Unsupervised Object Representation Learning using Translation and Rotation Group Equivariant VAE

Alireza Nasiri

Tristan Bepler

SIMONS MACHINE LEARNING CENTER

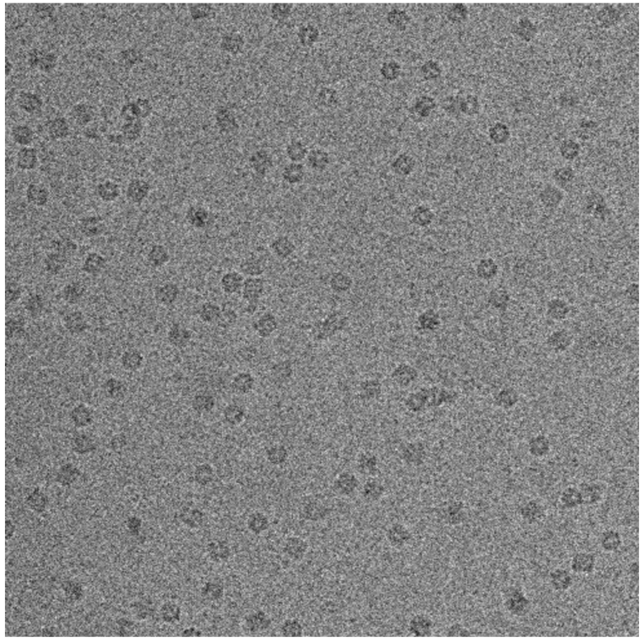
SIMONS ELECTRON MICROSCOPY CENTER

NEW YORK STRUCTURAL BIOLOGY CENTER

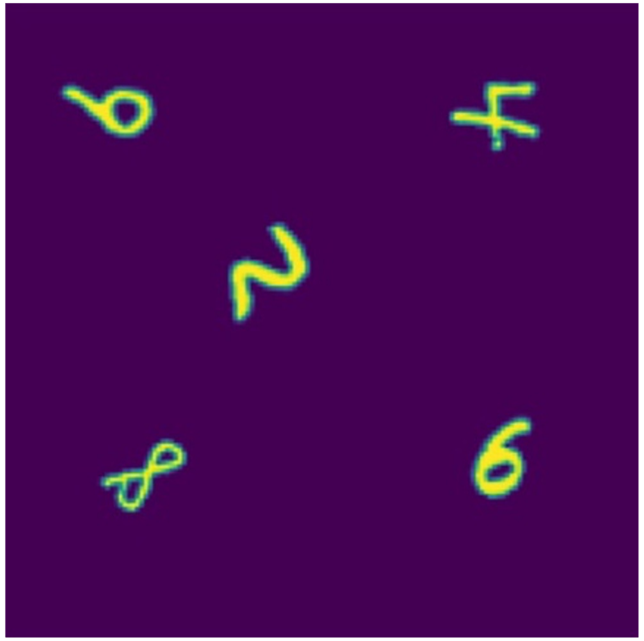
In natural images, objects often have unknown orientations

- The *pose* of an object does not change its nature
- How can we identify different objects independent of their location and pose in an unsupervised manner?

Aerial image
Source: wired.com



Electron-Microscopy Micrograph



Multi-digit hand-written numbers

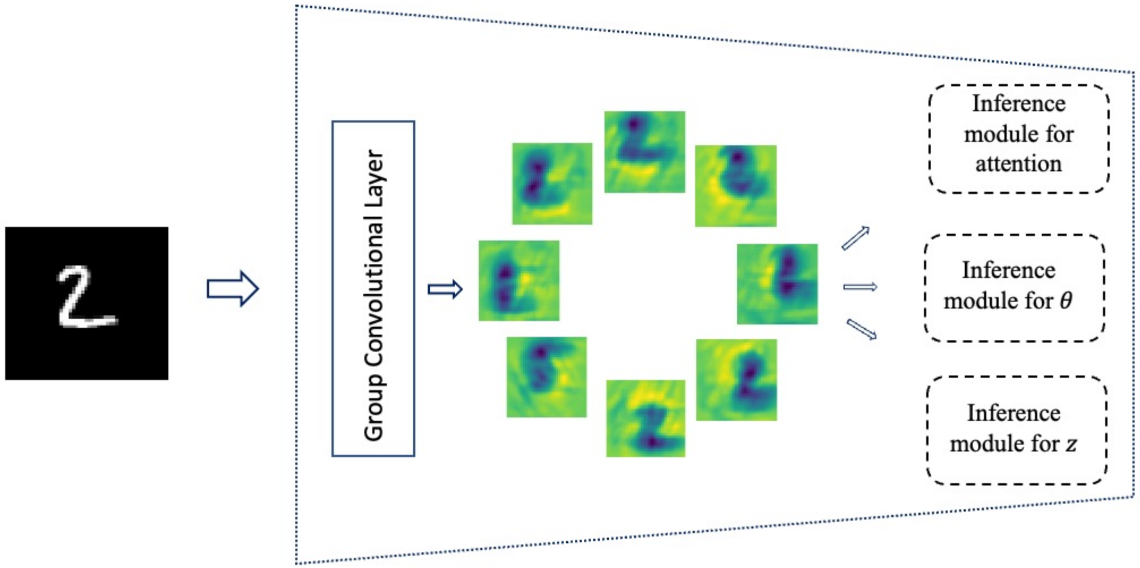
Our Proposed Method

- Goal:
  - Given images of arbitrary objects with unknown pose, learn semantic representations of those objects separately from their rotations and translations with no supervision
  - Perform efficient inference on these variables using a neural network
  - Enable controlled generation of object images from the semantic representations

- Our Proposed method, Translation and Rotation Group Equivariant Variational Auto-Encoder (TARGET-VAE), has three main components:
  1. Translation and rotation group equivariant encoder
  2. Structurally disentangled distributions over rotation, translation, and semantic representation
  3. Spatially equivariant generator

