Encoding Event-Based Data With a Hybrid SNN Guided Variational Auto-encoder in Neuromorphic Hardware

Kenneth M Stewart, Andreea Danielescu,
Timothy M Shea, Emre O Neftci

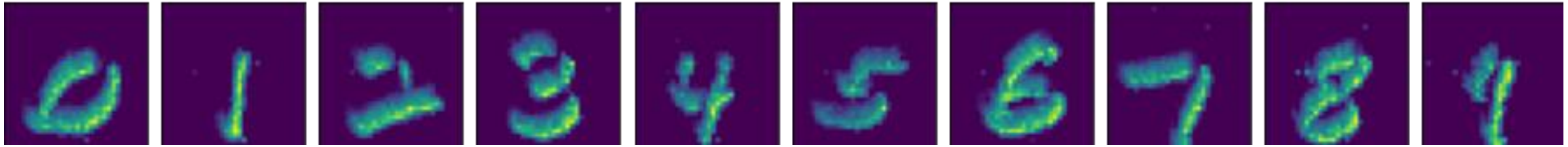
accenture

Accenture Technology

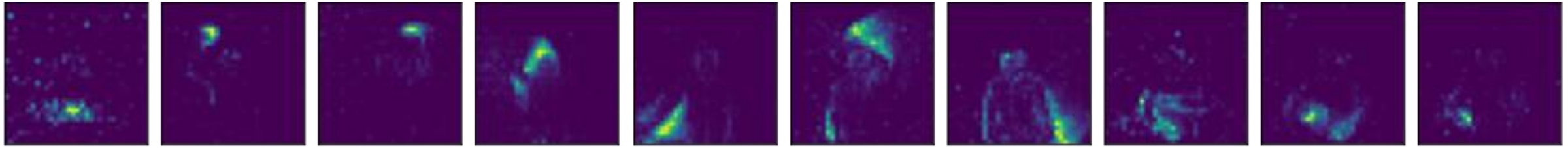
Intro and Motivation

Supervised Learning

- Achieves State of the Art Performance on a Variety of Tasks
- Needs labelled data
- Many iterations of training
- Retraining or transfer learning for learning new classes



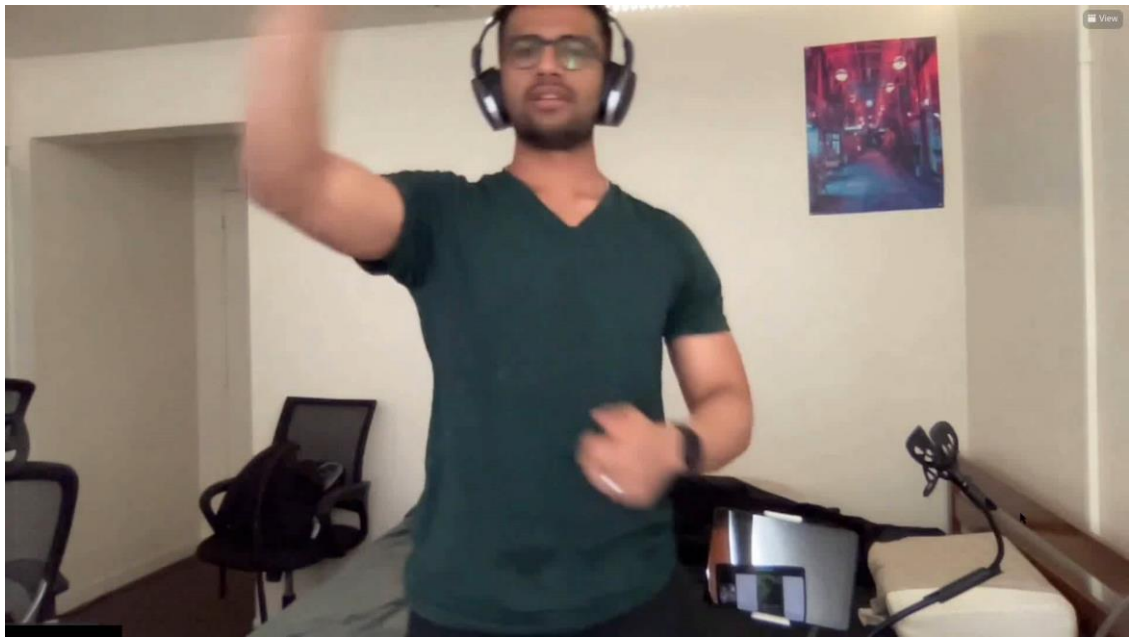
NMNIST: 99% Accuracy



DVSGesture: 96% Accuracy

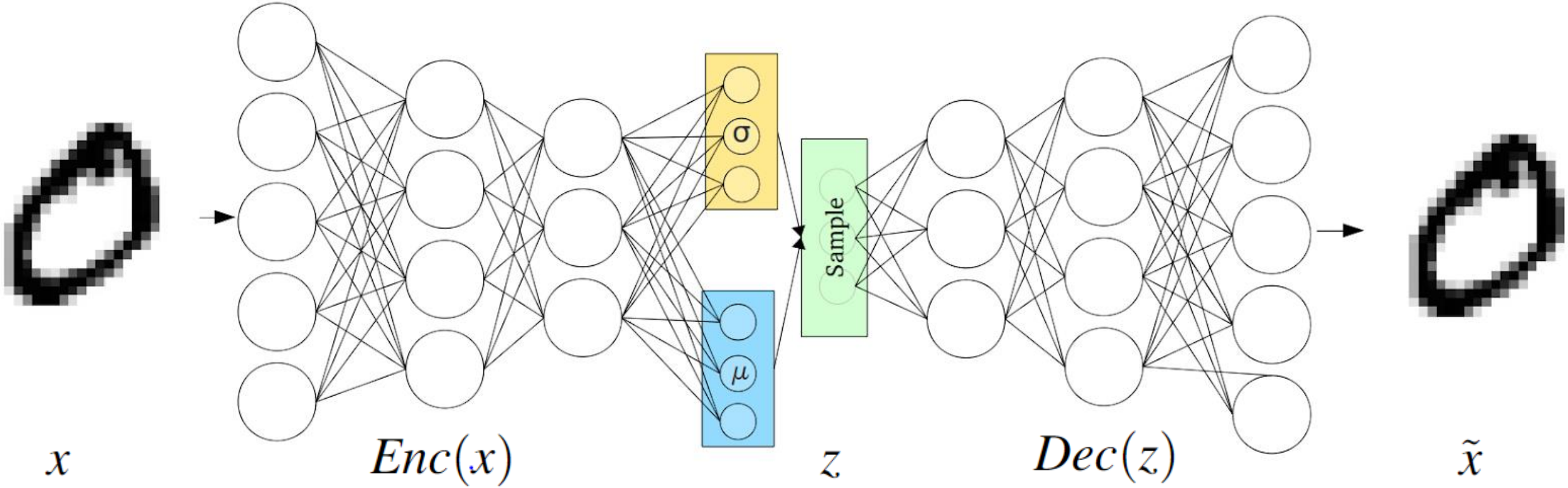
Why Learn from Unlabeled Data

- Difficult to collect sufficiently large data sets for supervised learning
- Limitations of data sets to cover all potential scenarios
- Potential to customize learning to users and scenarios



