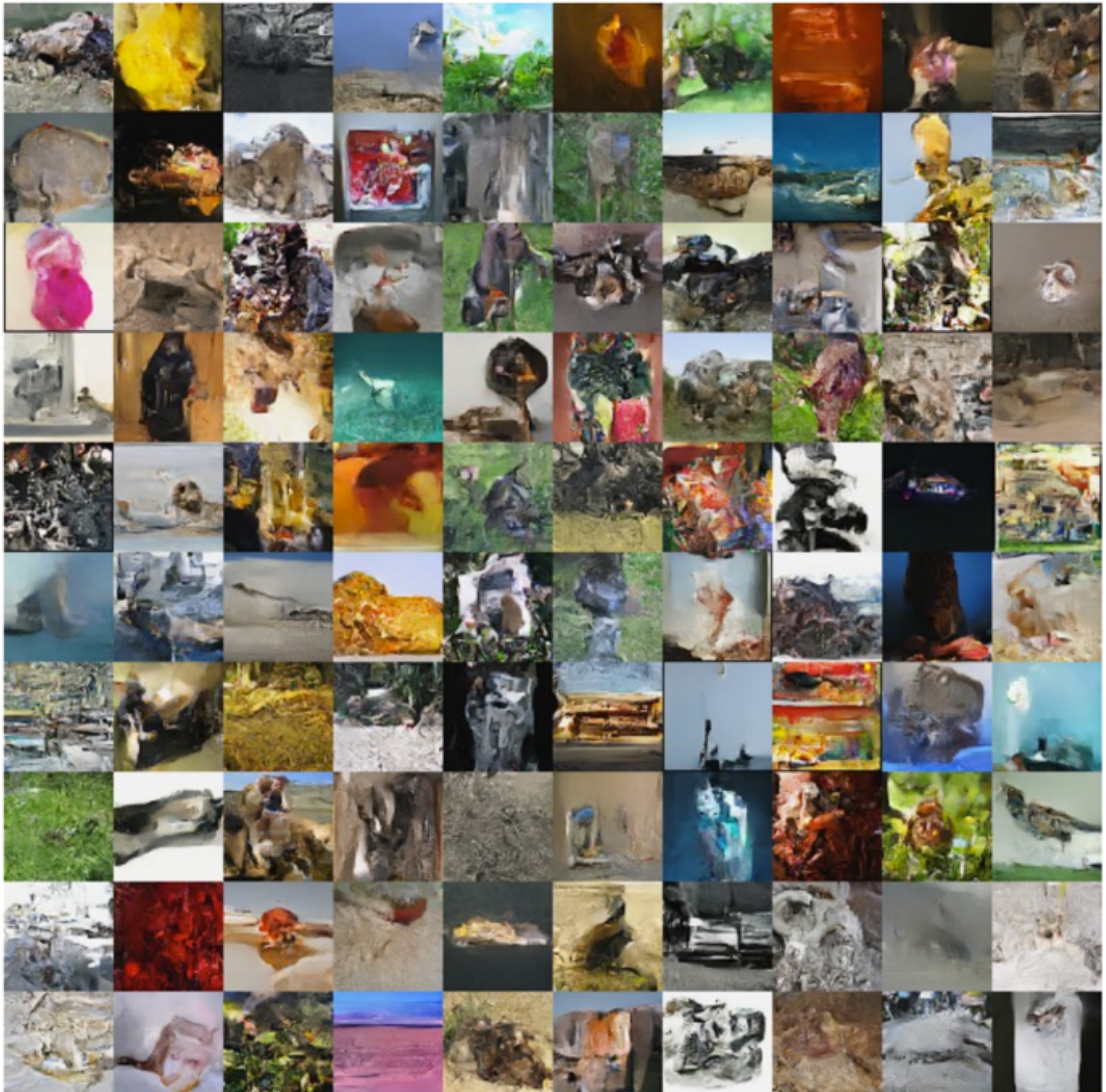
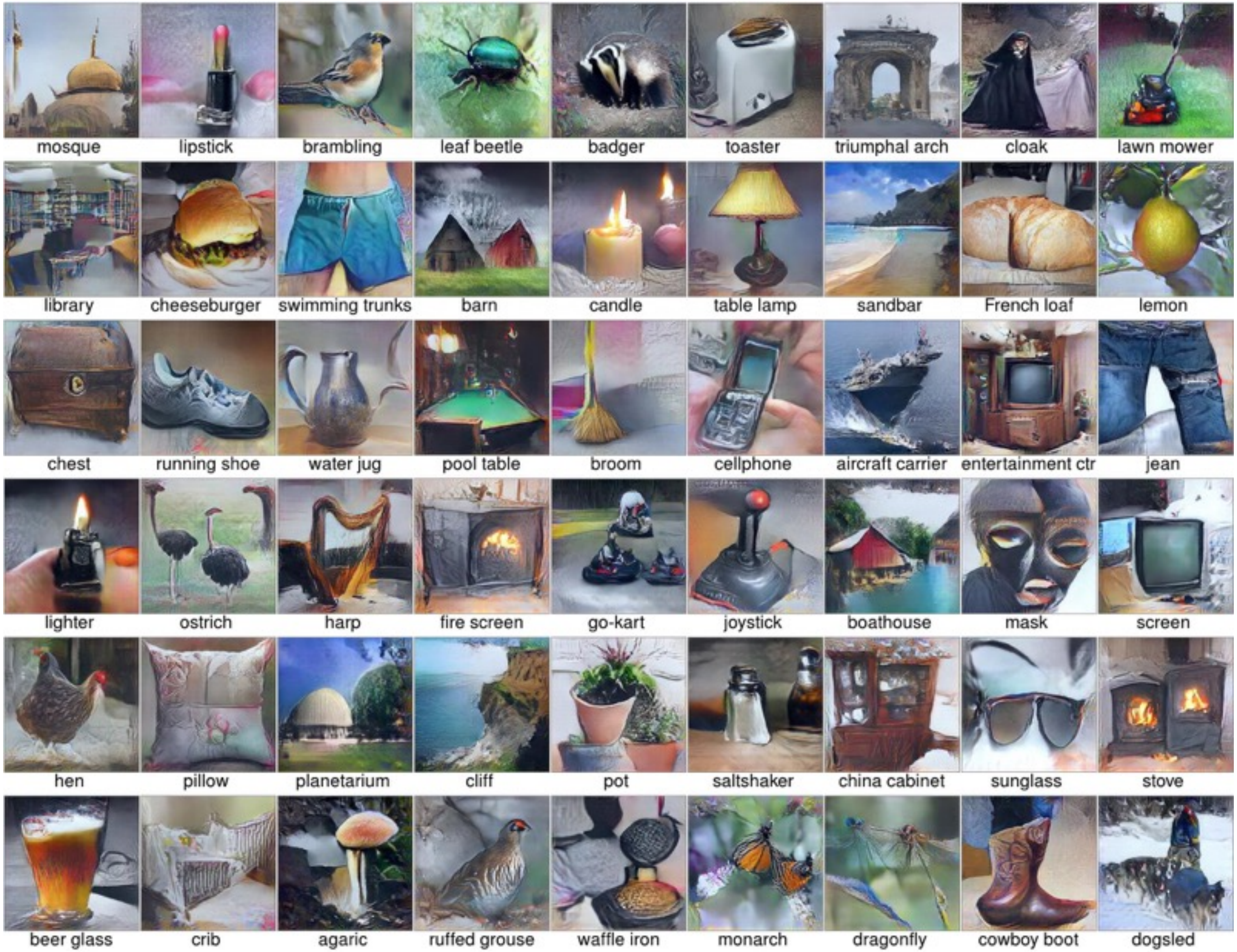
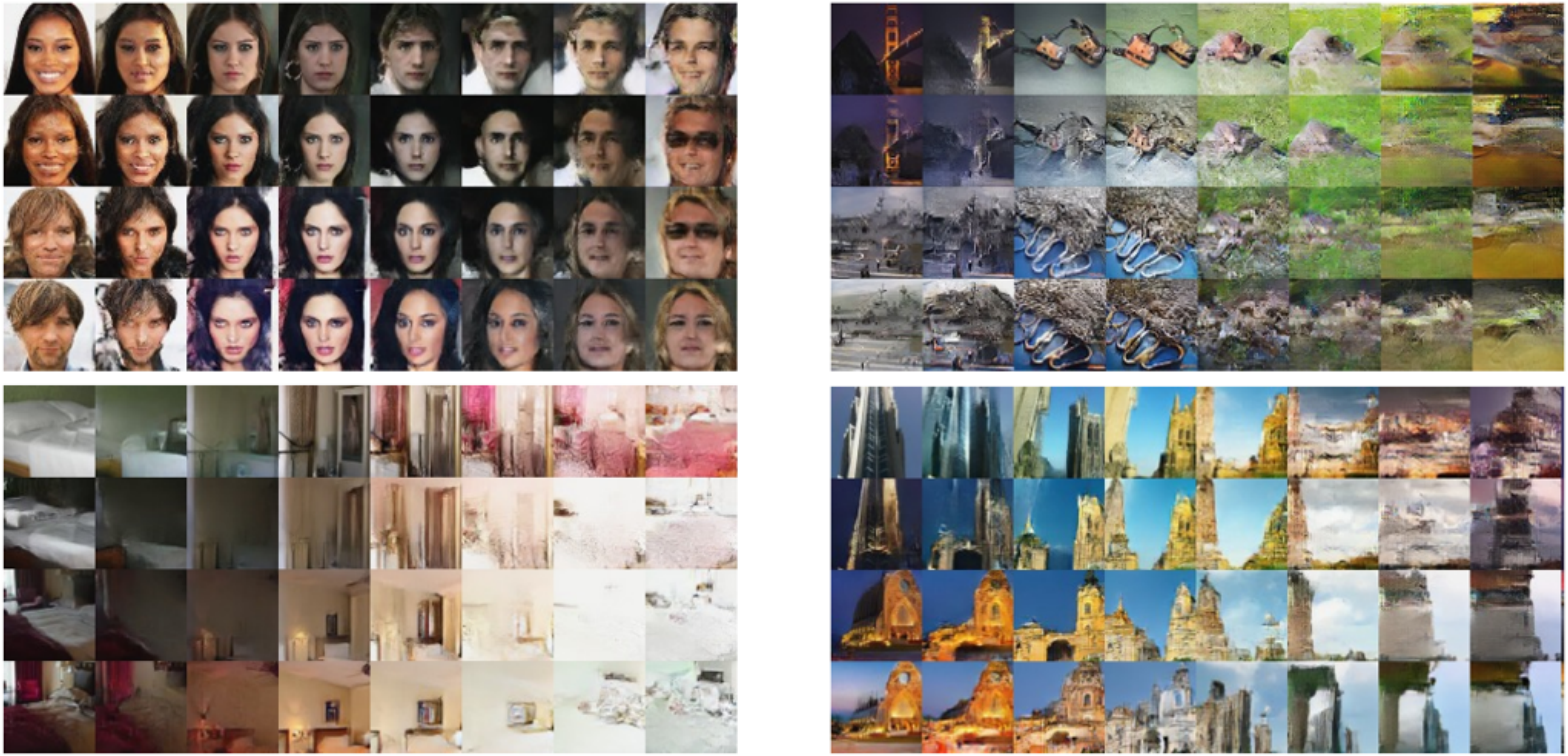
CSC2541: Differentiable Inference and Generative Models



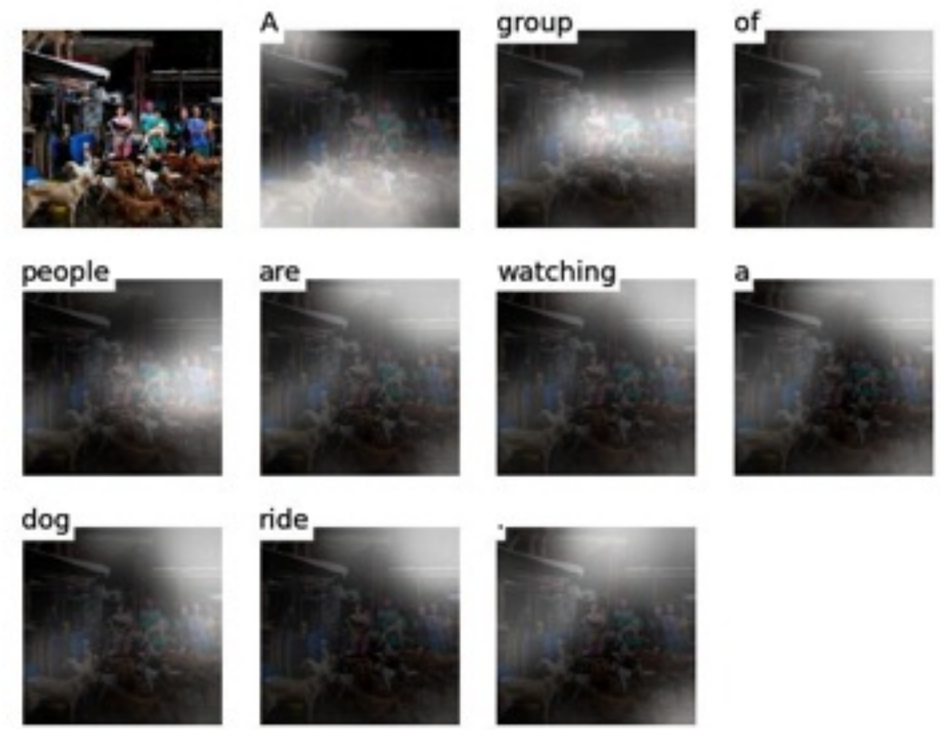
Density estimation using Real NVP. Ding et al, 2016



Nguyen A, Dosovitskiy A, Yosinski J, Brox T, Clune J (2016). *Synthesizing the preferred inputs for neurons in neural networks via deep generator networks.* Advances in Neural Information Processing Systems 29



Density estimation using Real NVP. Ding et al, 2016



A group of people are watching a dog ride

(Jamie Kyros)