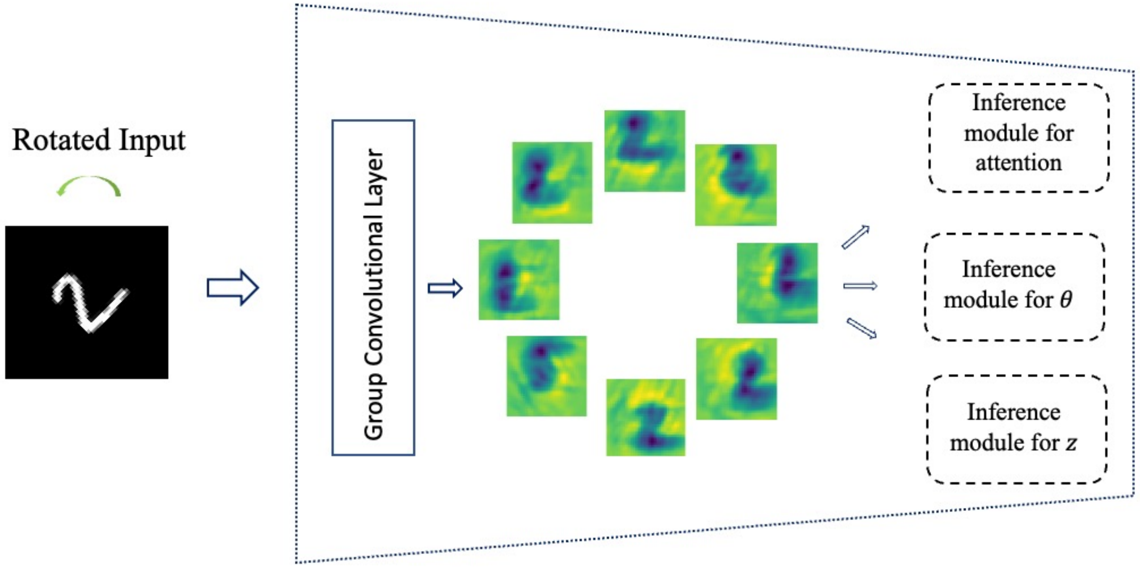
Main Components of TARGET-VAE

1. Translation and rotation group-equivariant encoder



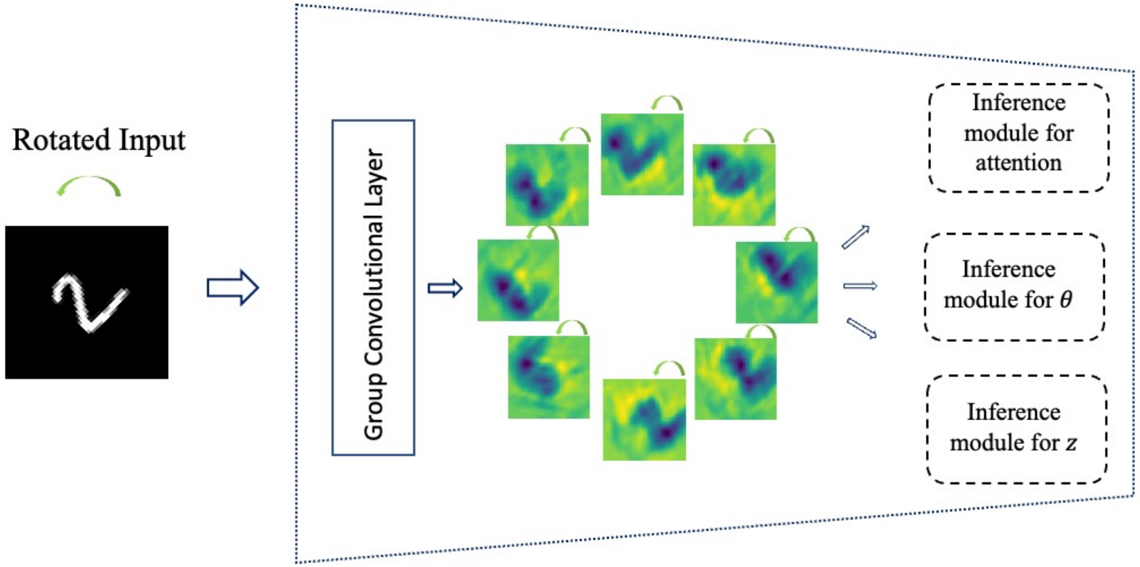
Main Components of TARGET-VAE

1. Translation and rotation group-equivariant encoder



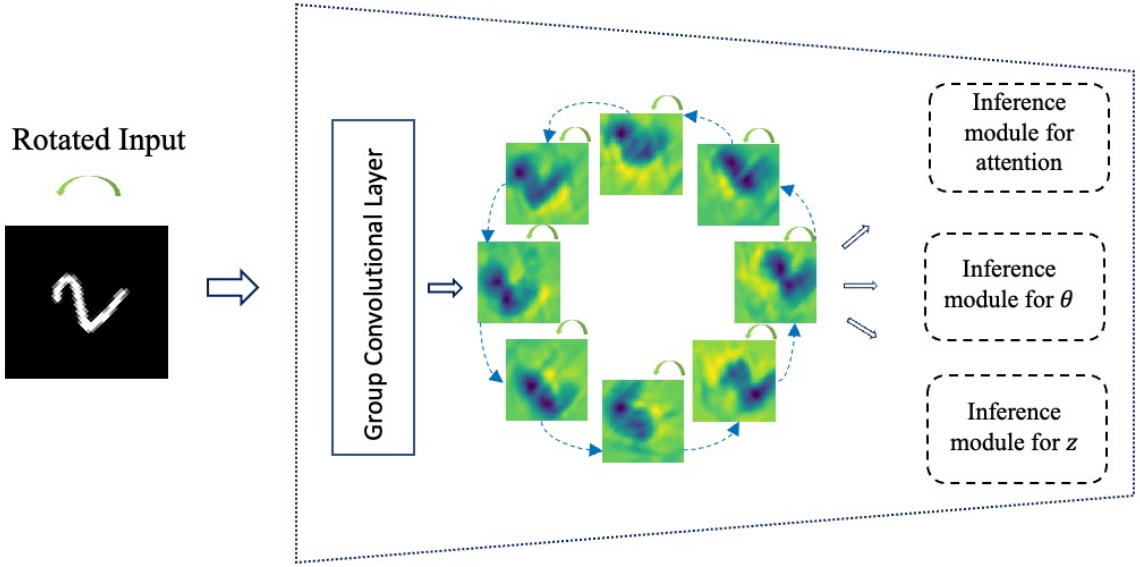