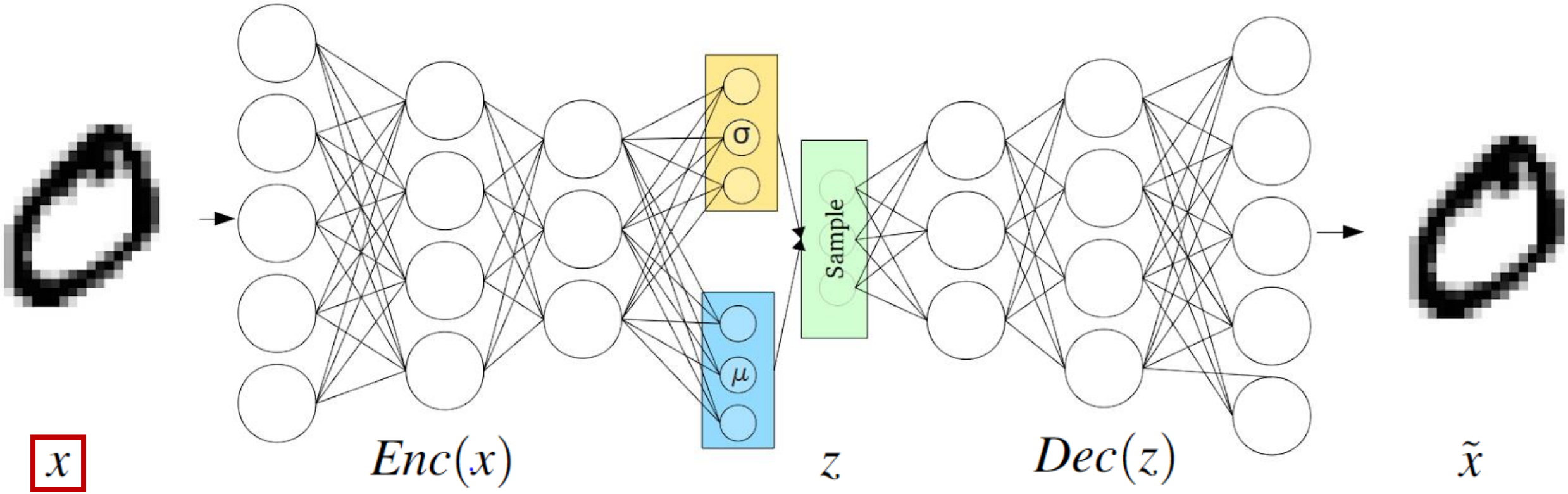
Variational Auto Encoders

- How to learn from new data without a dedicated training phase?
- Learn disentangled representation of the data



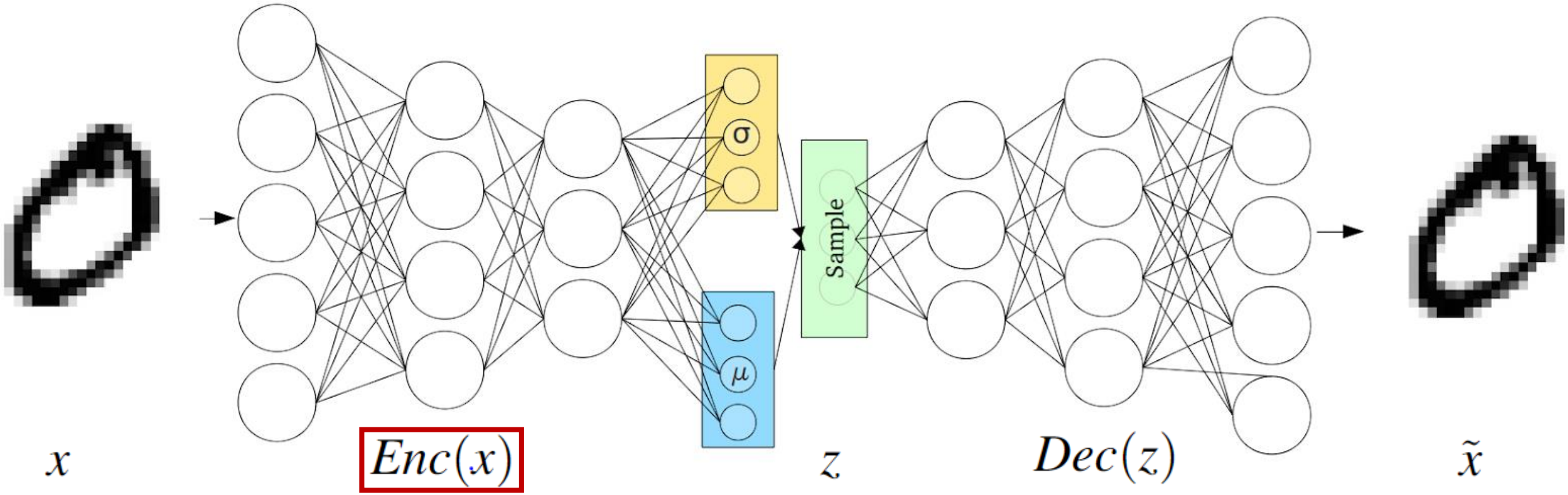
Variational Auto Encoders

- How to learn from new data without a dedicated training phase?
- Learn meaningful causal factors of variation