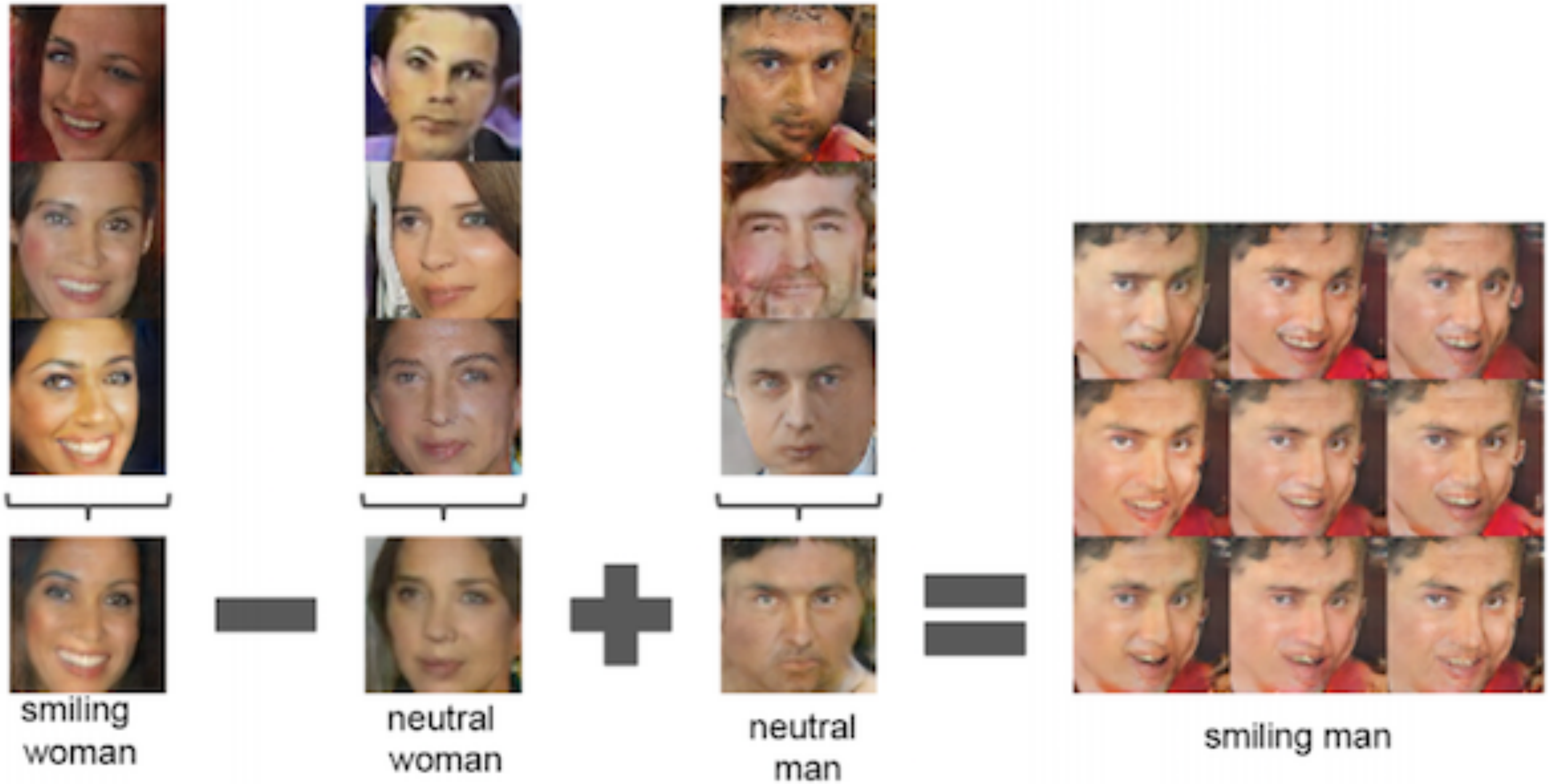
Pixel Recurrent Neural Networks
Aaron van den Oord, Nal Kalchbrenner, Koray Kavukcuoglu

Types of Generative Models

- Conditional probabilistic models
- Latent-variable probabilistic models
- GANs
- Invertible models

Advantages of latent variable models

- Model checking by sampling
- Natural way to specify models
- Compact representations
- Semi-Supervised learning
- Understanding factors of variation in data

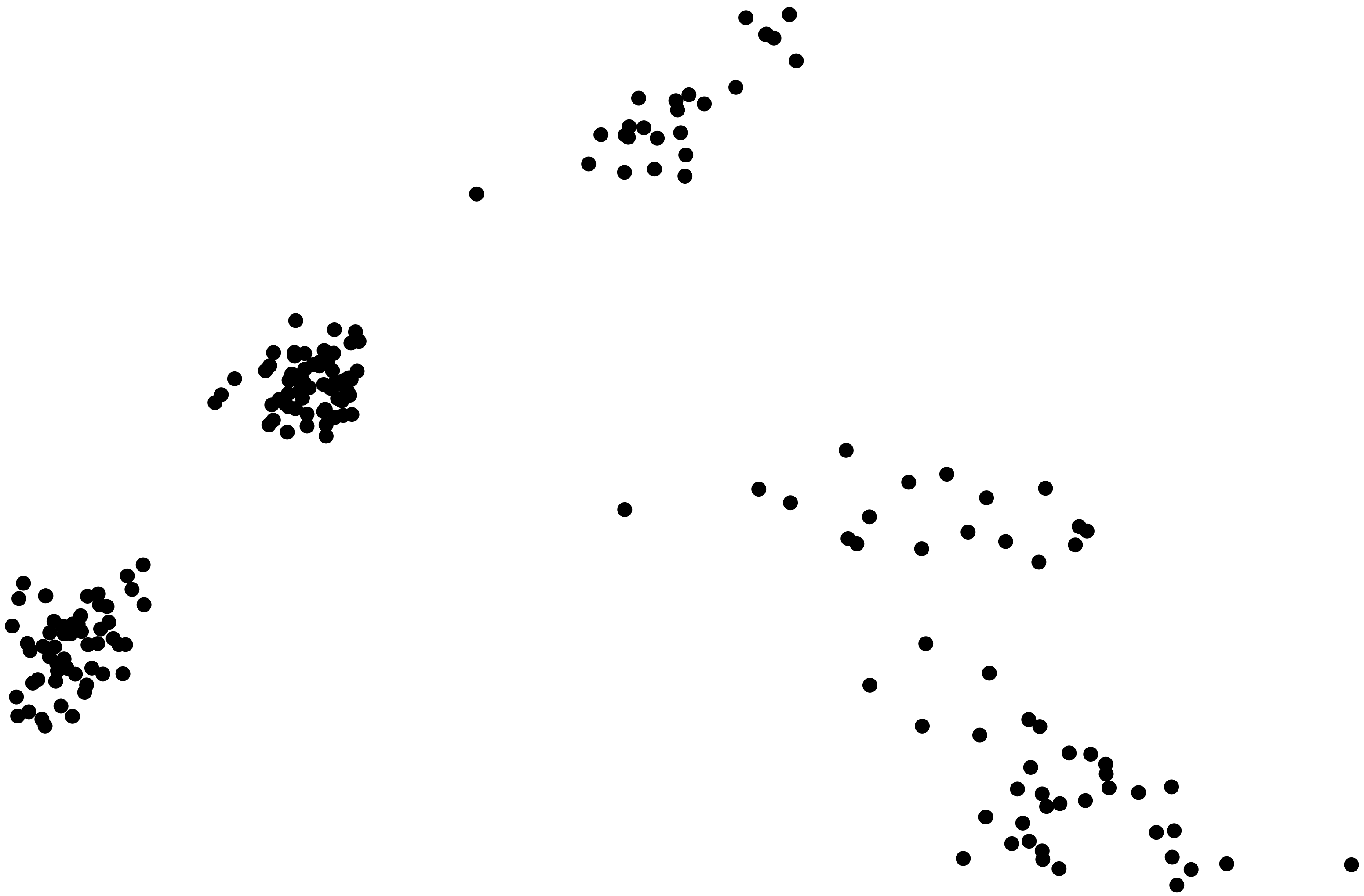


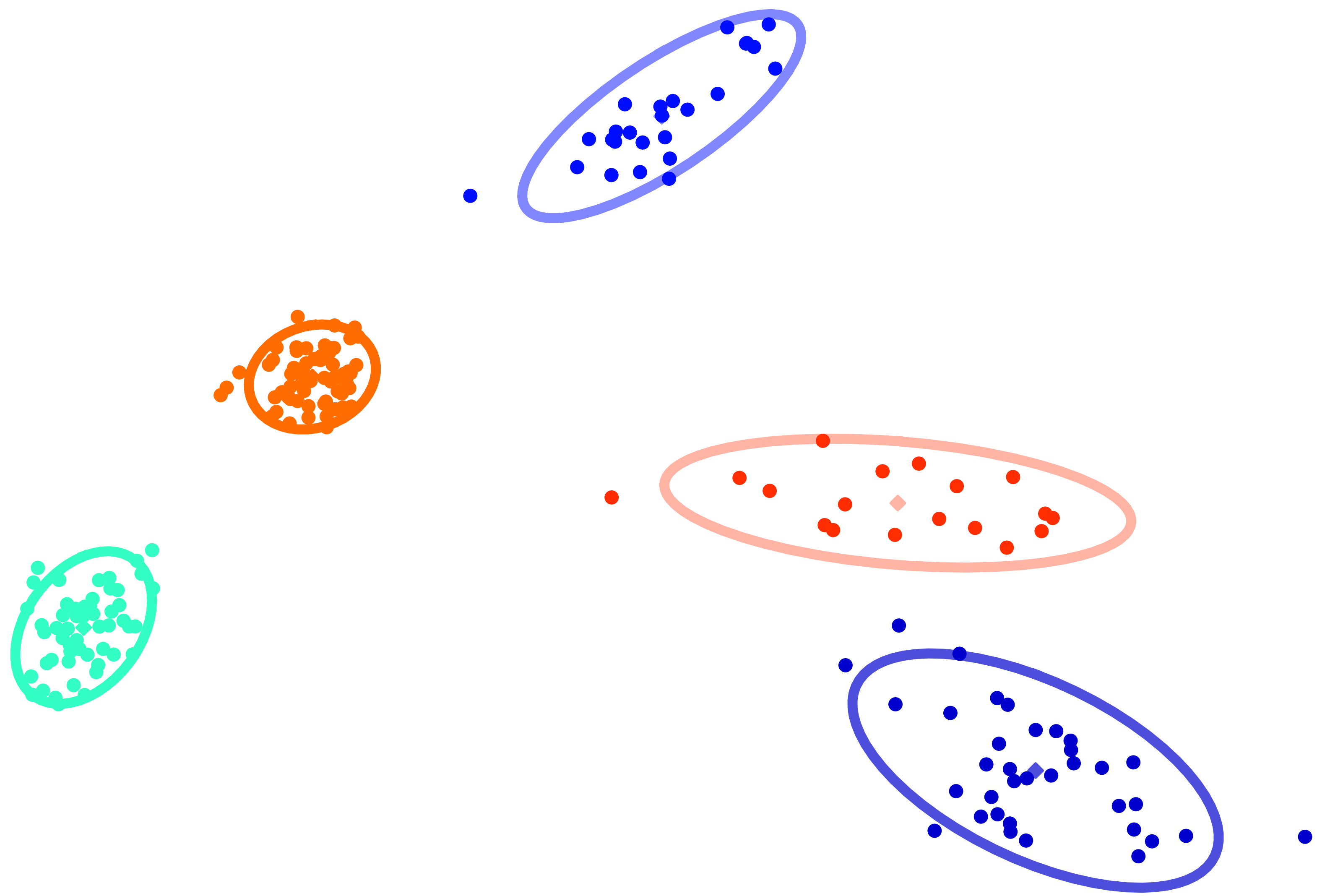
Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks Alec Radford, Luke Metz, Soumith Chintala

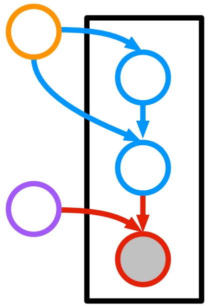
Advantages of *probabilistic* latent-variable models

- **Data-efficient learning** - automatic regularization, can take advantage of more information
- **Compose models** - e.g. incorporate data corruption model. Different from composing feedforward computations
- **Handle missing data** (without the standard hack of just guessing the missing values using averages).
- **Predictive uncertainty** - necessary for decision-making
- **conditional predictions** (e.g. if brexit happens, the value of the pound will fall)
- **Active learning** - what data would be expected to increase our confidence about a prediction

- Cons:
  - intractable integral over latent variables
- Examples: medical diagnosis, image modeling

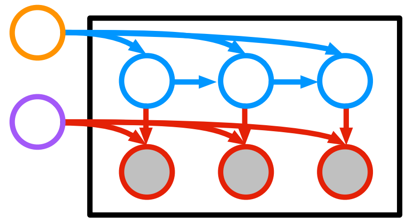






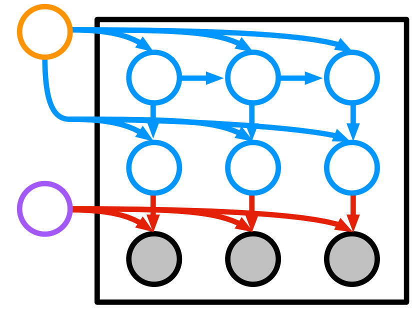
[1]

Gaussian mixture model



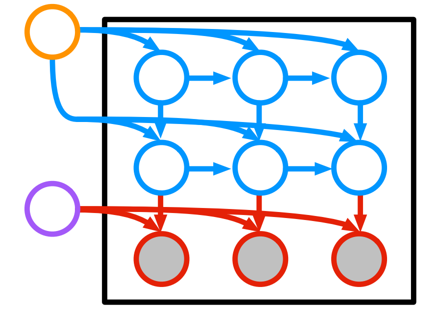
[2]