Main Components of TARGET-VAE

1. Translation and rotation group-equivariant encoder



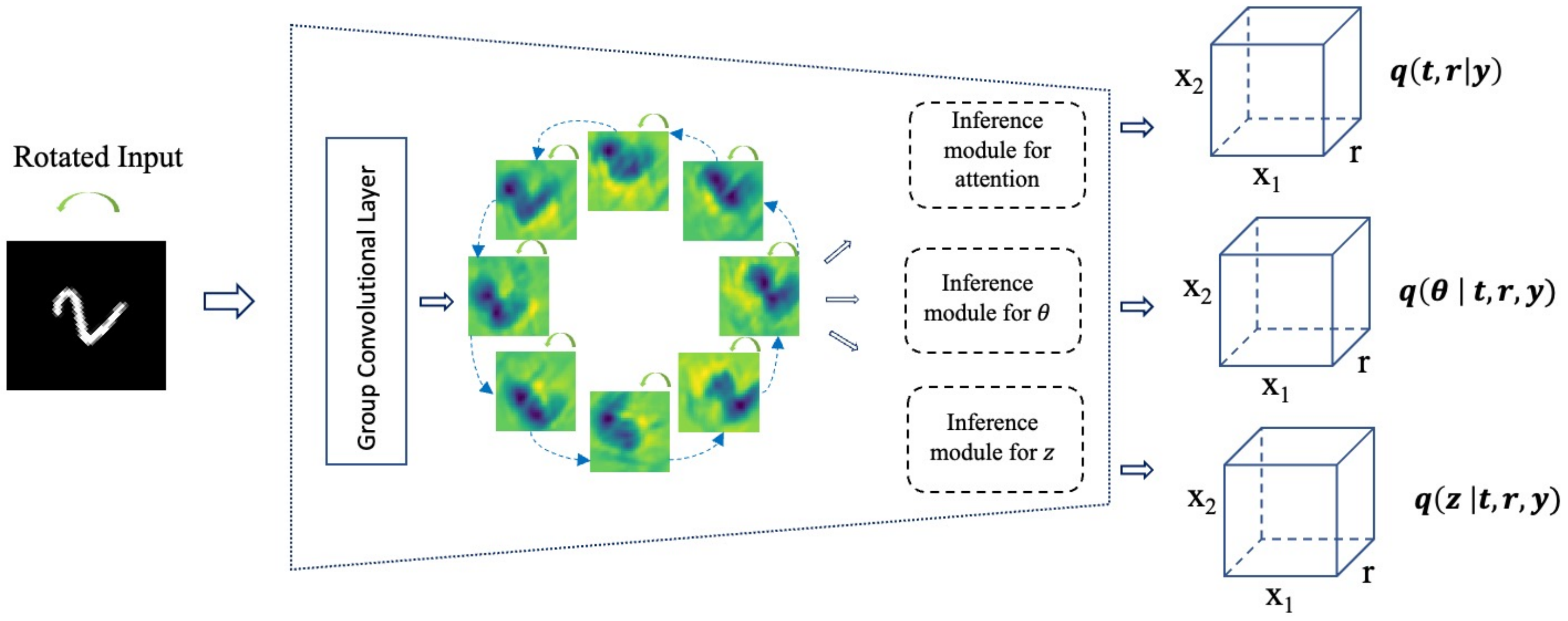
Main Components of TARGET-VAE

1. Translation and rotation group-equivariant encoder



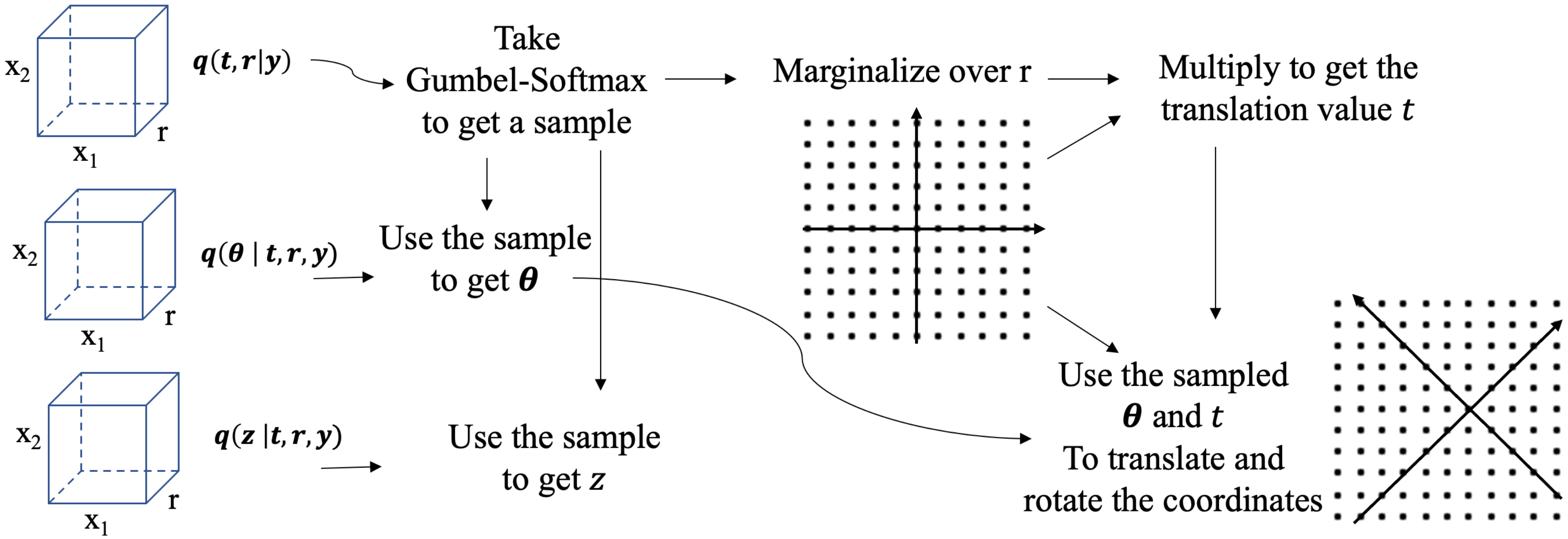
Main Components of TARGET-VAE