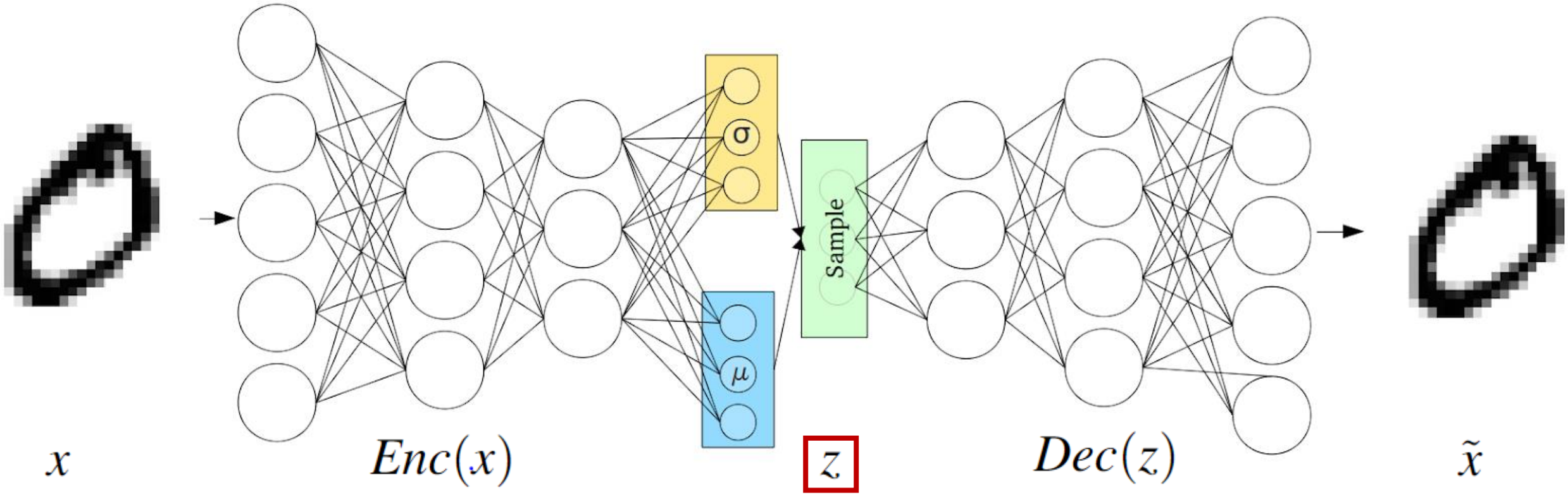
Variational Auto Encoders

- How to learn from new data without a dedicated training phase?
- Learn meaningful causal factors of variation



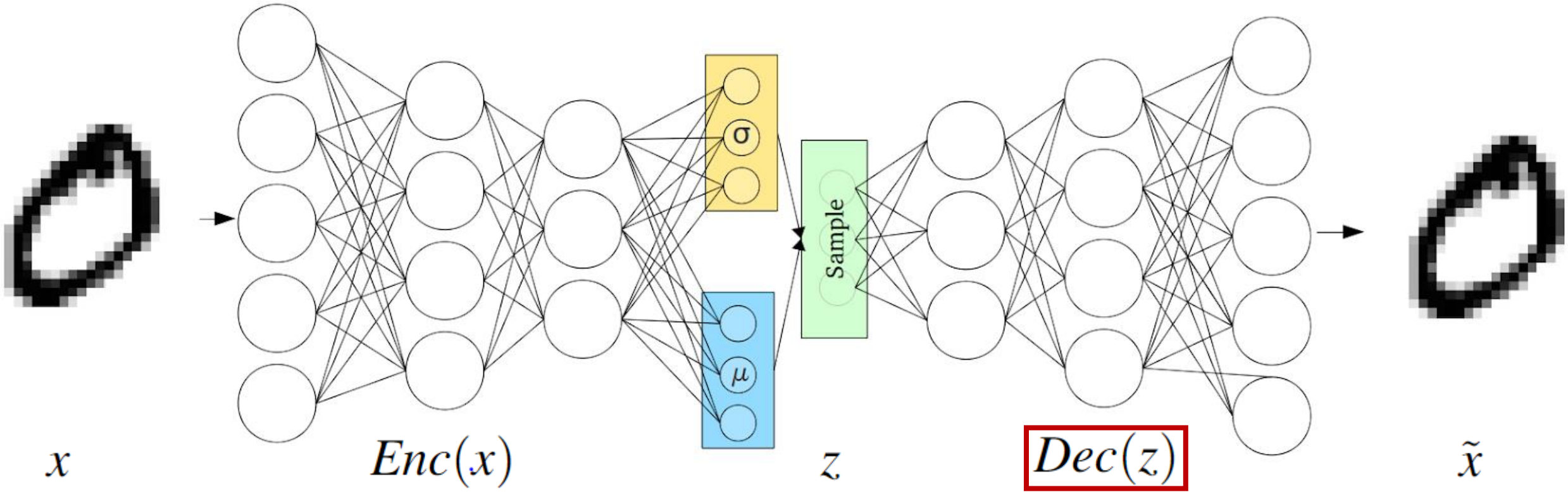
Variational Auto Encoders

- How to learn from new data without a dedicated training phase?
- Learn meaningful causal factors of variation