Linear dynamical system



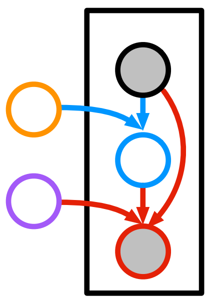
[3]

Hidden Markov model



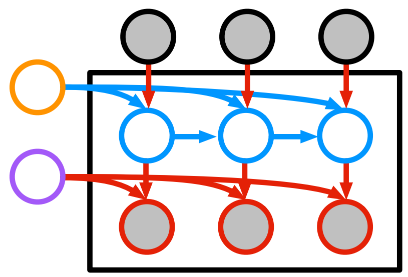
[4]

Switching LDS



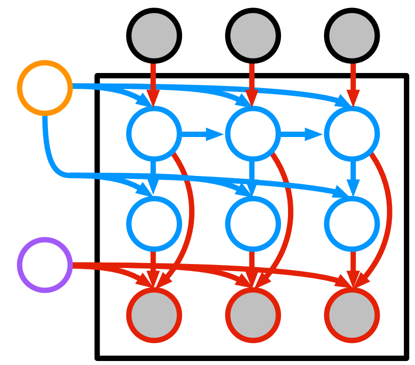
[5]

Mixture of Experts



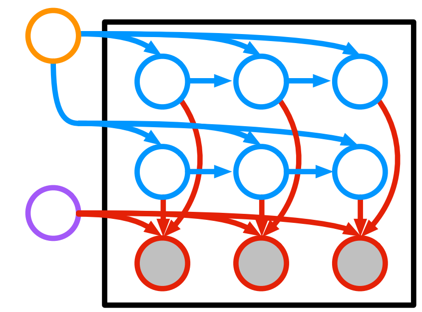
[2]

Driven LDS



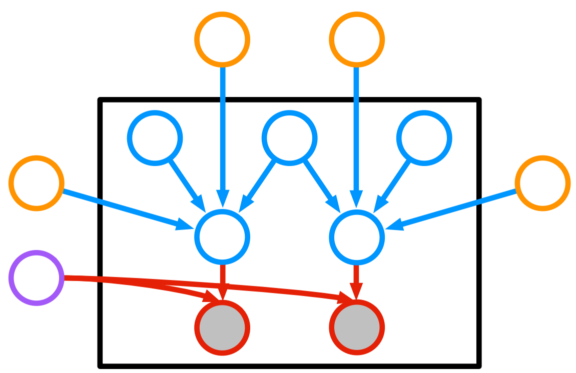
[6]

IO-HMM



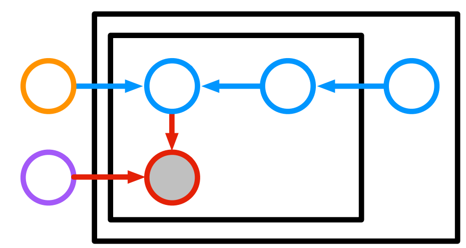
[7]

Factorial HMM



[8,9]

Canonical correlations analysis



[10]

admixture / LDA / NMF

[1] Palmer, Wipf, Kreutz-Delgado, and Rao. Variational EM algorithms for non-Gaussian latent variable models. NIPS 2005.
[2] Ghahramani and Beal. Propagation algorithms for variational Bayesian learning. NIPS 2001.
[3] Beal. Variational algorithms for approximate Bayesian inference, Ch. 3. U of London Ph.D. Thesis 2003.
[4] Ghahramani and Hinton. Variational learning for switching state-space models. Neural Computation 2000.
[5] Jordan and Jacobs. Hierarchical Mixtures of Experts and the EM algorithm. Neural Computation 1994.
[6] Bengio and Frasconi. An Input Output HMM Architecture. NIPS 1995.
[7] Ghahramani and Jordan. Factorial Hidden Markov Models. Machine Learning 1997.
[8] Bach and Jordan. A probabilistic interpretation of Canonical Correlation Analysis. Tech. Report 2005.
[9] Archambeau and Bach. Sparse probabilistic projections. NIPS 2008.
[10] Hoffman, Bach, Blei. Online learning for Latent Dirichlet Allocation. NIPS 2010.

Courtesy of Matthew Johnson

Differentiable models

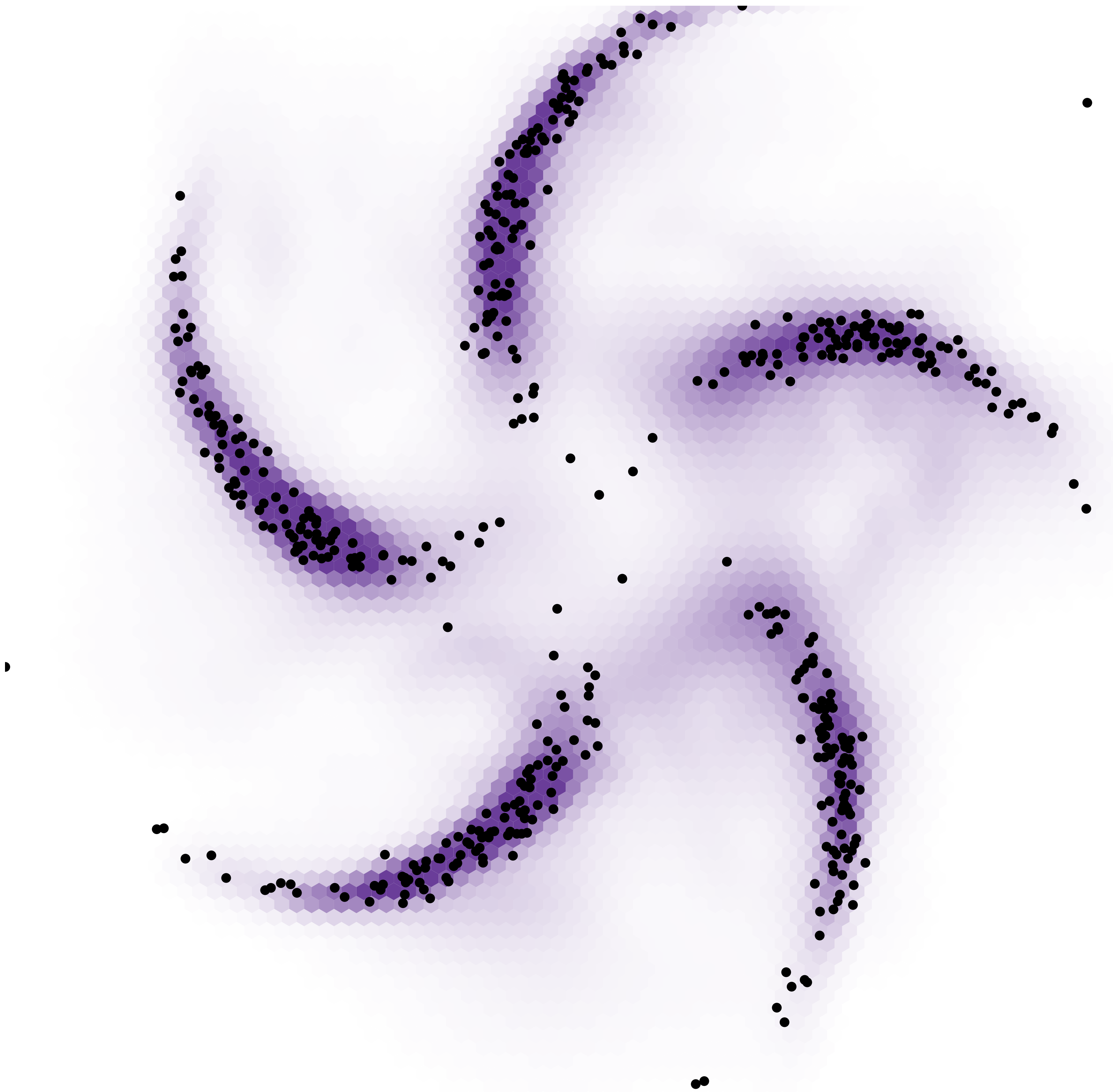
- Model distributions implicitly by a variable pushed through a deep net:

$$y = f_\theta(x)$$

- Approximate intractable distribution by a tractable distribution parameterized by a deep net:

$$p(y|x) = \mathcal{N}(y|\mu = f_\theta(x), \Sigma = g_\theta(x))$$

- Optimize all parameters using stochastic gradient descent

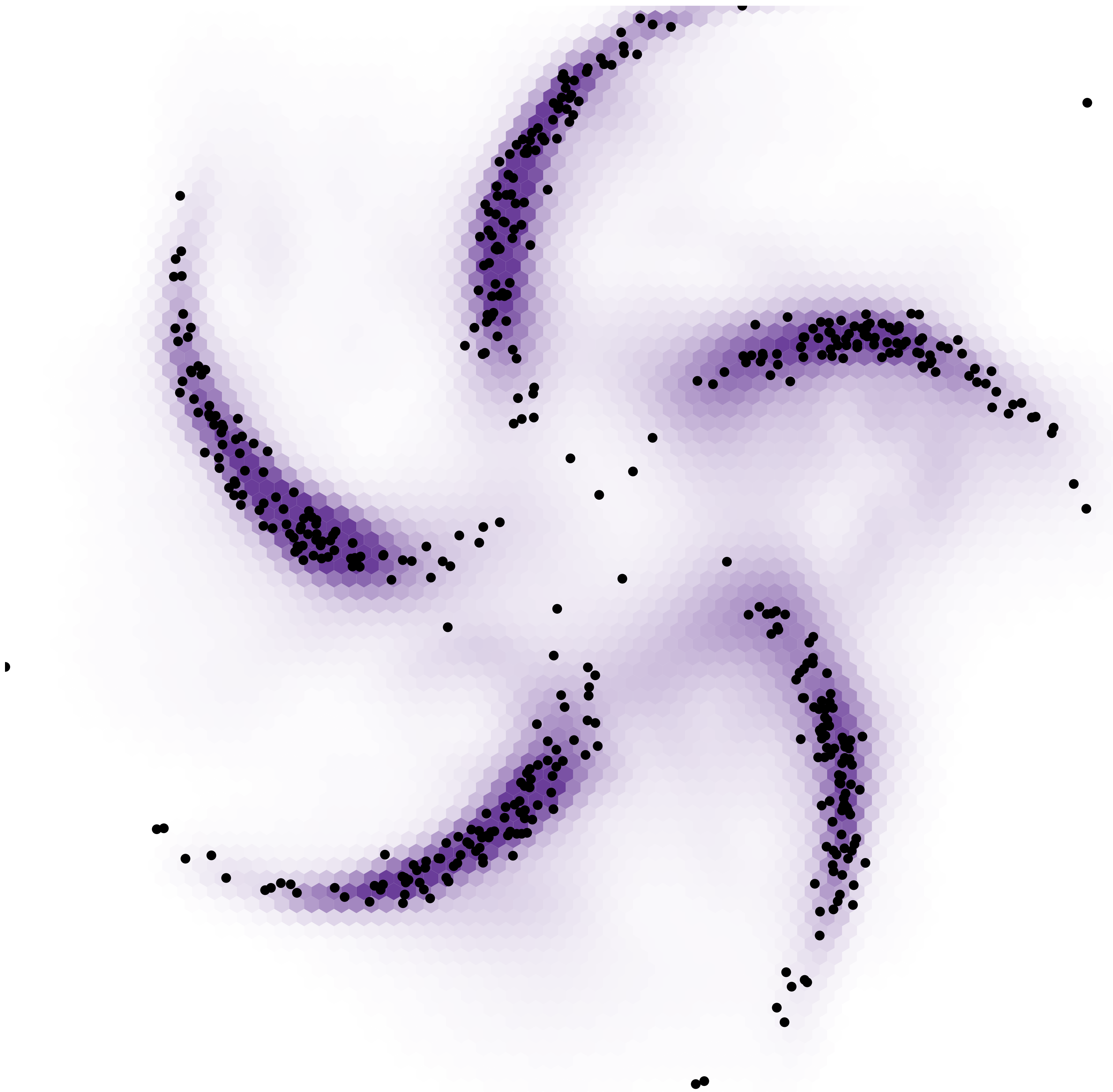


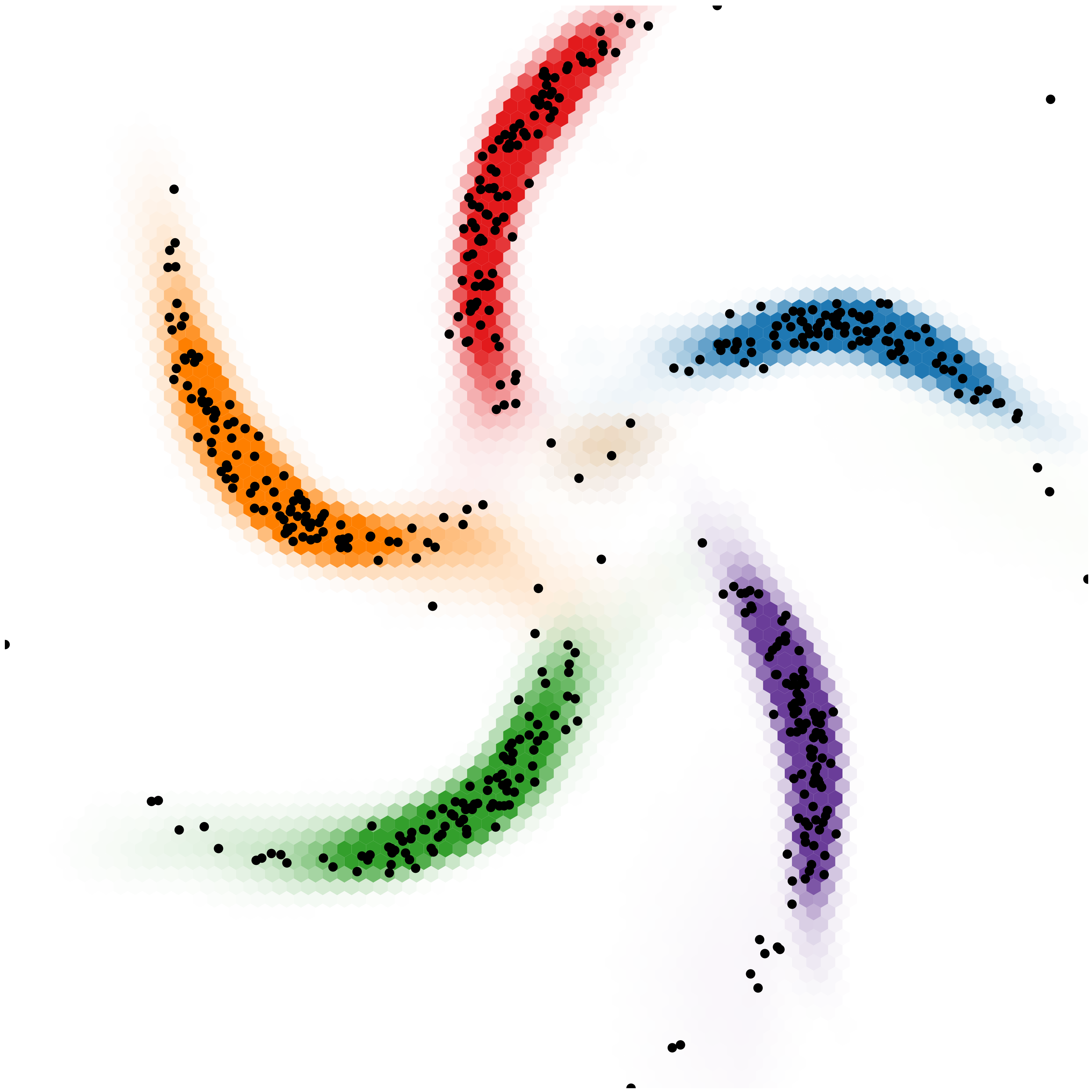
Probabilistic graphical models

+ structured representations
+ priors and uncertainty
+ data and computational efficiency
– rigid assumptions may not fit
– feature engineering
– top-down inference