1. Translation and rotation group-equivariant encoder



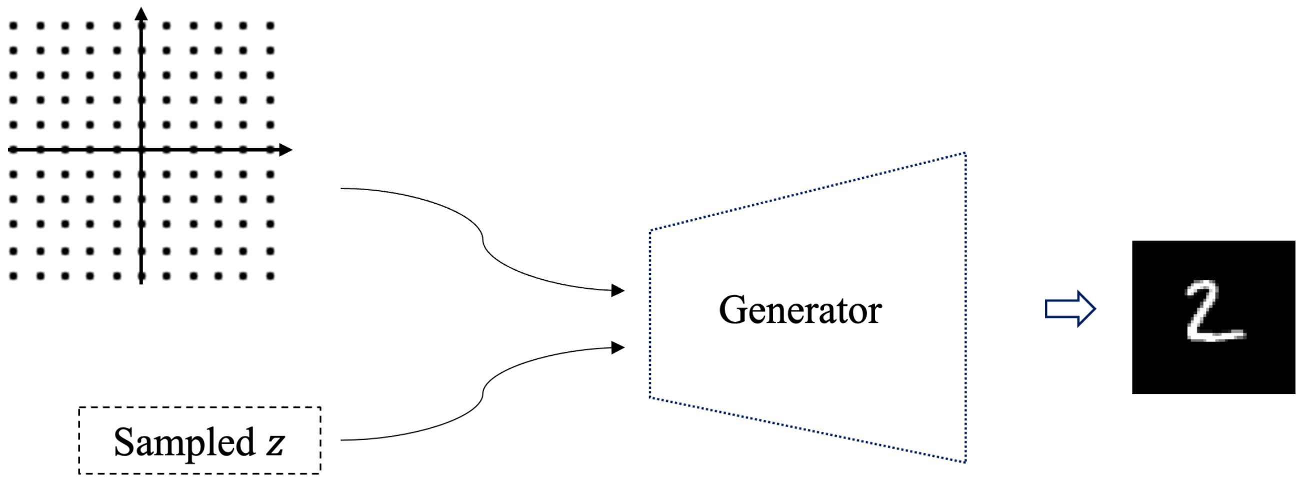
Main Components of TARGET-VAE

2. Structurally disentangled distribution over latent rotation, translation, and a rotation-translation-invariant semantic object representation