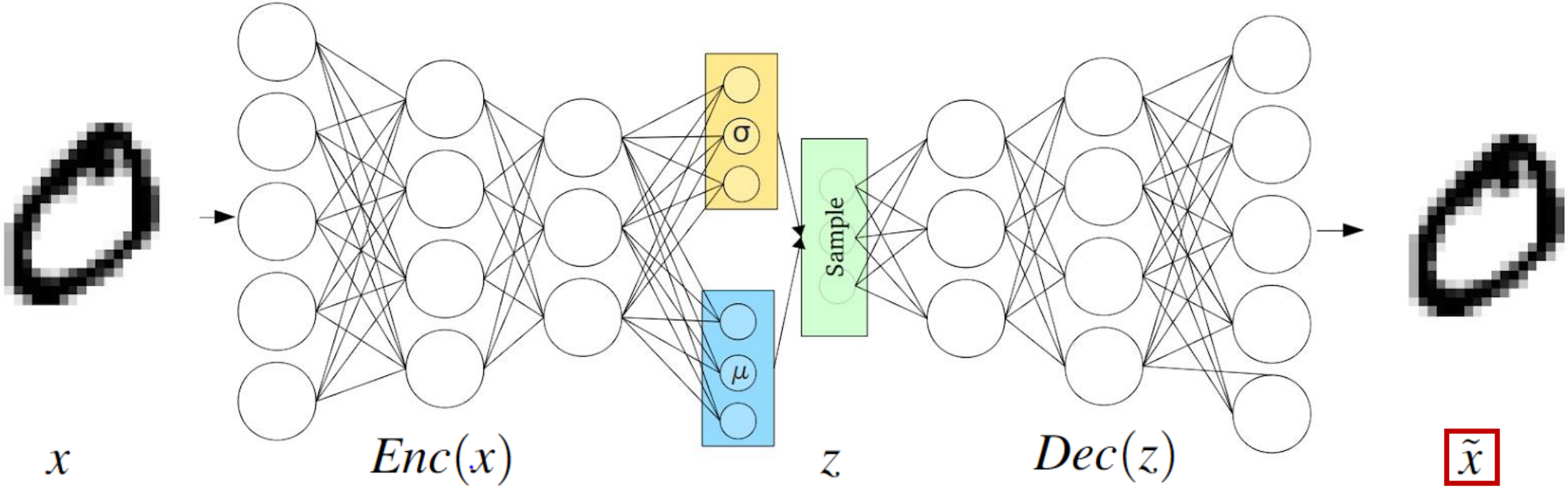
Variational Auto Encoders

- How to learn from new data without a dedicated training phase?
- Learn meaningful causal factors of variation



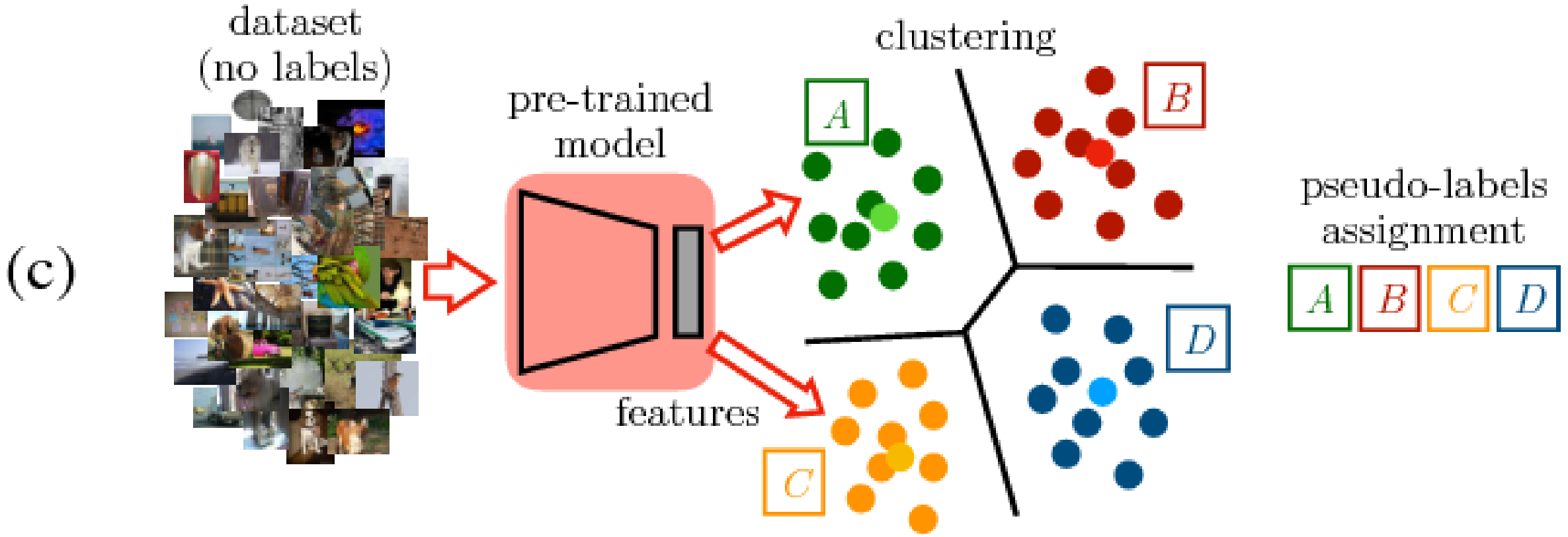
Variational Auto Encoders

- How to learn from new data without a dedicated training phase?
- Learn meaningful causal factors of variation