Deep learning

– neural net “goo”
– difficult parameterization
– can require lots of data
+ flexible
+ feature learning
+ recognition networks



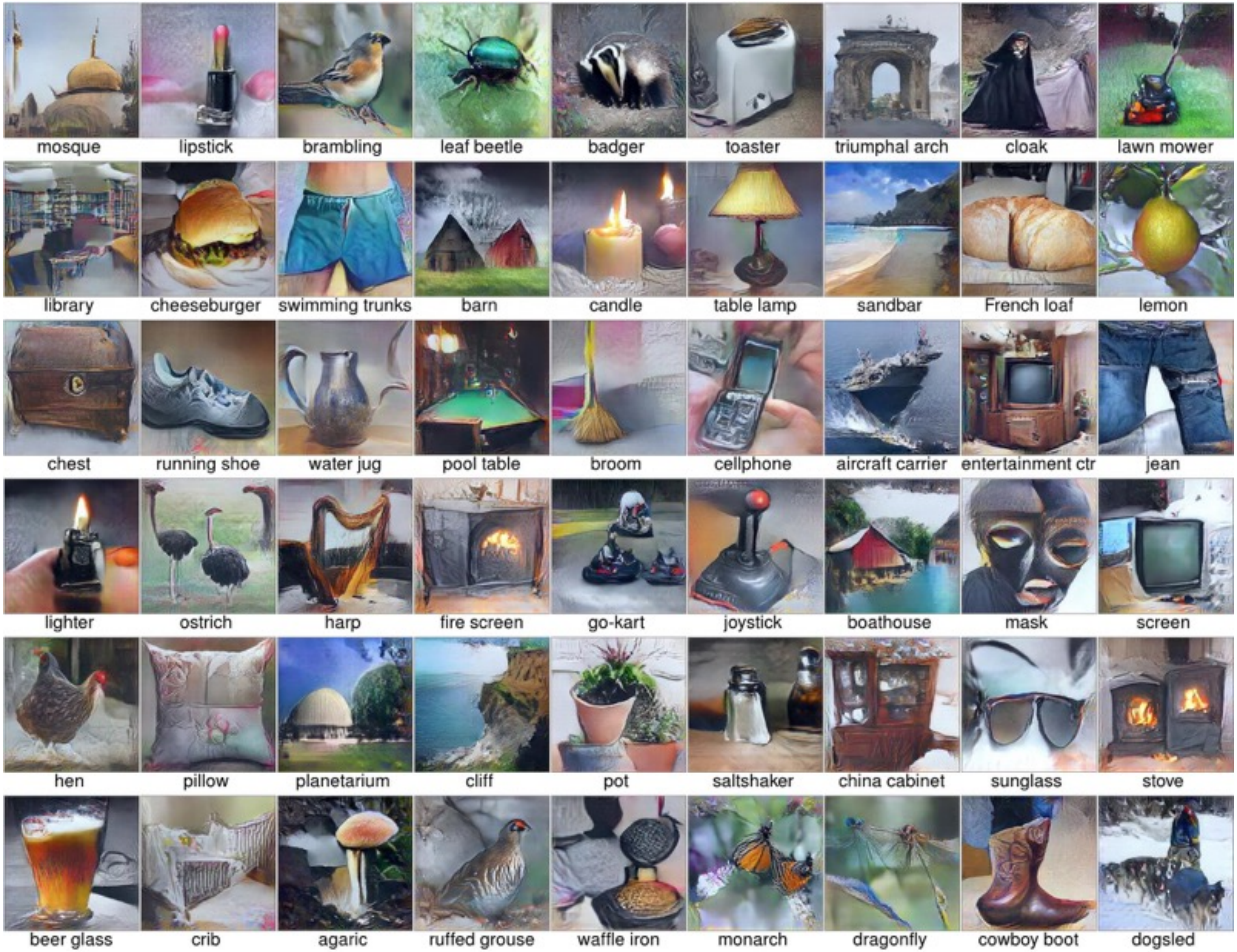


Machine-learning-centric History of Generative Models

- **1940s - 1960s** Motivating probability and Bayesian inference
- **1980s - 2000s** Bayesian machine learning with MCMC
- **1990s - 2000s** Graphical models with exact inference
- **1990s** - **present** Bayesian Nonparametrics with MCMC (Indian Buffet process, Chinese restaurant process)
- **1990s - 2000s** Bayesian ML with mean-field variational inference
- **1995** Helmholtz machine (*almost* invented variational autoencoders)
- **2000s - present** Probabilistic Programming
- **2000s - 2013** Deep undirected graphical models (RBMs, pretraining)
- **2010s** - **present** Stan - Bayesian Data Analysis with HMC
- **2000s** - **2013** Autoencoders, denoising autoencoders
- **2000s** - **present** Invertible density estimation
- **2013** - **present** Variational autoencoders
- **2014** - **present** Generative adversarial nets