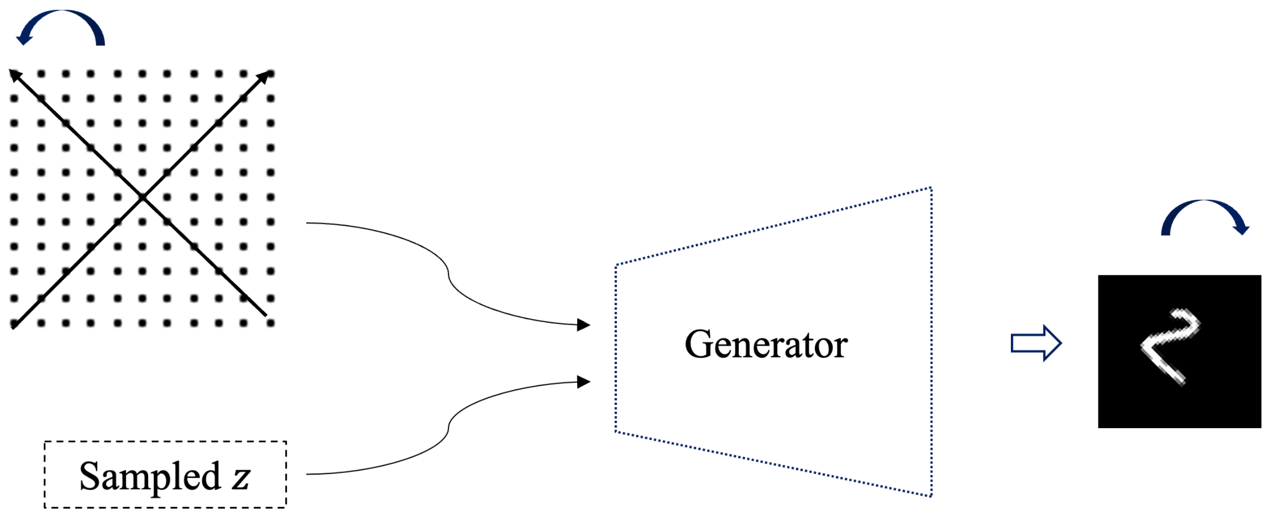
Main Components of TARGET-VAE

3. Spatially equivariant generator network



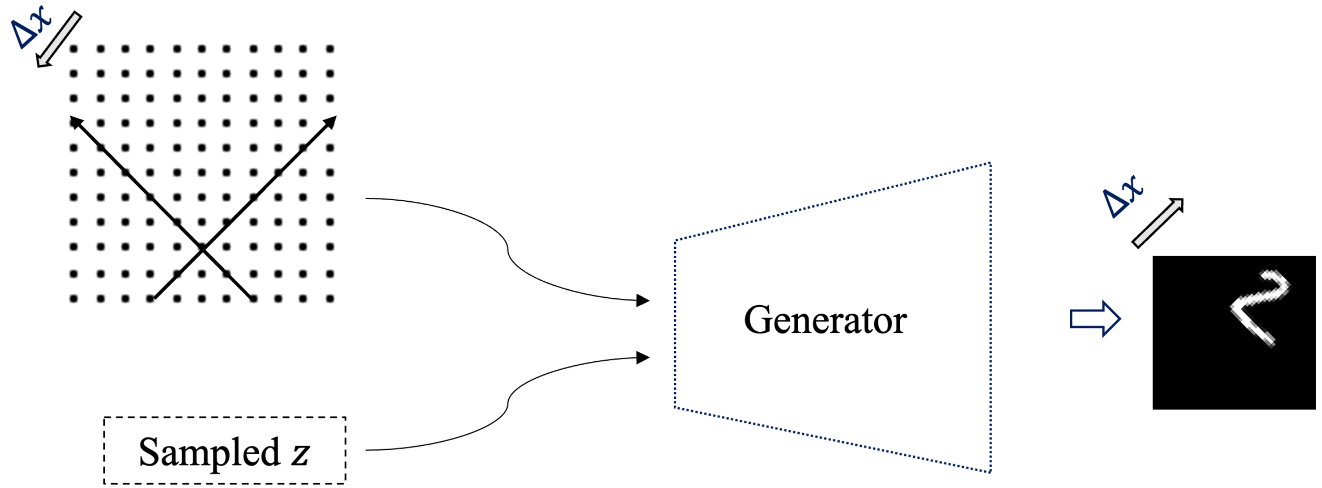
Main Components of TARGET-VAE

3. Spatially equivariant generator network

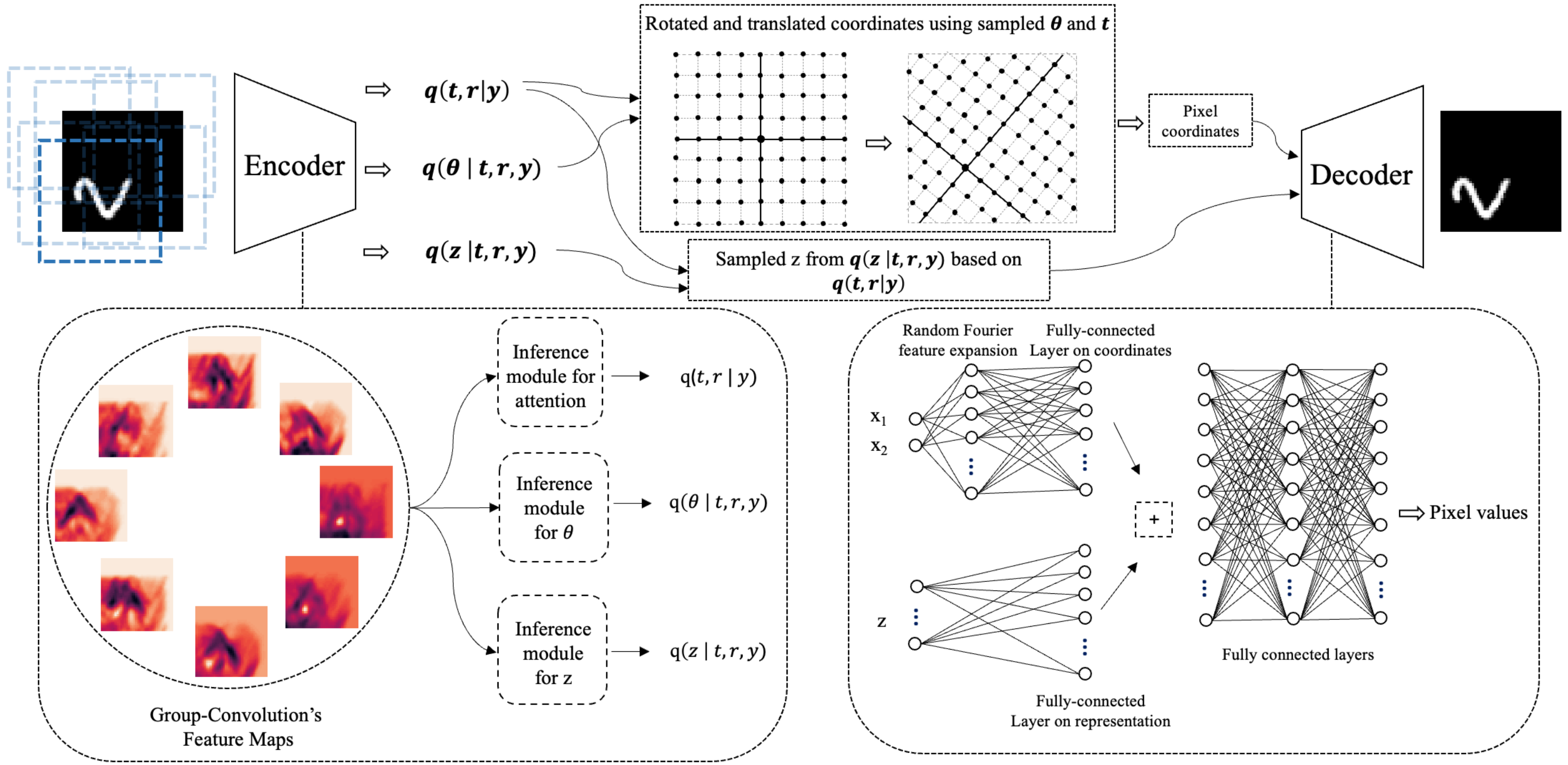


Main Components of TARGET-VAE