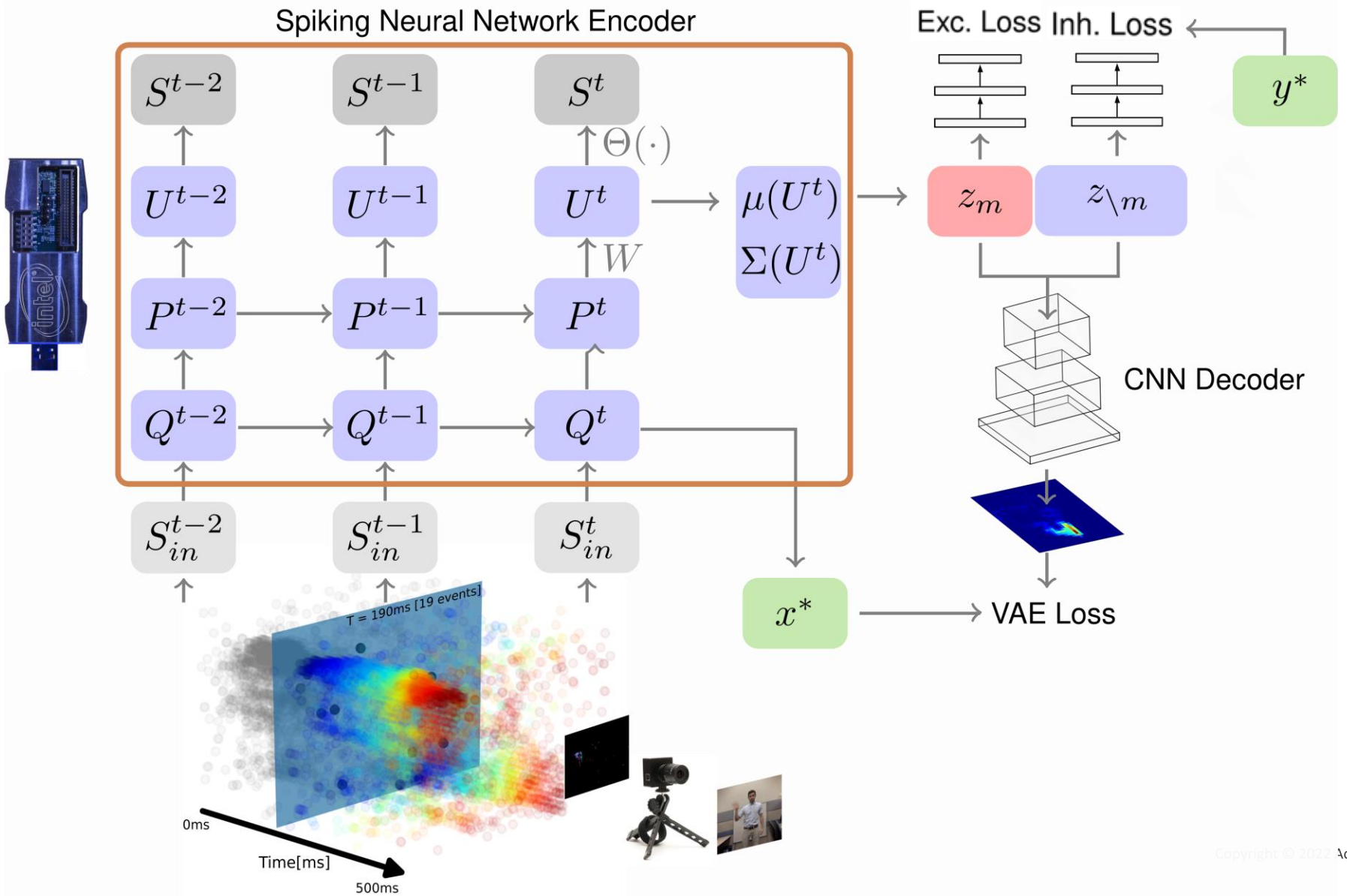
Self-Labelling Data

- Supervised to self-supervised learning with latent representation learning



Hybrid Guided Variational Auto-Encoder

Hybrid Guided Variational Auto-Encoder

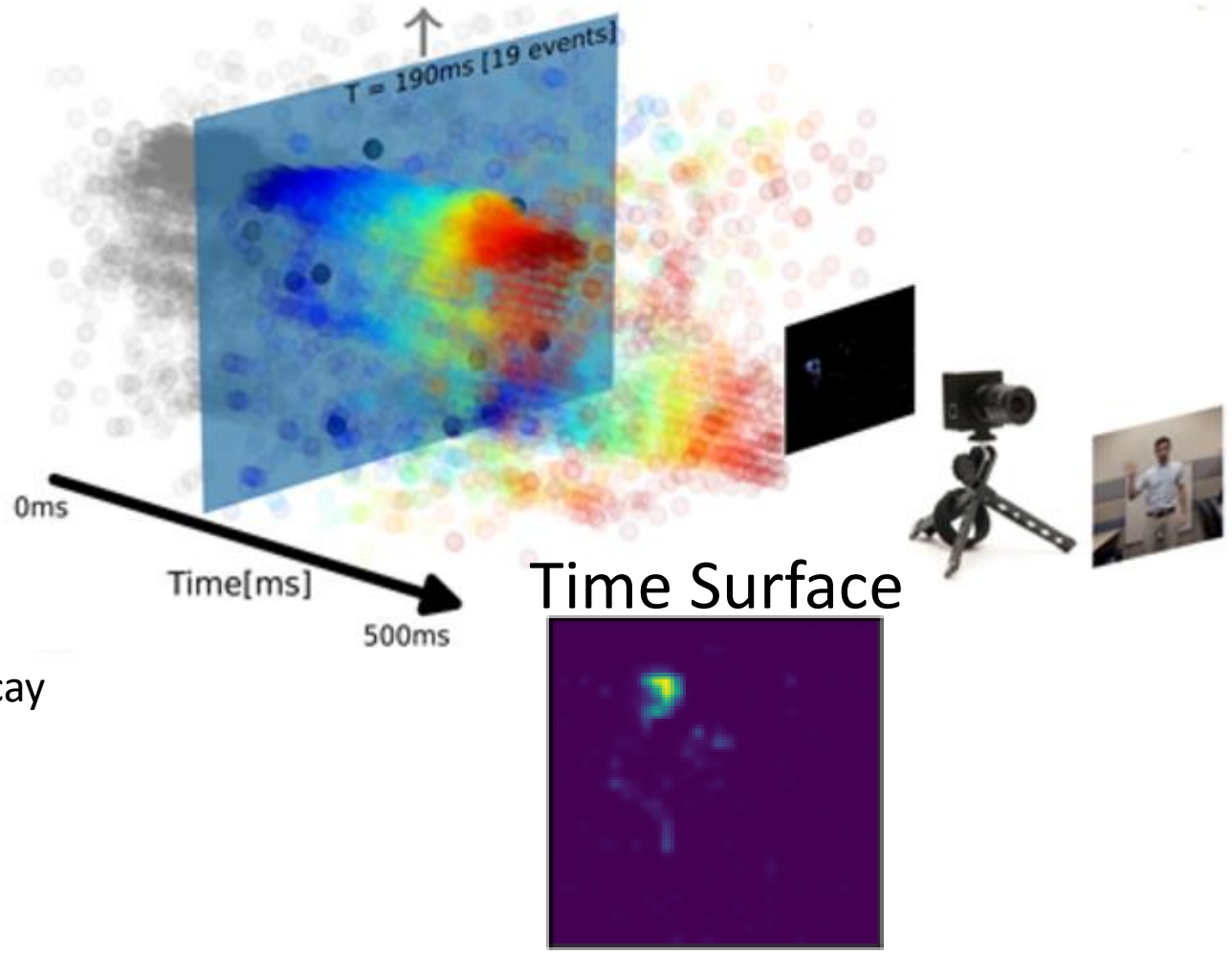


Hybrid Guided Variational Auto-Encoder

Event Sensor Data Streams and Time Surfaces (TS)

- DVS record event streams at a high temporal resolution
- Compatible with SNNs
- Detect brightness changes on a log scale
- Event: *x*, *y*, time *t*, polarity *p*
- Event stream: $S^t_{DVS,x,y,p} \in \mathbb{N}^+$
- Use Time Surfaces (TS) for VAE targets
- TS constructed by convolving an exponential decay kernel through time in the event stream