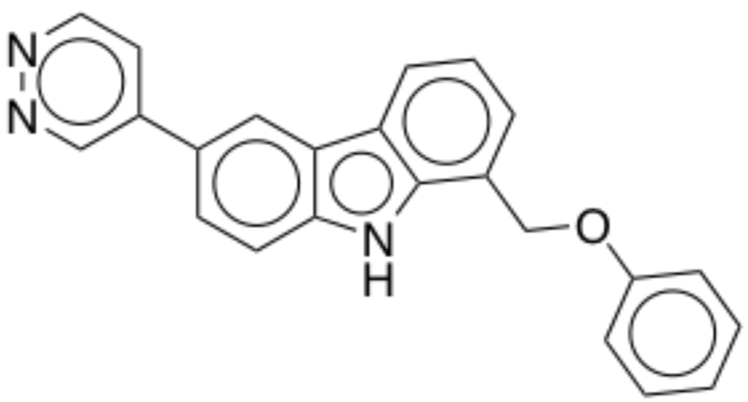
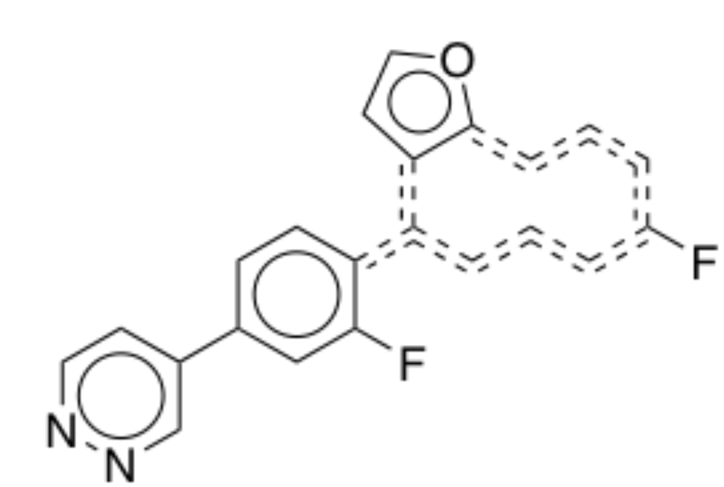
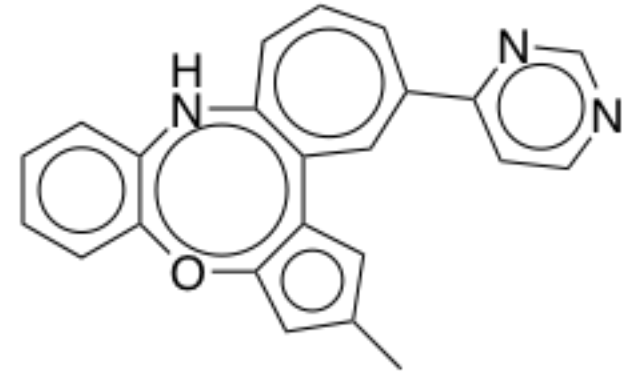
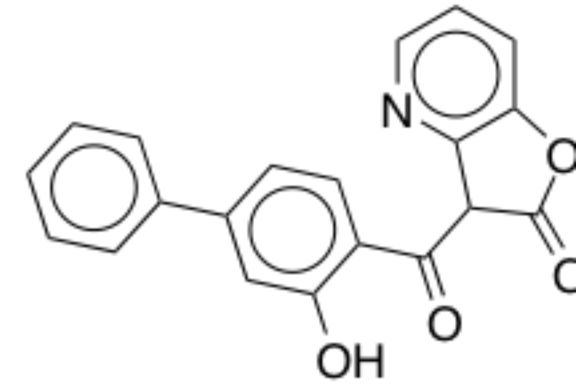
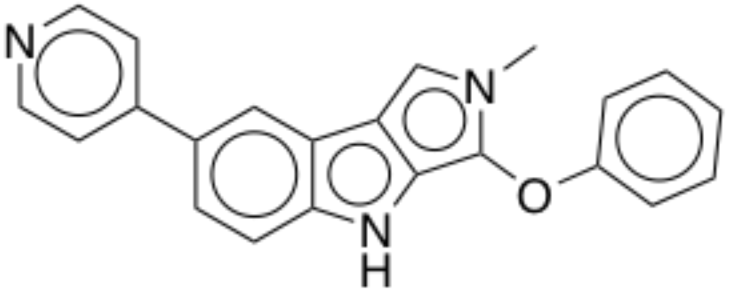
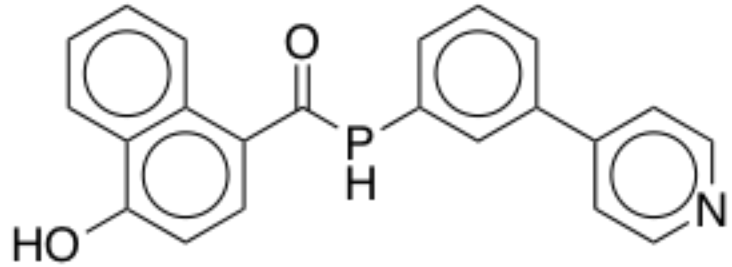
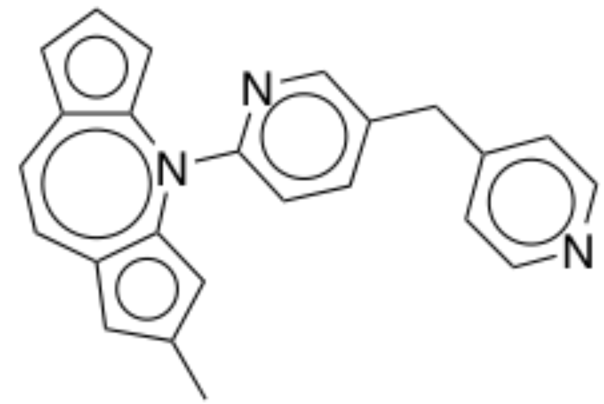
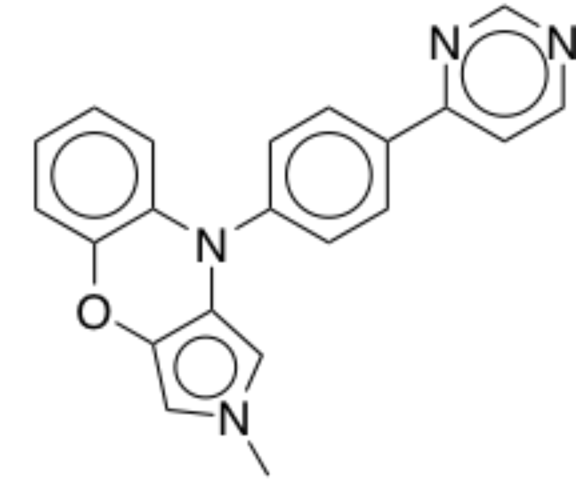
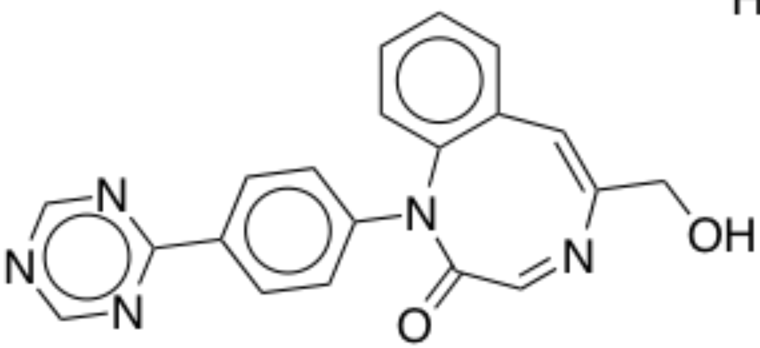
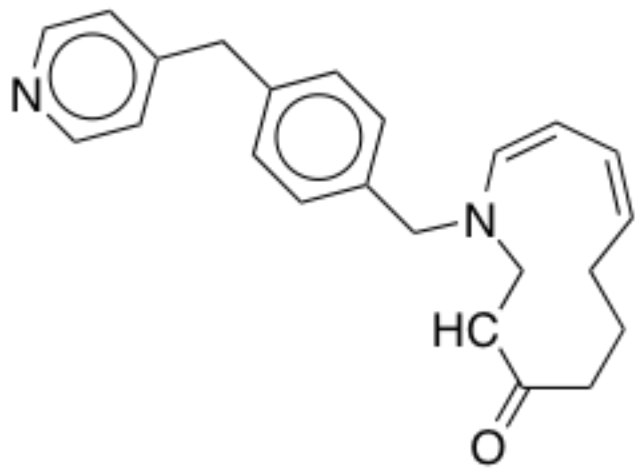
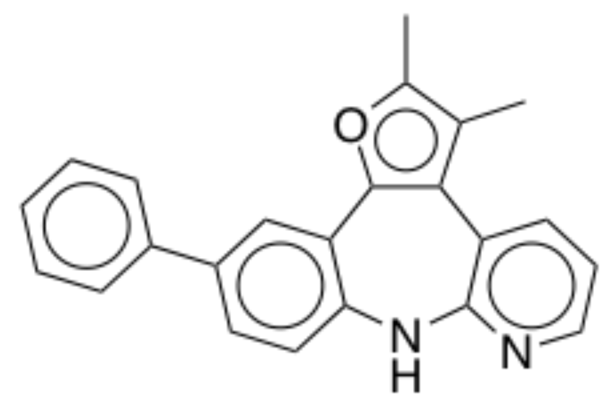
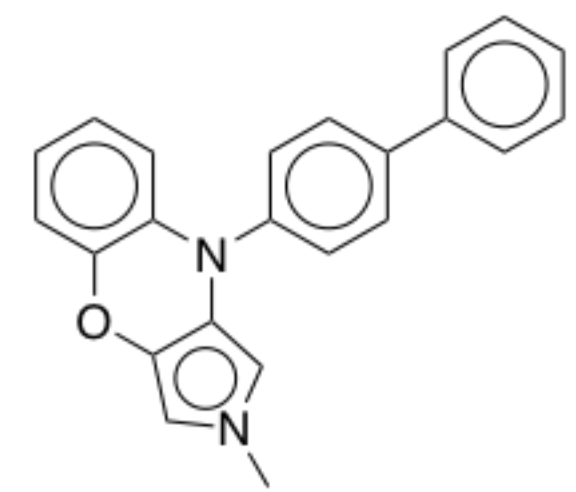
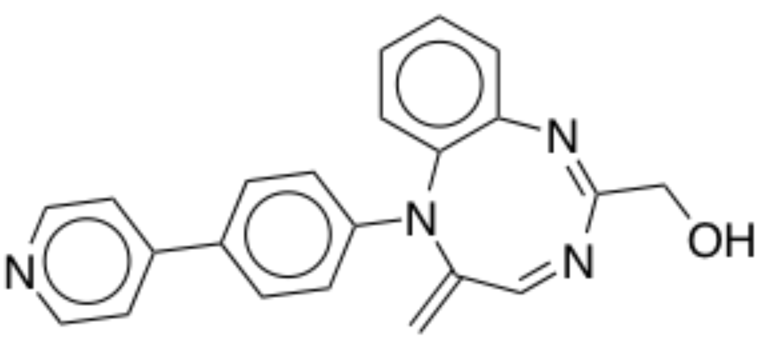
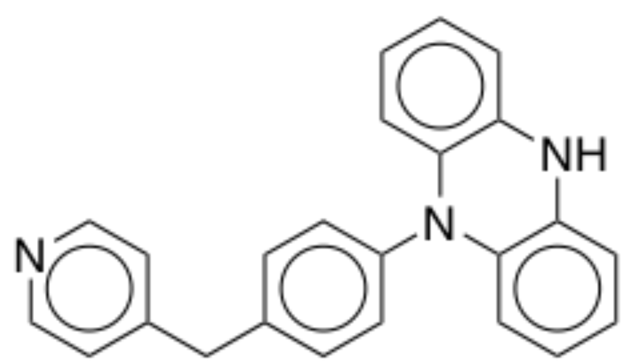
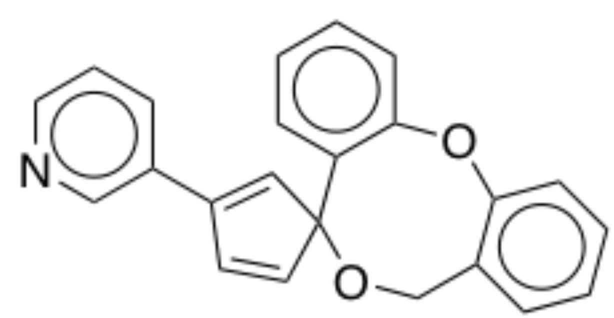
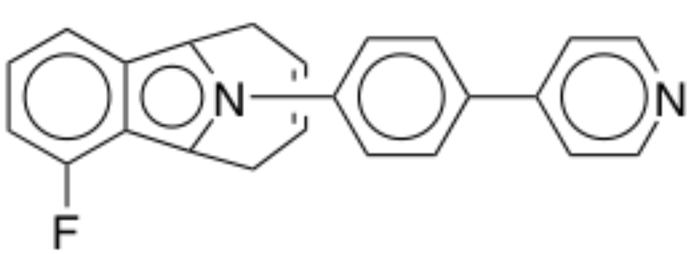
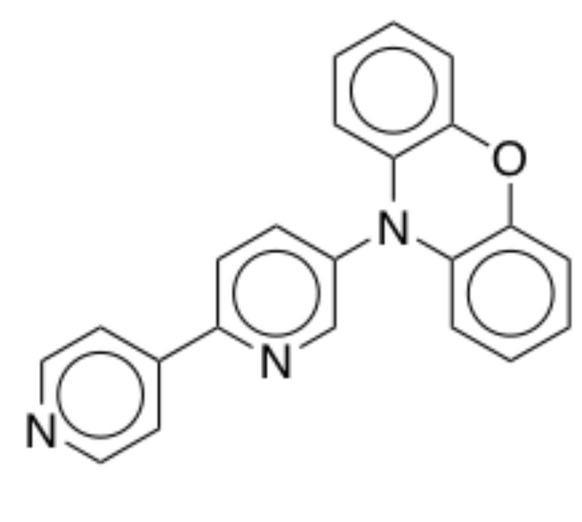
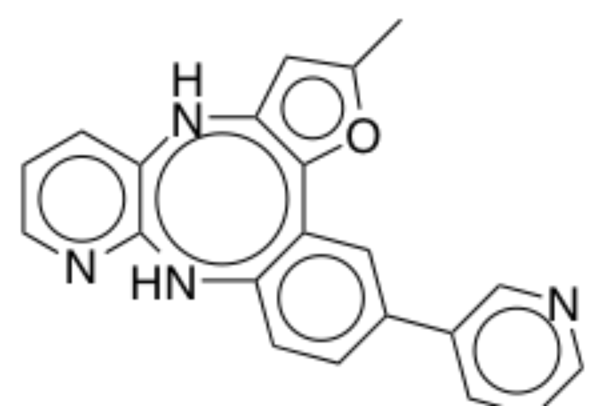
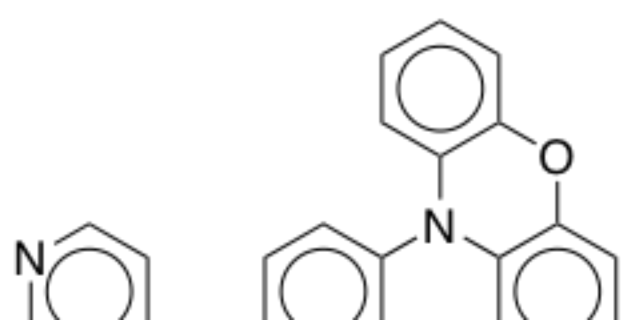
Frontiers

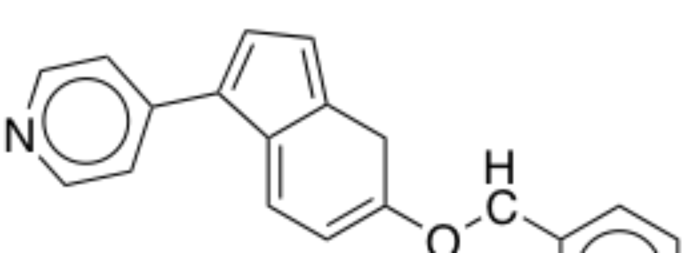
- Generate images given captions
- Generating large structures
  - images with consistent internal structure and not blurry
  - videos
  - long texts
- Discrete latent random variables
- Generate complex discrete structures
- Time-series models for reinforcement learning



Nguyen A, Dosovitskiy A, Yosinski J, Brox T, Clune J (2016). *Synthesizing the preferred inputs for neurons in neural networks via deep generator networks.* Advances in Neural Information Processing Systems 29

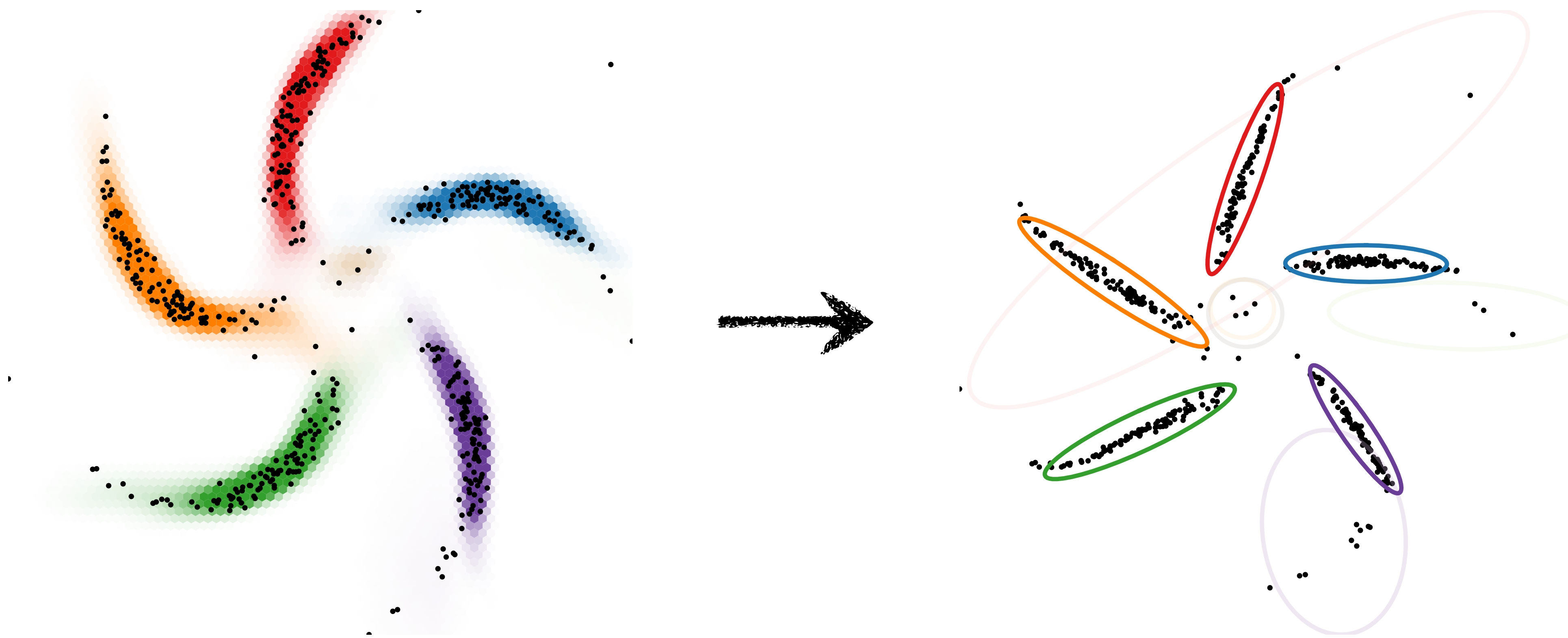
Density estimation using Real NVP. Ding et al, 2016