3. Spatially equivariant generator network

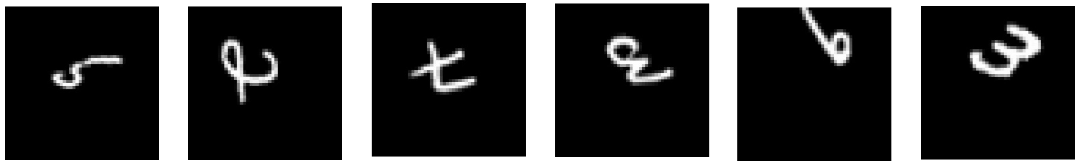


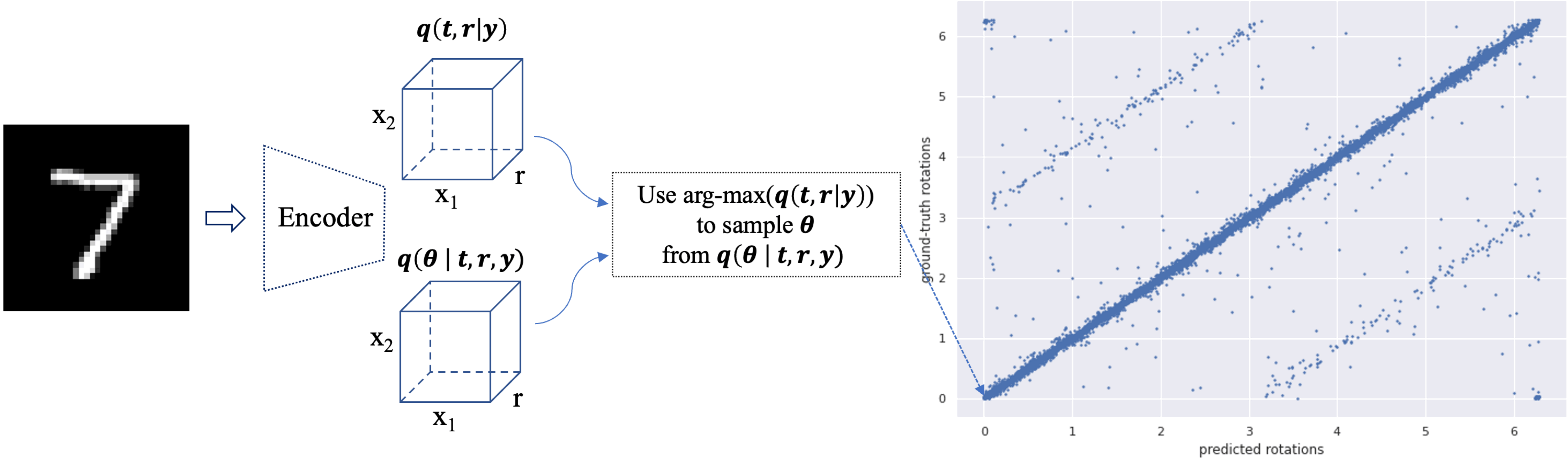
TARGET-VAE: Translation and Rotation Group Equivariant VAE



Results – TARGET-VAE Accurately Predicts Rotation without Supervision

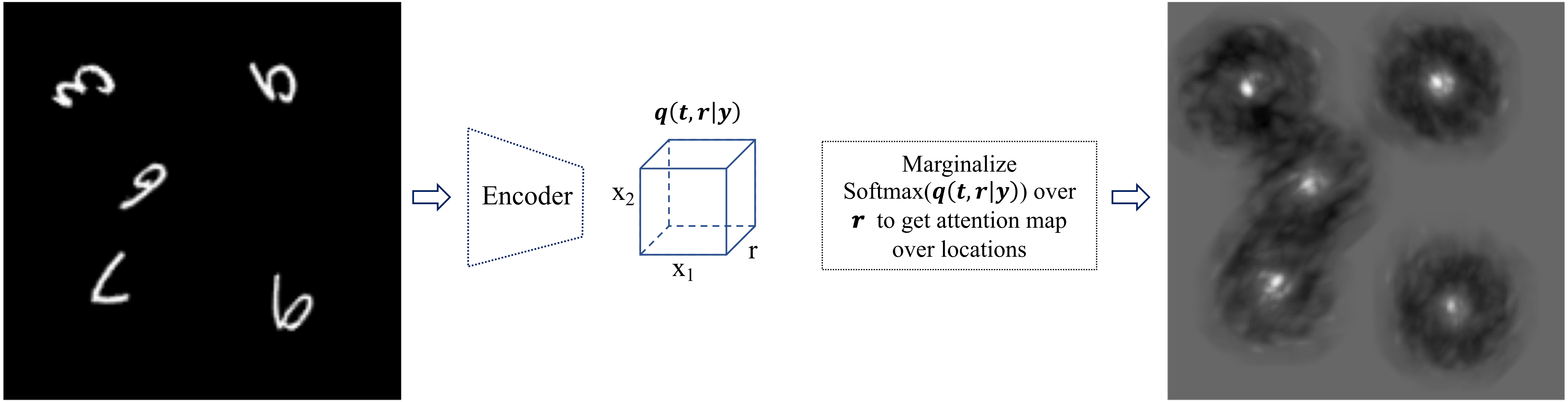
Rotated and Translated Digits:



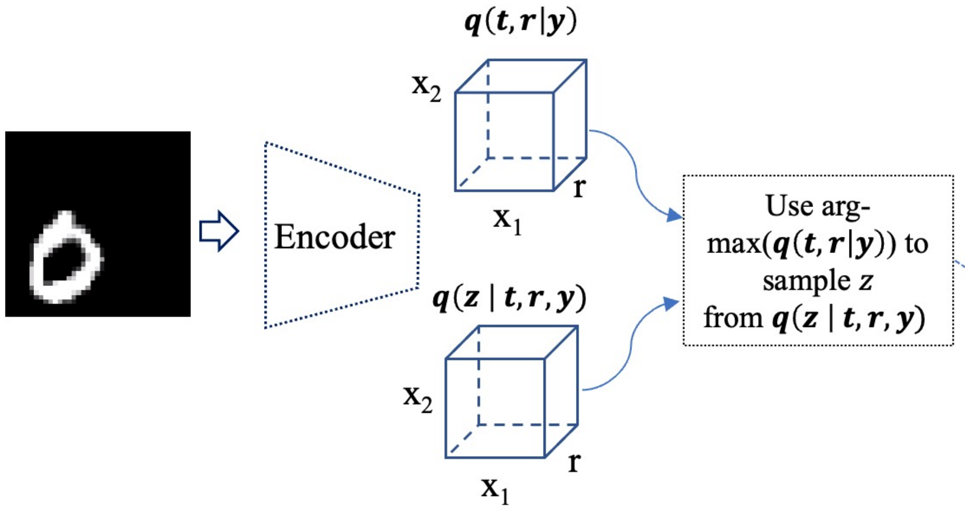


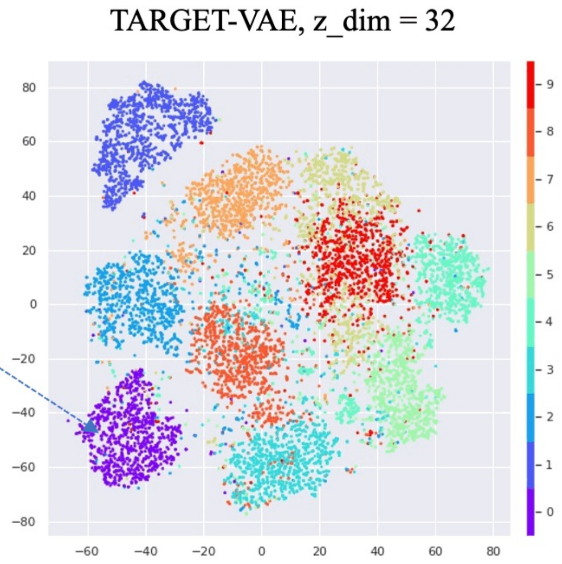
Circular Correlation coefficient = 0.86

Results – TARGET-VAE Accurately Predicts Translation without Supervision



Results – TARGET-VAE Learns Rotation and Translation Invariant Semantic Representation





Clustering Accuracy: 71.6 %

Table 2: Clustering accuracy (%) on MNIST(N) and MNIST(U)

Model	MNIST(N)	MNIST(U)
VAE (z_dim=4) [3]	15.3	12.8
Beta-VAE (z_dim=4) [4]	15.1	18.0
Spatial-VAE (z_dim=2) [6]	37.1	28.2
TARGET-VAE P_4 (z_dim=2)	56.4	56.6
TARGET-VAE P_8 (z_dim=2)	60.1	57.1
TARGET-VAE P_{16} (z_dim=2)	60.1	63.4
TARGET-VAE P_4 (z_dim=32)	65.1	64.3
TARGET-VAE P_8 (z_dim=32)	**77.7**	69.1
TARGET-VAE P_{16} (z_dim=32)	75.2	**71.2**

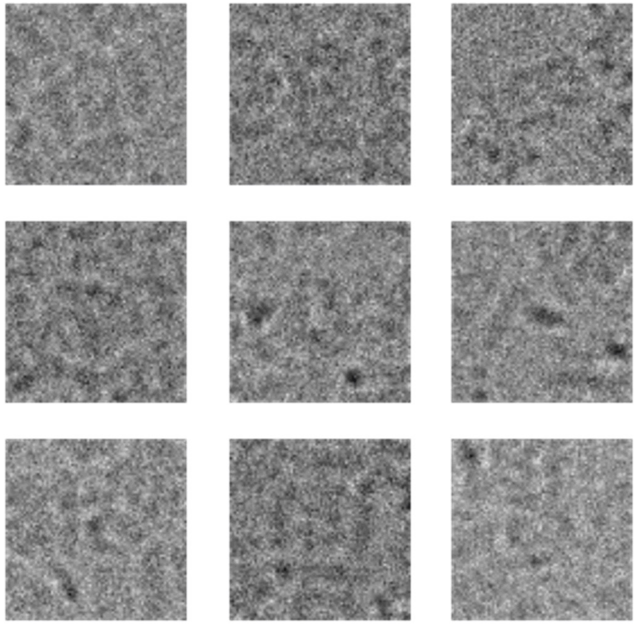
Results – Identifying Protein Heterogeneity on Cryo-EM Particle Stack

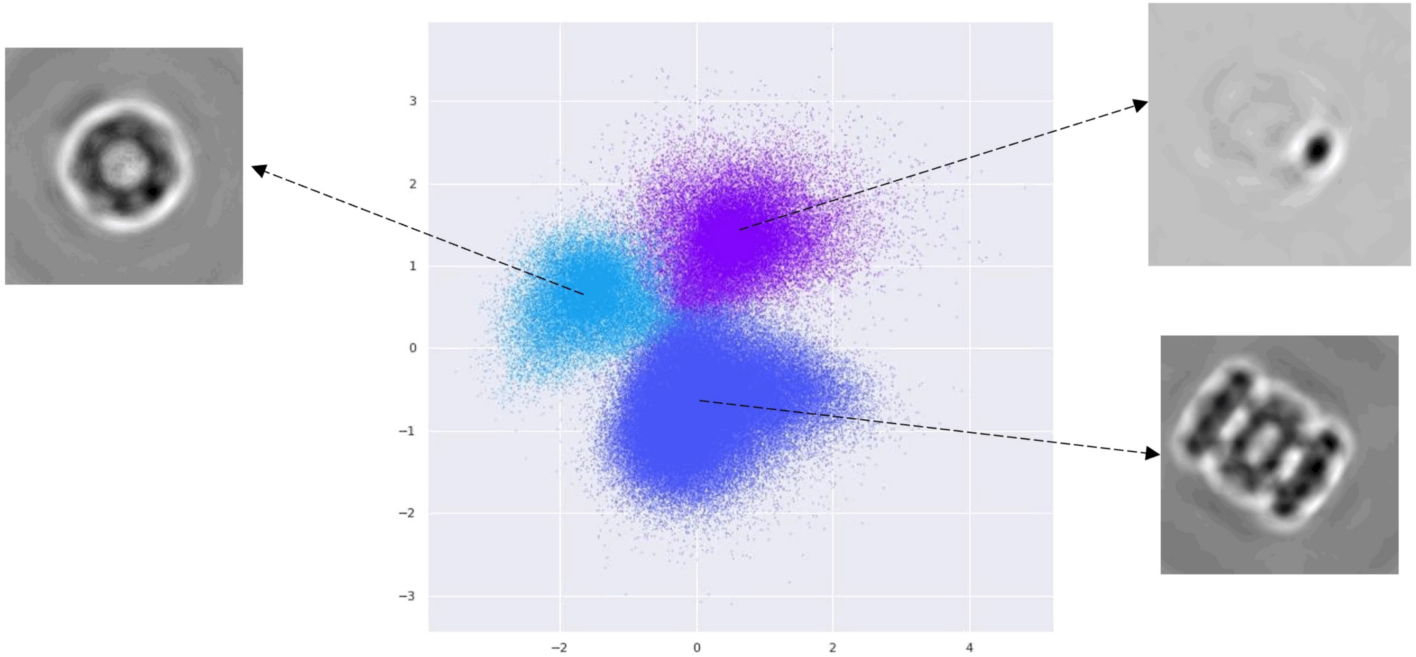
EMPIAR 10025 - T20S Proteasome at 2.8 Å Resolution