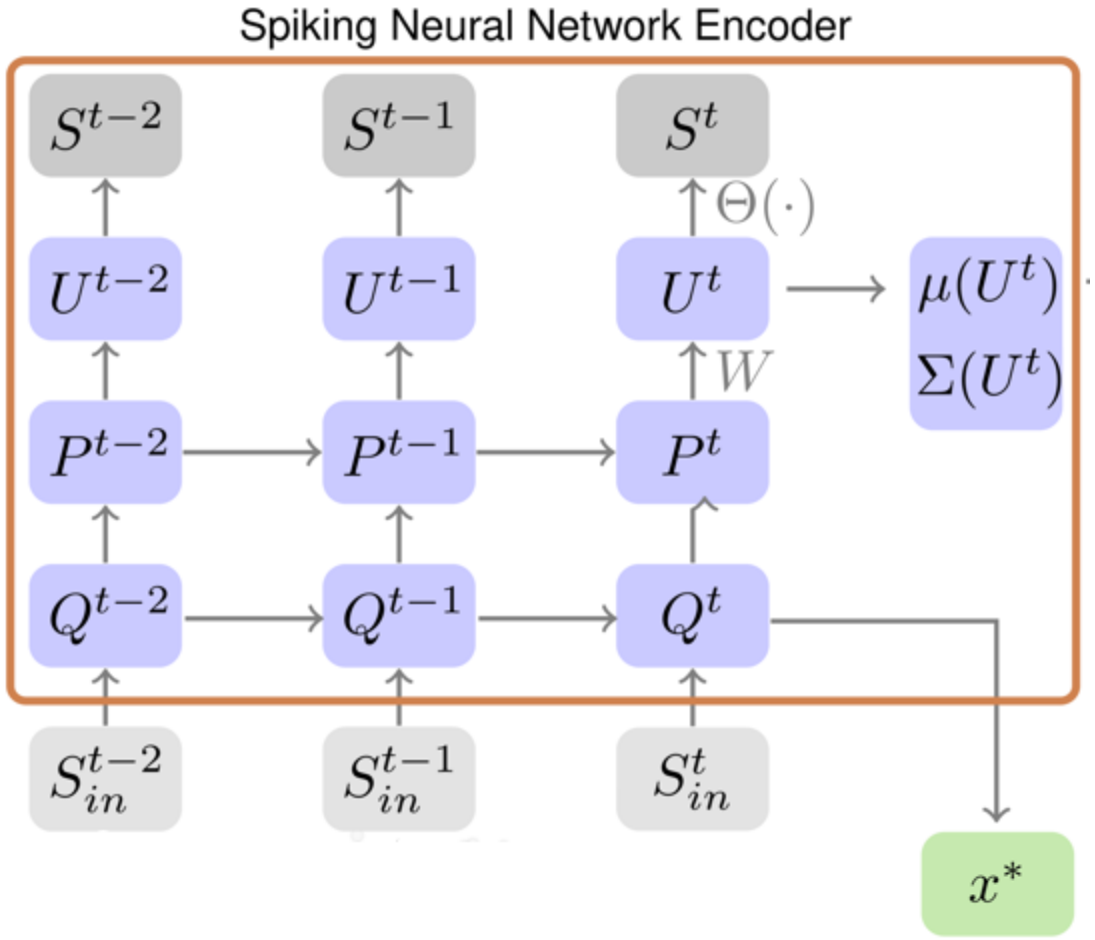
$$TS^t_{x,y,p} = \epsilon^t * S^t_{DVS,x,y,p} \text{ with } \epsilon^t = \mathrm{e}^{-\frac{t}{\tau}}$$



Hybrid Guided Variational Auto-Encoder

DECOLLE SNN Encoder (GPU)

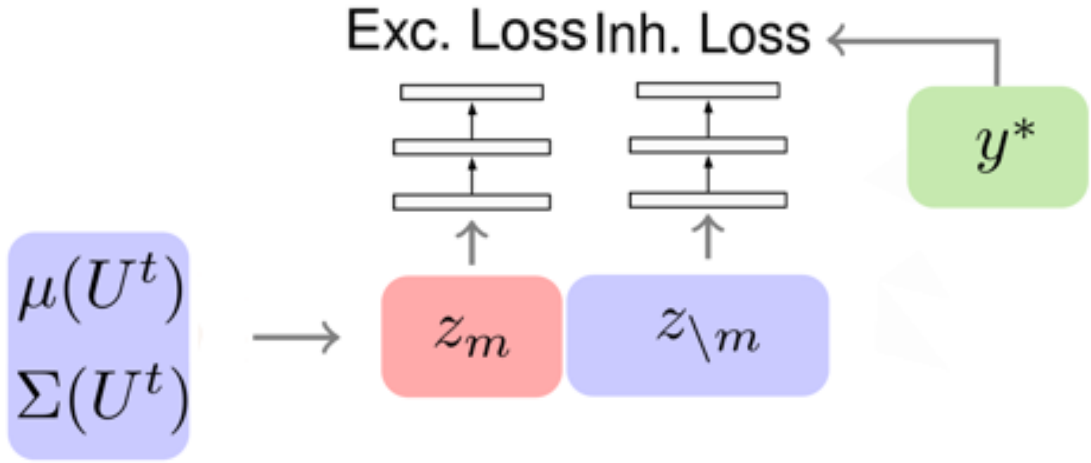
- Encoder SNN trainable through gradient descent
- Convolutional SNN Layers encode spatio-temporal streams into a latent space
- SNN can bridge computational time scales by extracting slow and relevant factors of variation from fast event streams recorded by the DVS
- Reconstructed TS *x** is equivalent to pre-synaptic trace *Qt*
- LIF neuron model with time step Δt
- Training done on GPU



Hybrid Guided Variational Auto-Encoder

Guiding Adversarial Classifiers

- VAEs do not necessarily disentangle all factors of variation
- Need a disentangled, interpretable latent space
- Supervised Guided-VAE trains subset of latent variables to encode ground-truth labels
- Remaining latent variables uncorrelated with the label



$$\mathcal{L}_{Exc}(z,m) = \max_{c_m} \left(\sum_{n=1}^{N} \mathbb{E}_{q(z_m|x_n)} \log p_{c_m}(y = y_m(x_n)|z_m) \right),$$