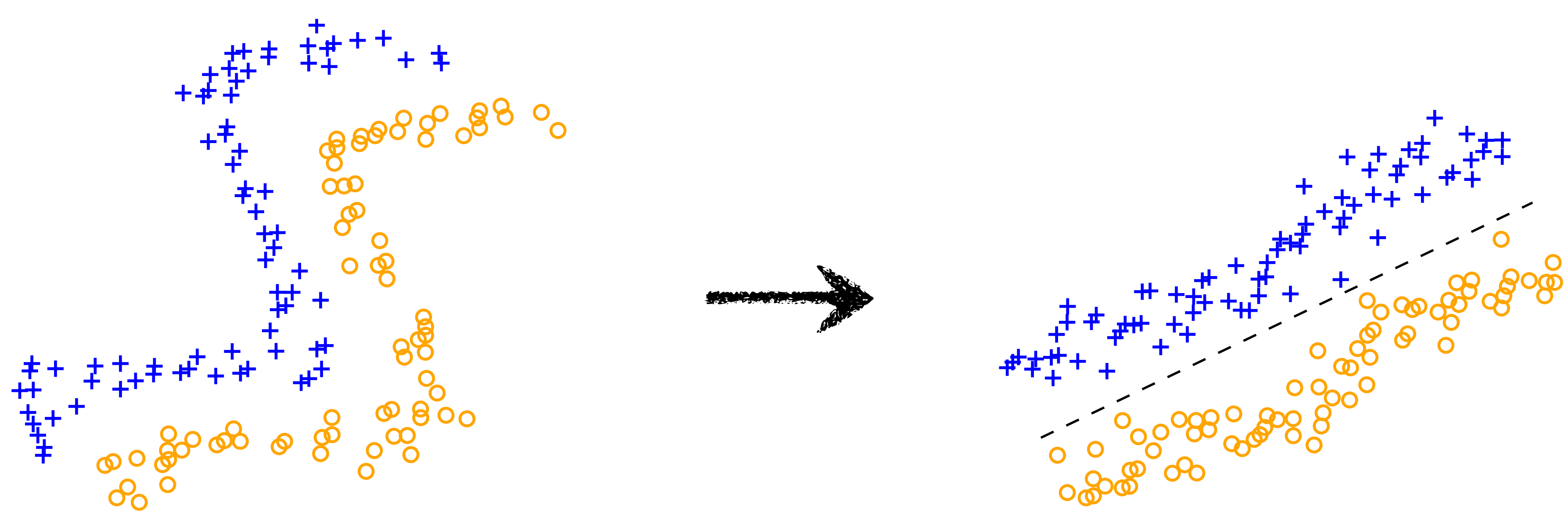


Modeling idea: graphical models on latent variables,
neural network models for observations

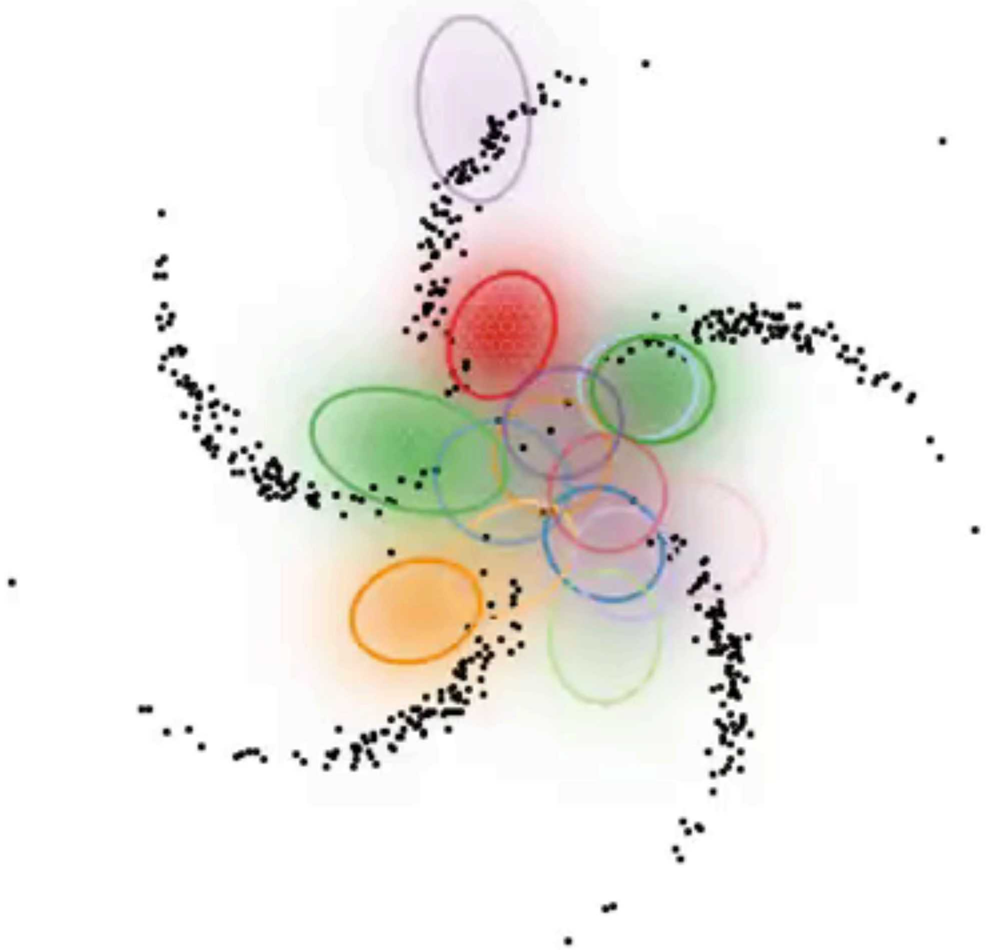
Composing graphical models with neural networks for structured representations and fast inference. Johnson, Duvenaud, Wiltschko, Datta, Adams, NIPS 2016 unsupervised learning