Particle stack of 161,292 images 400x400 downsampled to 100x100

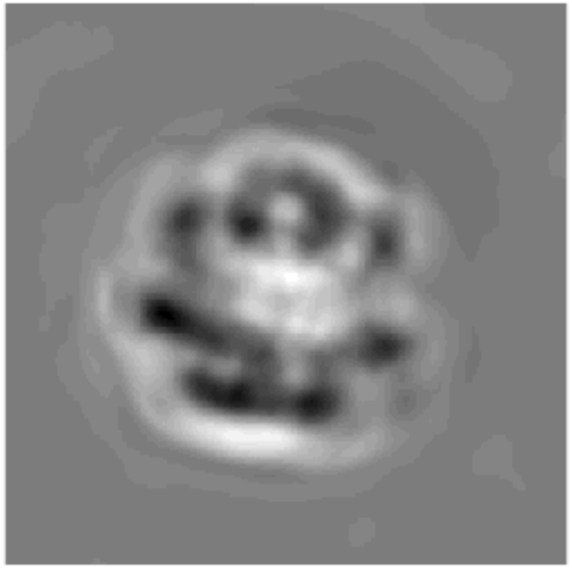
Pixel spacing: (0.66 Å, 0.66 Å)

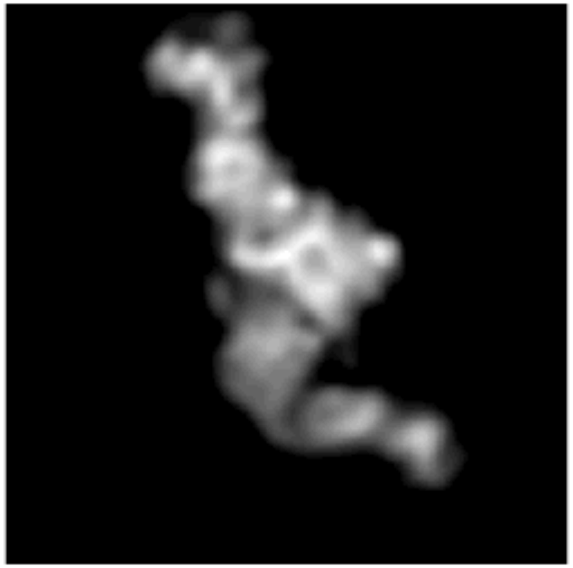
Latent dimension = 2

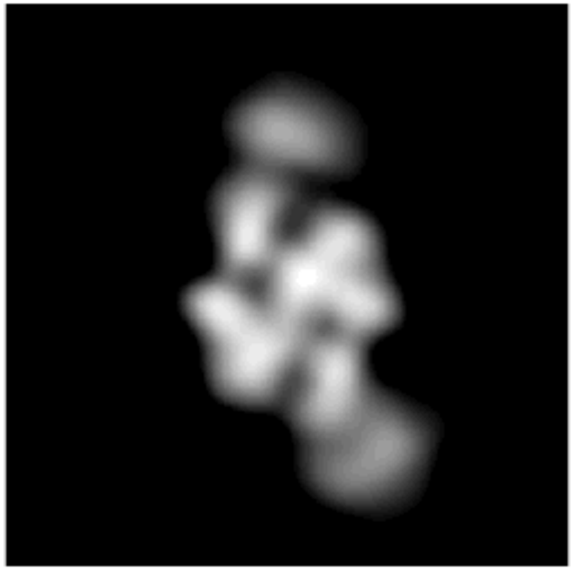




Results – Identifying Protein Heterogeneity on Cryo-EM Data







Conclusions and future work

- By designing models to capture relevant equivariances, we are able to better disentangle content from rotation and translation
- Accurate amortized inference on rotation and translation
- Learn disentangled semantic representations of objects with the ability to generate new images from the object manifold
- In the future
    - Multi object detection and unsupervised object tracking over time
    - Amortized pose inference for 3D reconstruction
    - Fully unsupervised particle picking + 2D classification

Thanks!

SMLC

- **Tristan Bepler**
- Robert Kiewisz
- Paul Kim
- Darnell Granberry
- Jiayi Shou

SEMC

- Bridget Carragher and Clint Potter
- Alex Noble

York University

- Marcus Brubaker

- Funded by Simons Foundation

SIMONS MACHINE LEARNING CENTER

SIMONS ELECTRON MICROSCOPY CENTER

NEW YORK STRUCTURAL BIOLOGY CENTER

SIMONS FOUNDATION