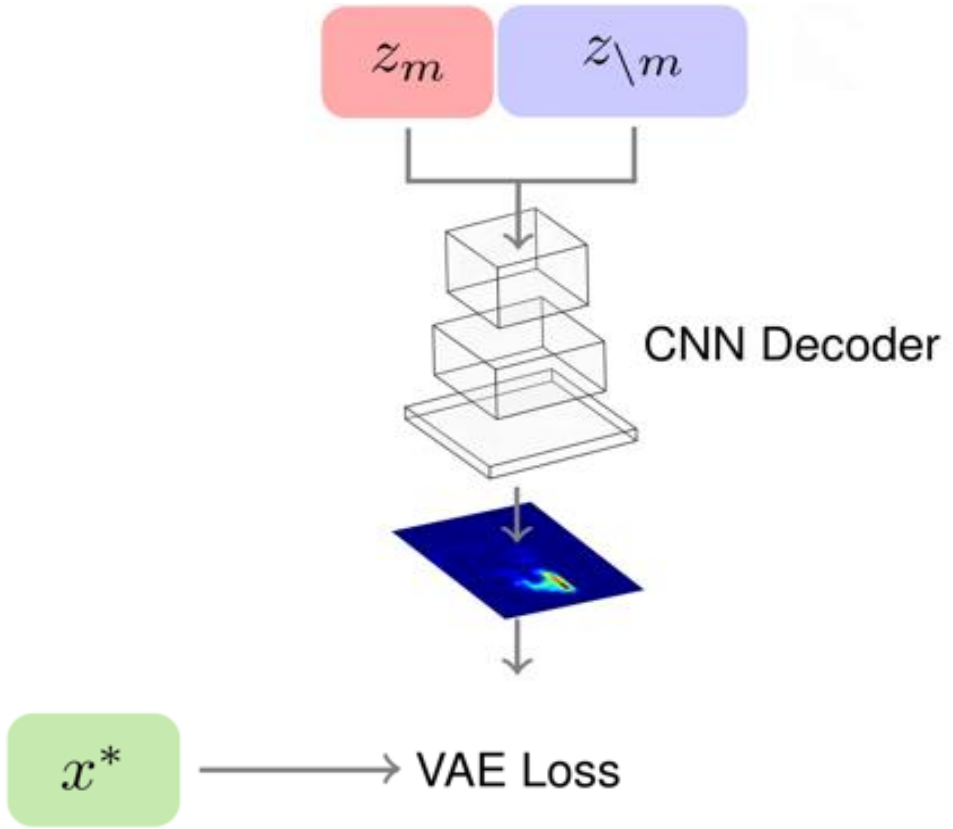
$$\mathcal{L}_{Inh}(z,m) = \max_{k_m} \left(\sum_{n=1}^{N} \mathbb{E}_{q(z_{\setminus m}|x_n)} \log p_{k_m}(y = y_m(x_n)|z_{\setminus m}), \right)$$

Hybrid Guided Variational Auto-Encoder

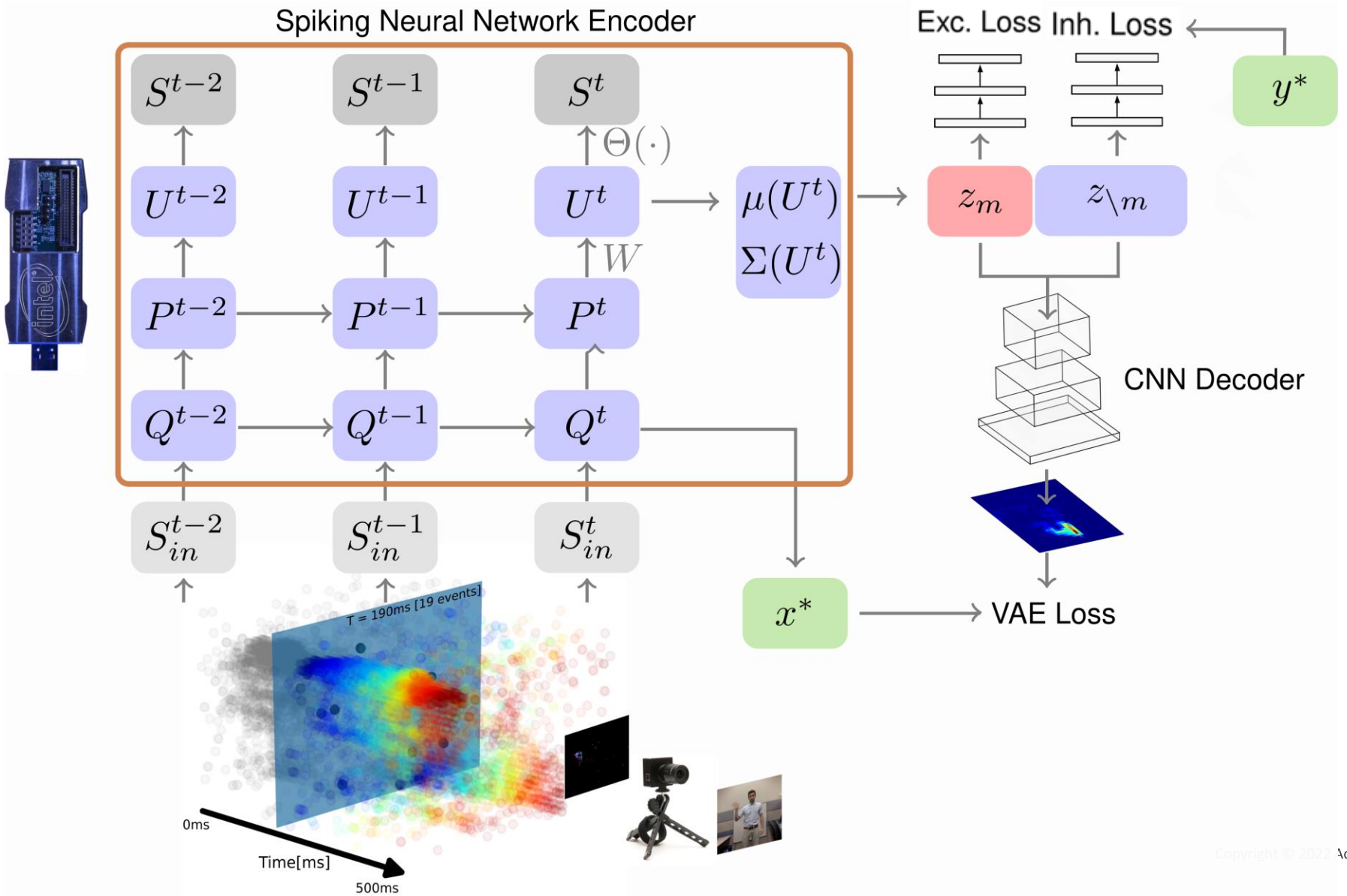
Non-Spiking CNN Decoder

- Non-spiking decoder
- Only interested in latent structure produced by encoder, rather than the generative features of the network
- Dedicated neuromorphic processor only requires the encoder to produce latent structure
- More resources can be dedicated to SNN encoder
- Reconstructs TS, use TS in reconstruction loss

$$\log p(x) \geq \underbrace{\mathbb{E}_{z\sim q} \log p(x|z)}_{\mathcal{L}_{ll}} - \underbrace{D_{KL}(q(z|x)||p(z))}_{\mathcal{L}_{prior}}$$