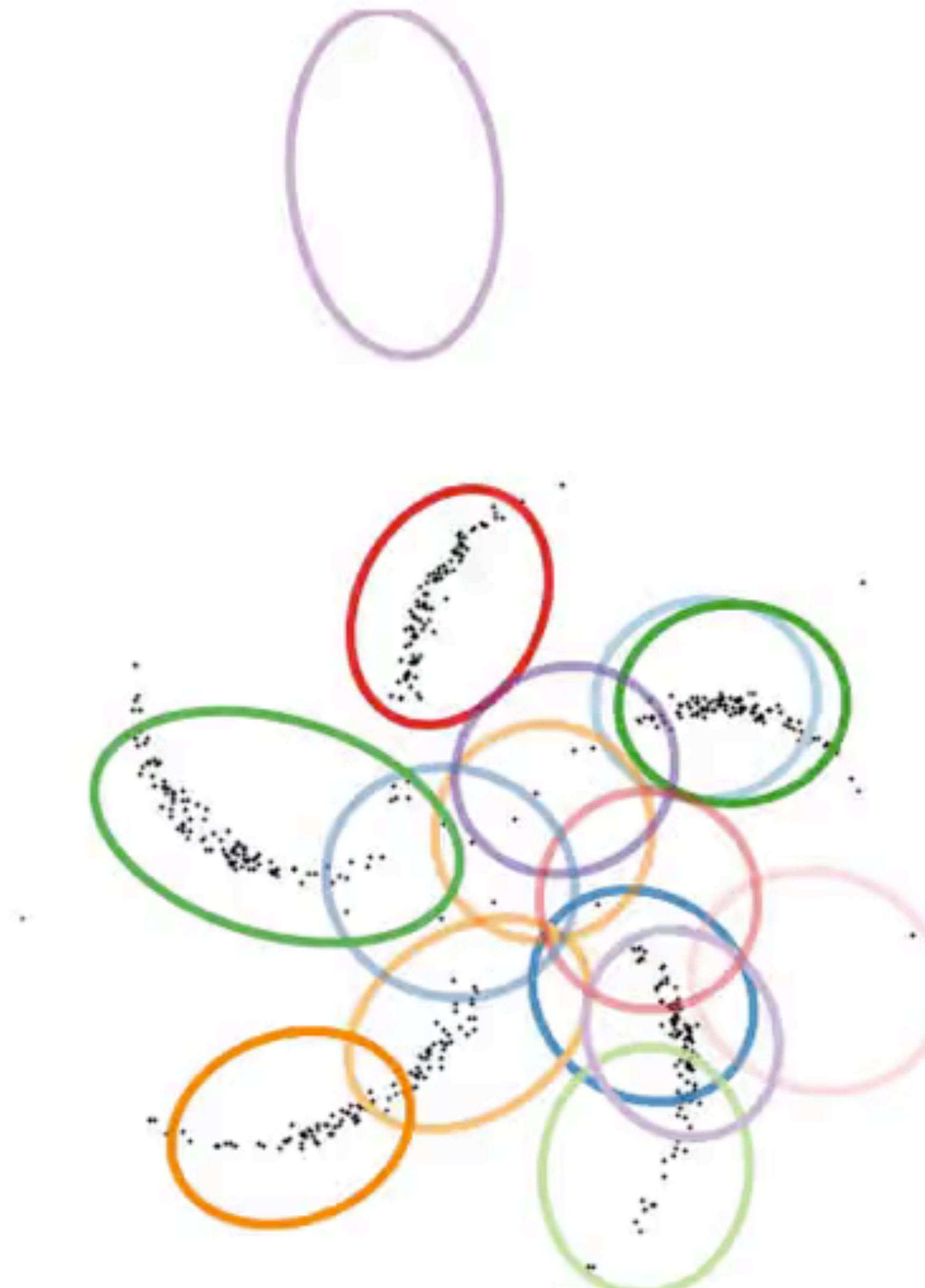
supervised learning



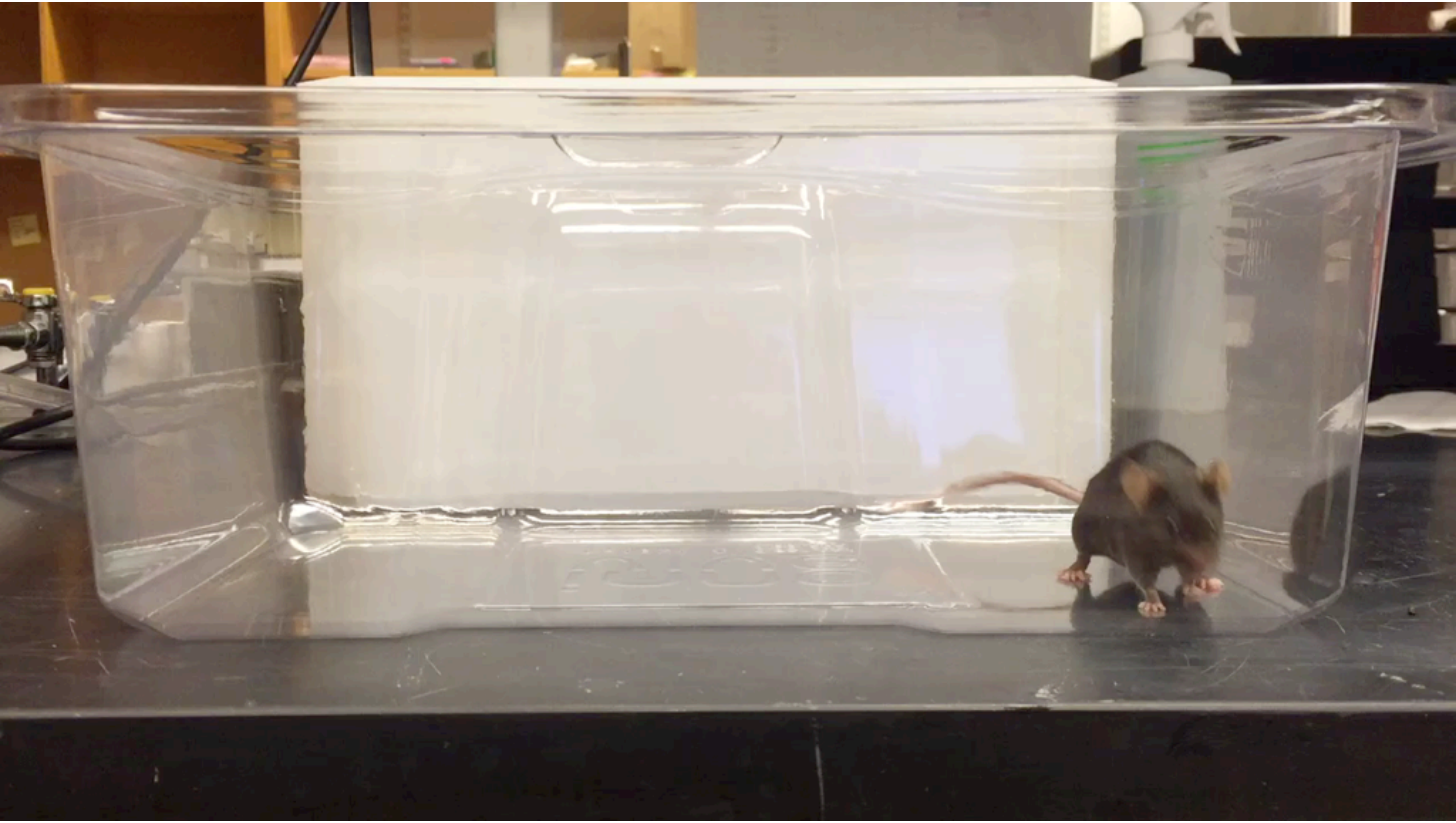
Courtesy of Matthew Johnson



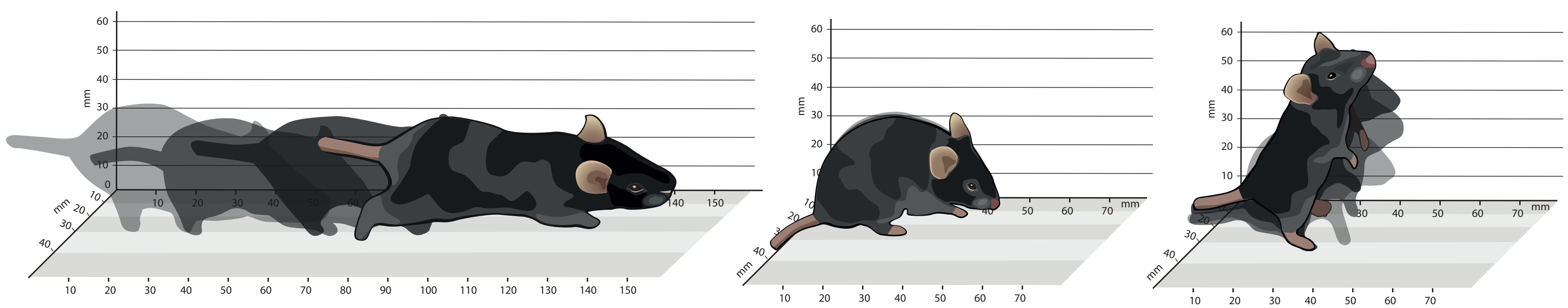
data space

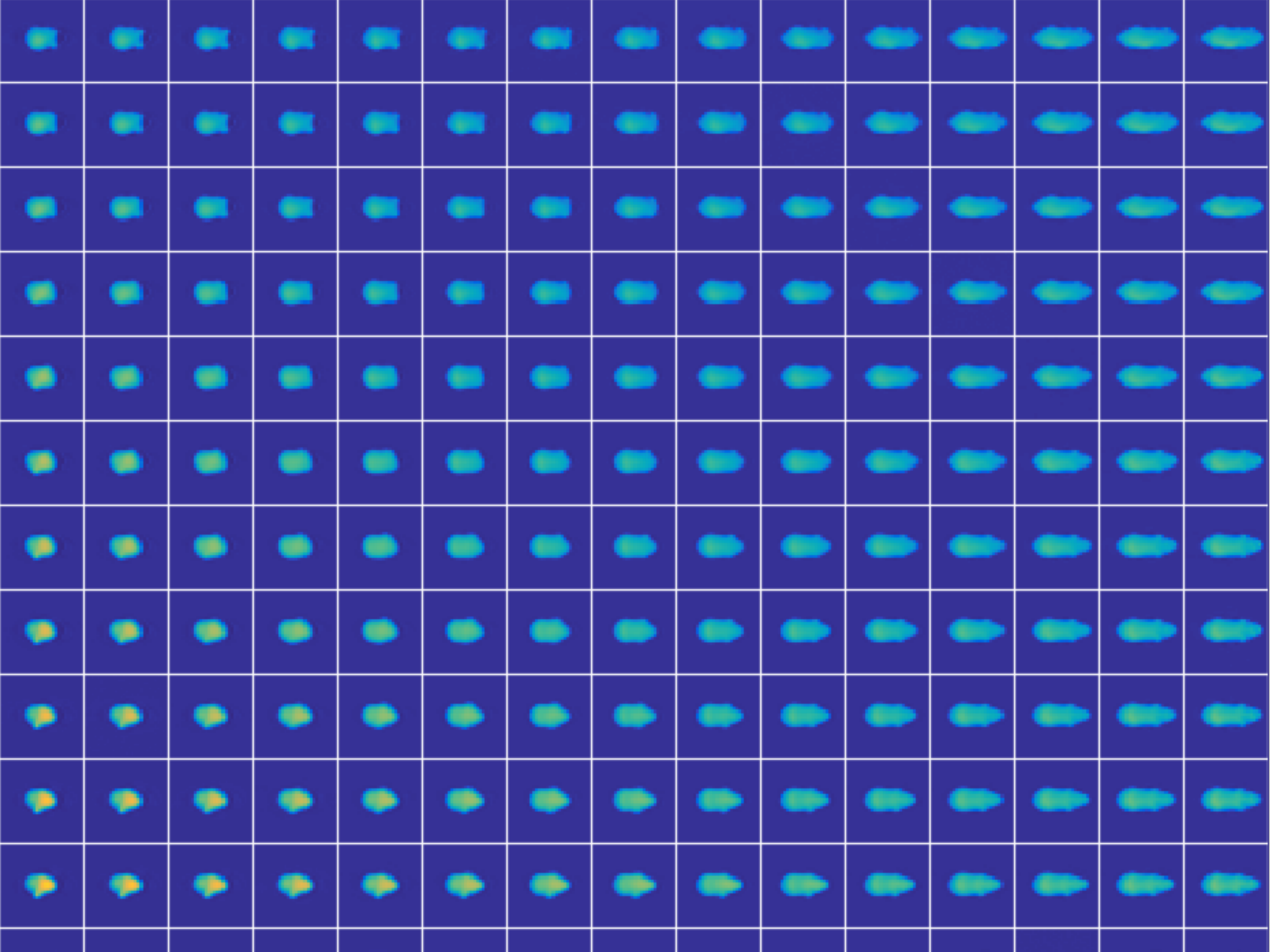


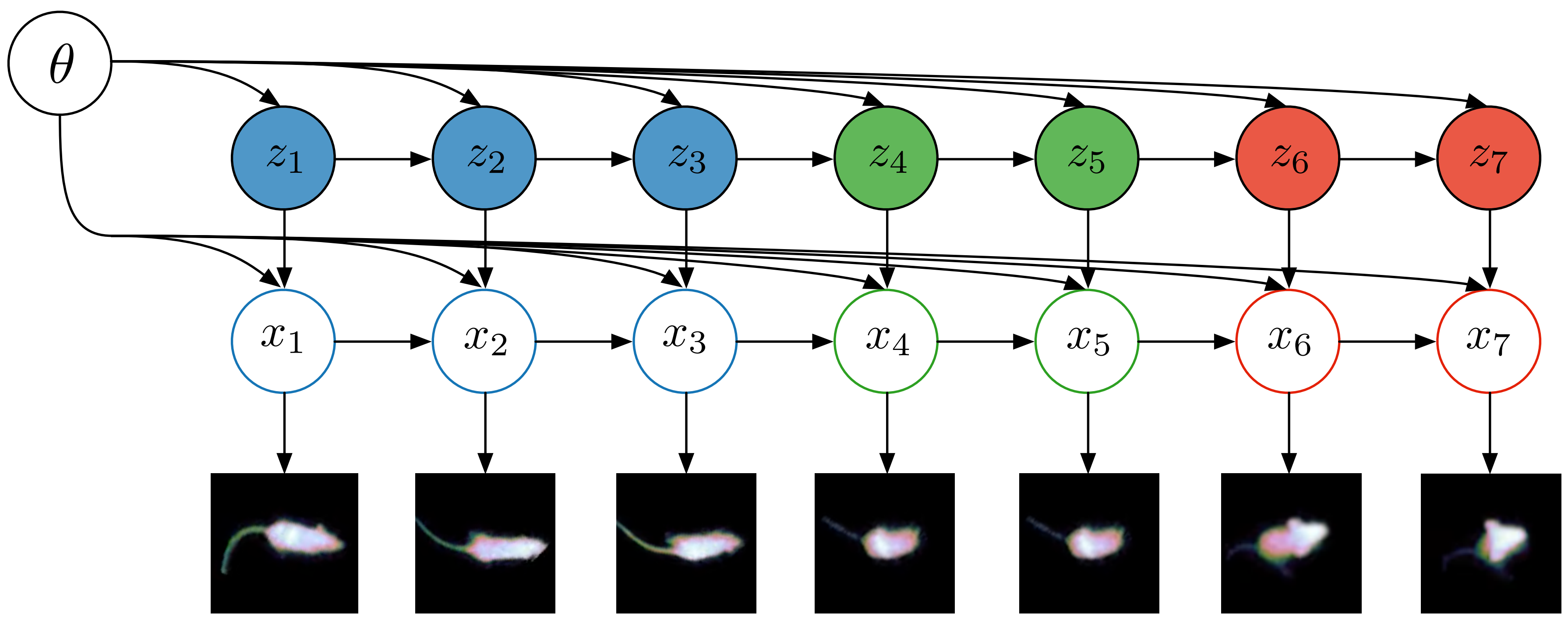
latent space

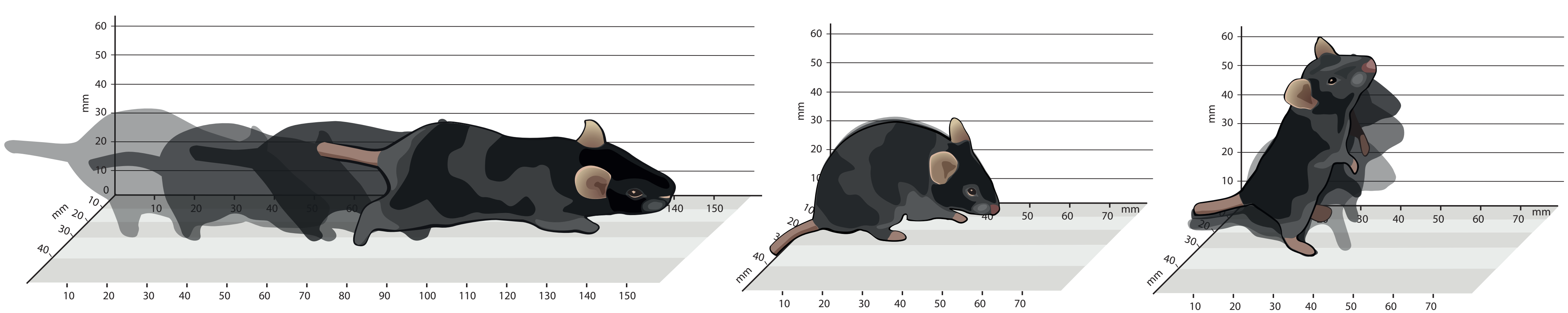


Application: learn syllable representation of behavior from video

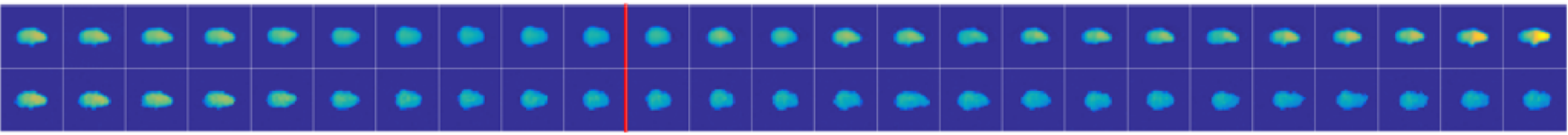


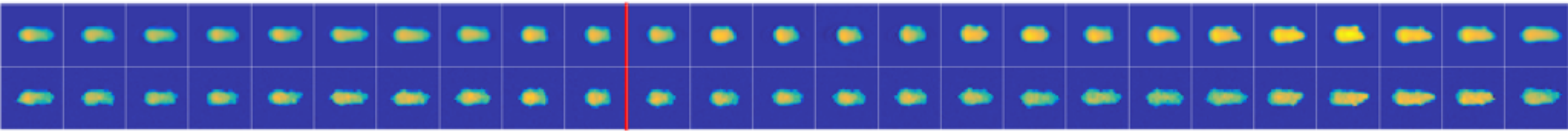


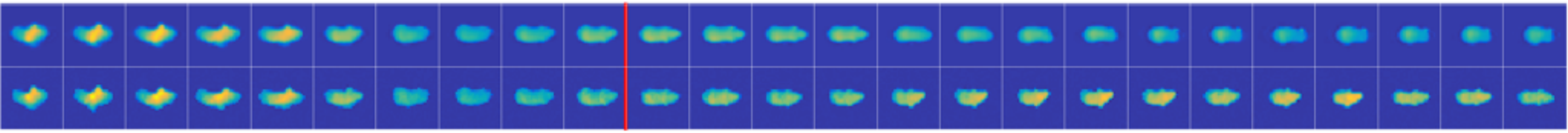


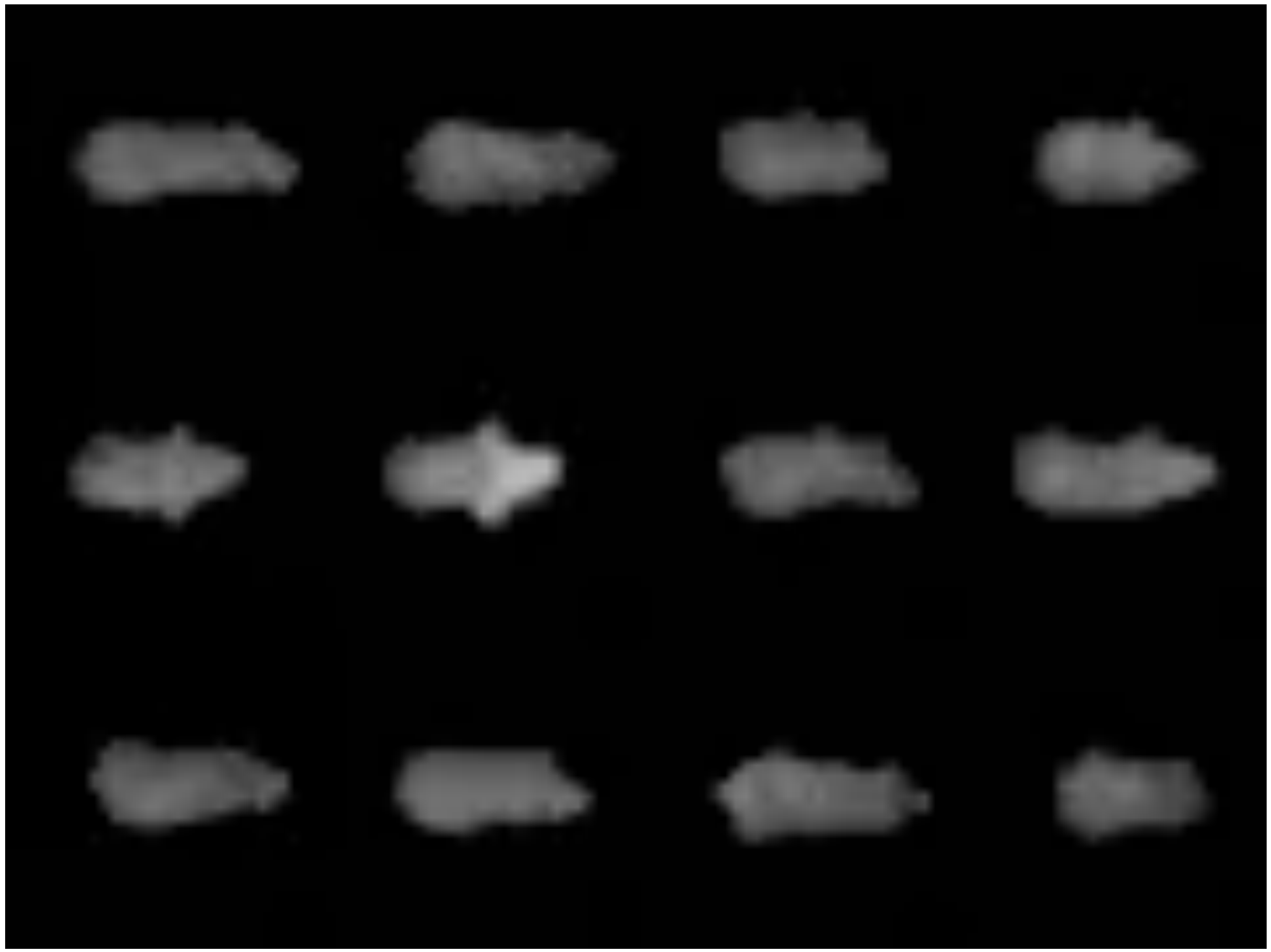


Courtesy of Matthew Johnson

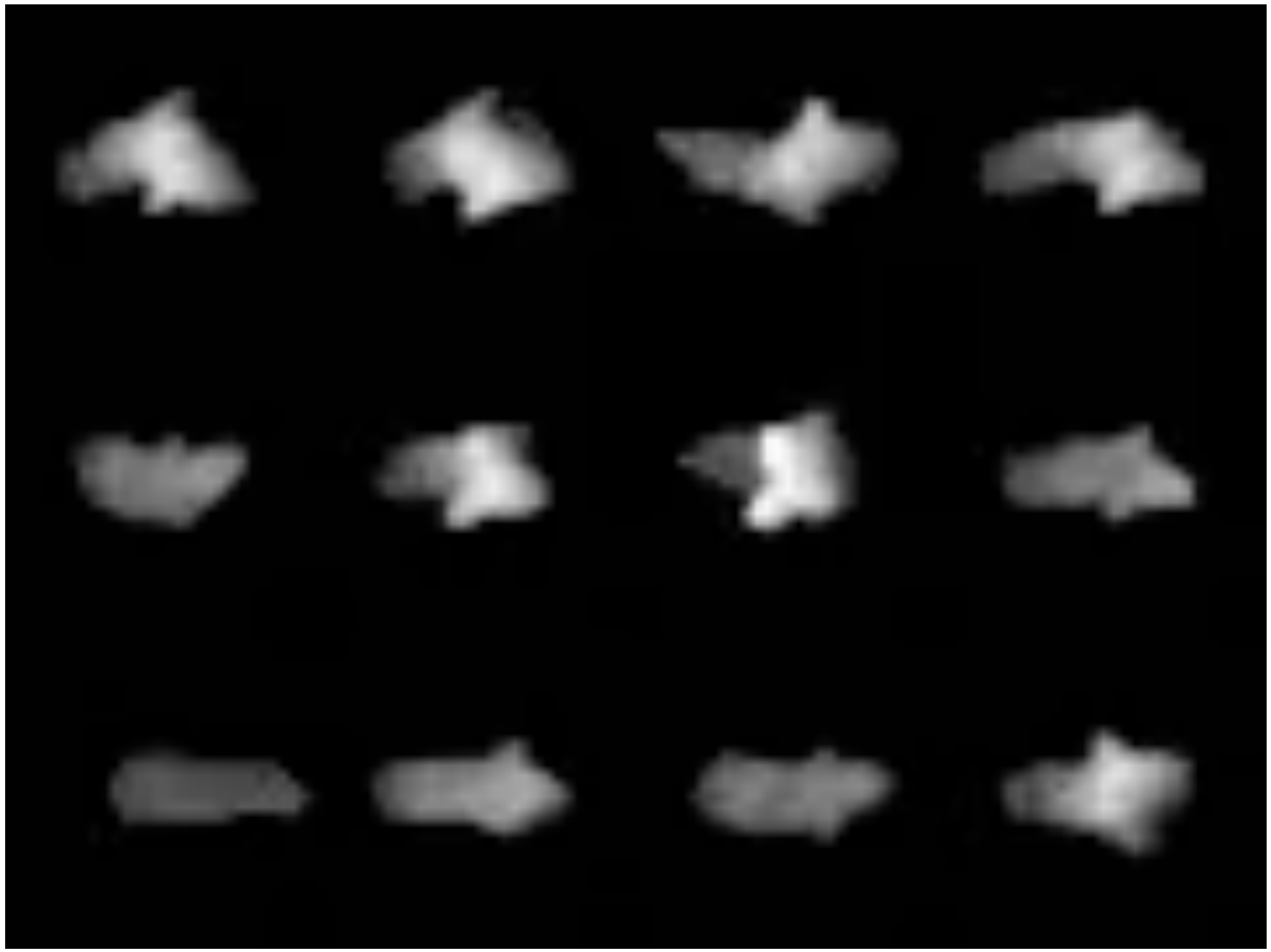




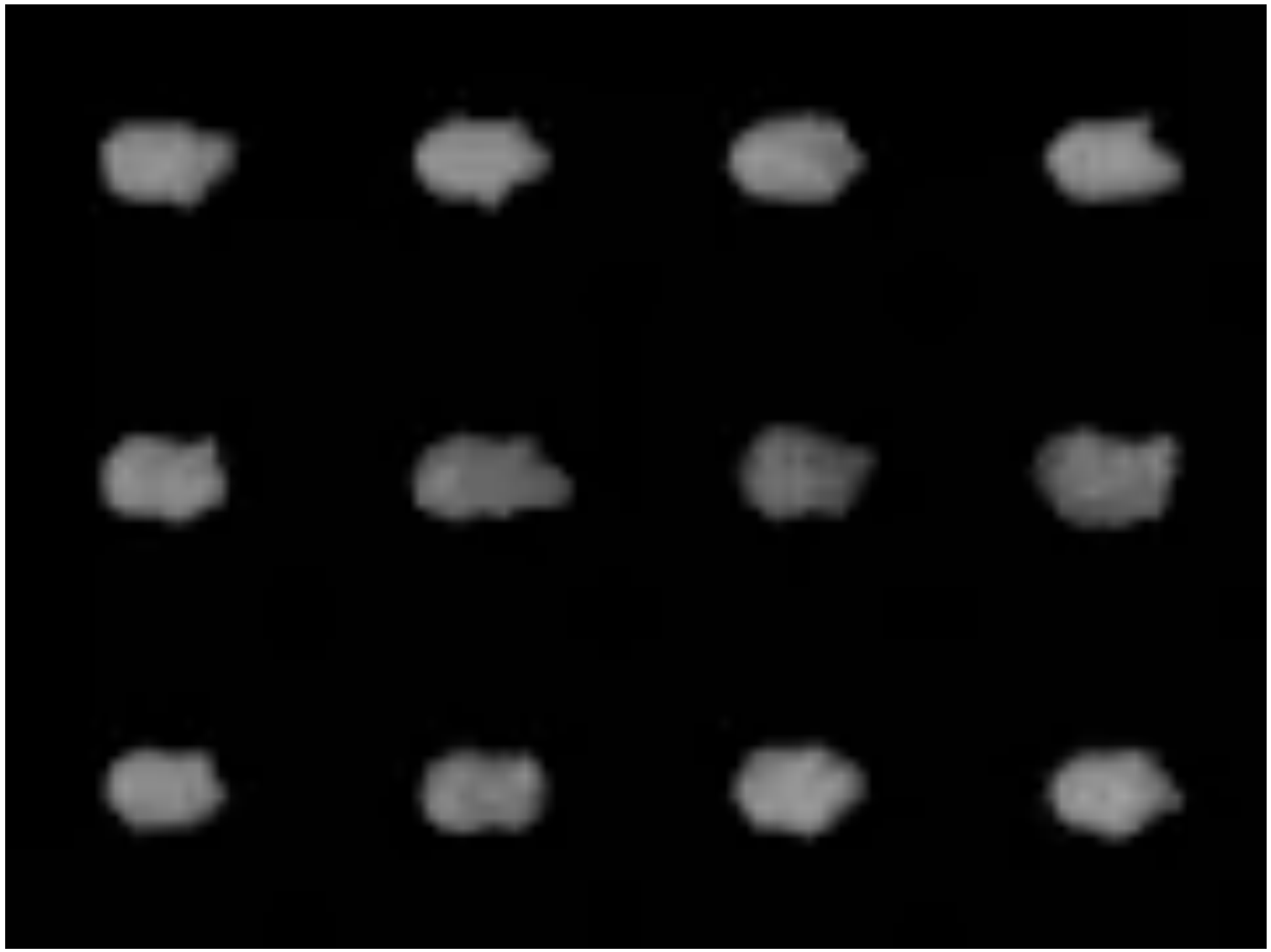




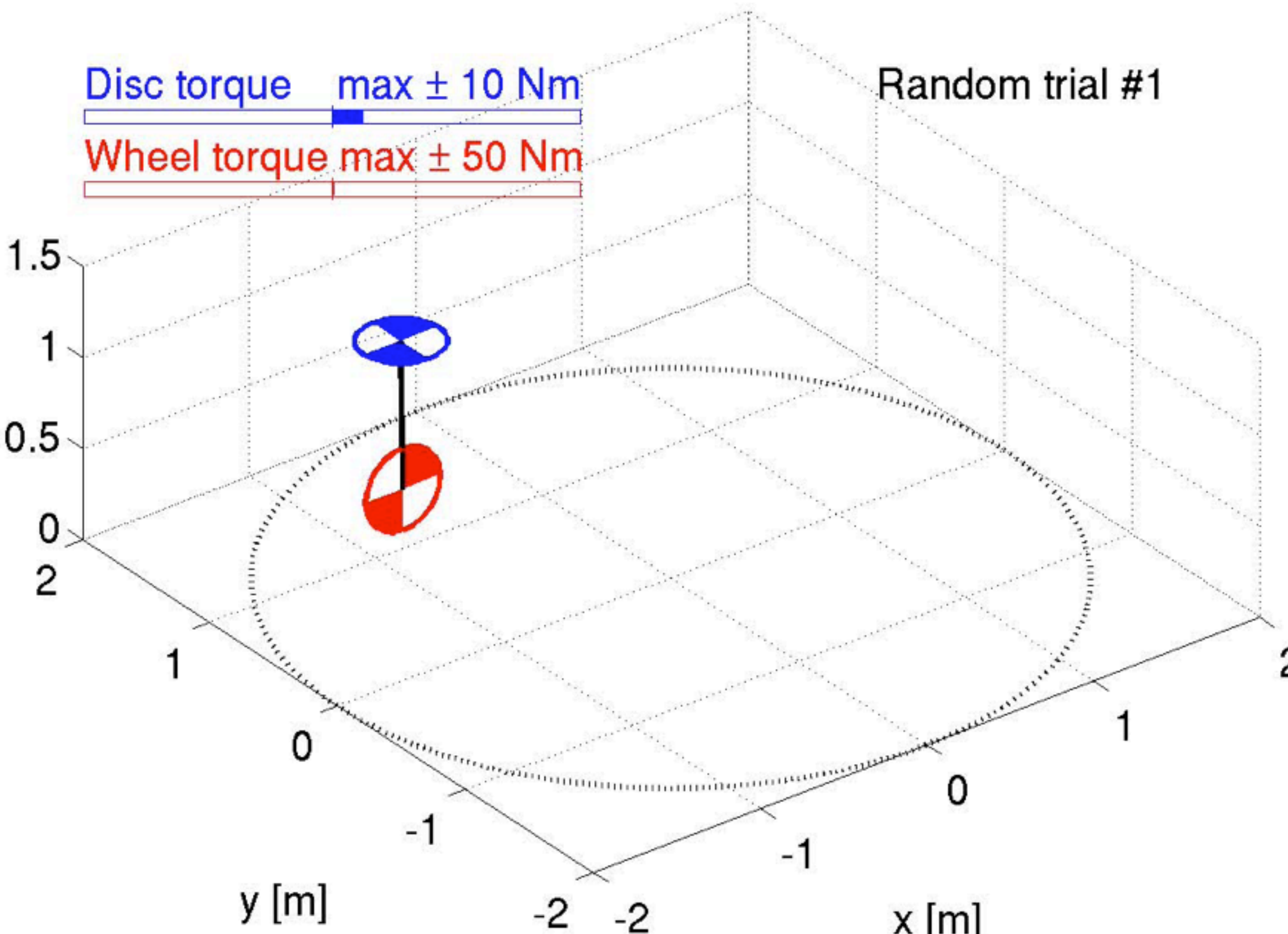
start rear



fall from rear



grooming



From Carl Rasmussen

Seminars

- 7 weeks of seminars, about 8 people each
- Each day will have one or two major themes, 3-6 papers covered
- Divided into 2-3 presentations of about 30 mins each
- Explain main idea, relate to previous work and future directions

Class Projects

- **Develop a generative model for a new medium**.
  - Generate sound given video (hard to generate raw sound)
  - Automatic onomatopoeia: Generate text ‘ka-bloom-kshhhh’ given a sound of an explosion.
  - Generating text of a specific style. For instance, generating SMILES strings representing organic molecules

Class Projects

- **Extend existing models, inference, or training**.
  For instance:

  - Extending variational autoencoders to have infinite capacity in some sense (combining Nonparametric Bayesian methods with variational autoencoders)

  - Train a VAE or GAN for matrix decomposition

  - Explore the use of mixture distributions for approximating distributions

Class Projects

- **Apply an existing approach in a new way**.
    - Missing data (not at random)
    - Automatic data cleaning (flagging suspect entries)
    - Simultaneous localization and mapping (SLAM) from scratch

Class Projects

- **Review / comparison / tutorials:**
    - Approaches to generating images
    - Approaches to generating video
    - Approaches to handling discrete latent variables
    - Approaches to building invertible yet general transformations
    - Variants of the GAN training objective
    - Different types of recognition networks
- clearly articulate the differences between different approaches, and their strengths and weaknesses.
- Ideally, include experiments highlighting the different properties of each method on realistic problems.

Class Project Dates

- Project proposal due Oct 14th
  - about 2 pages, include prelim. lit search
- Presentations: Nov 18th and 25th
- Projects due: Dec 10th

Grades

- Class presentations - 20%
- Project proposal - 20%
- Project presentation - 20%
- Project report and code - 40%

Quiz