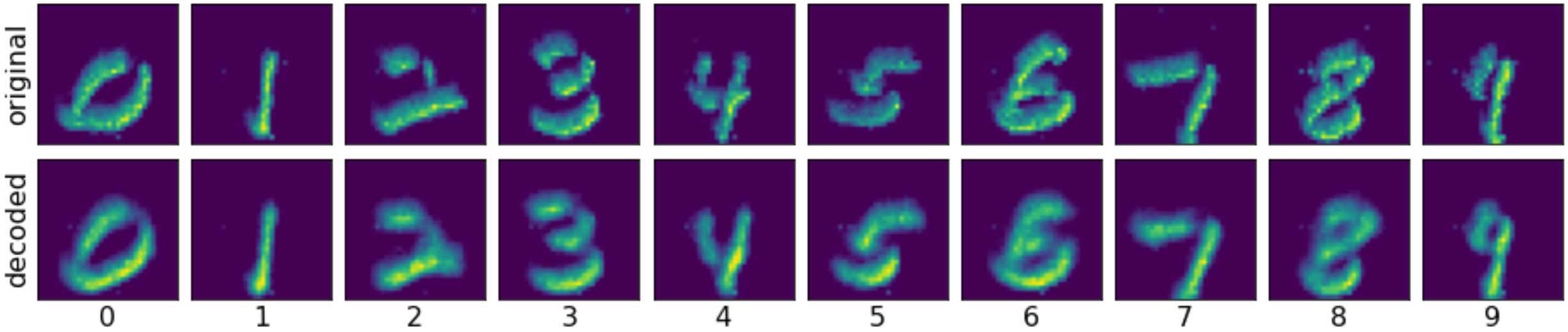


Hybrid Guided Variational Auto-Encoder



NMNIST Dataset

Comparison of original TS with reconstructed TS



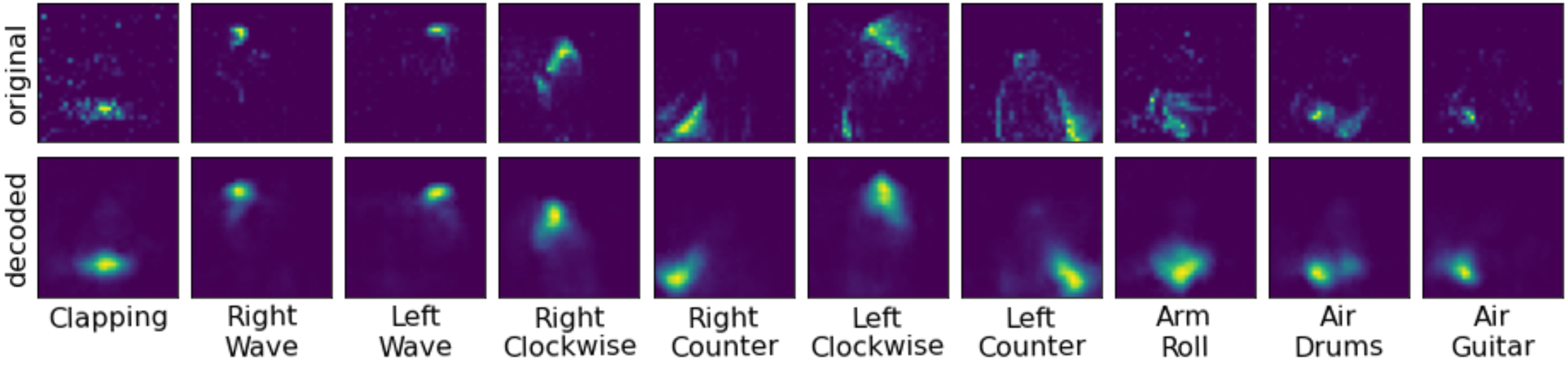
NMNIST Latent Space T-SNE

- T-SNE to visualize the learned representations and disentanglement of the classes
- T-SNE embeds both the local and global topology of the latent space into a two-dimensional space for visualization
- Each digit representation, coded by color, is clearly disentangled and separable in the latent space

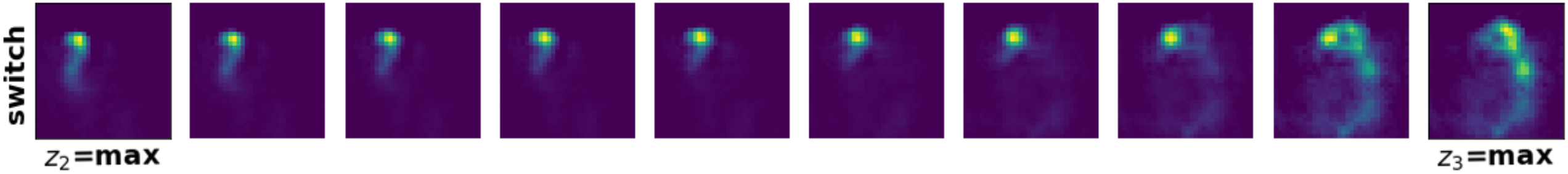
Labeling Unlabeled Gestures

DVSGesture Dataset

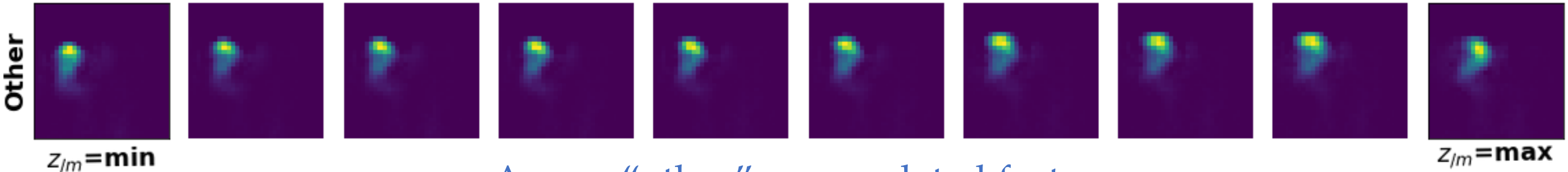
Comparison of original TS with reconstructed TS



Latent Space Traversal



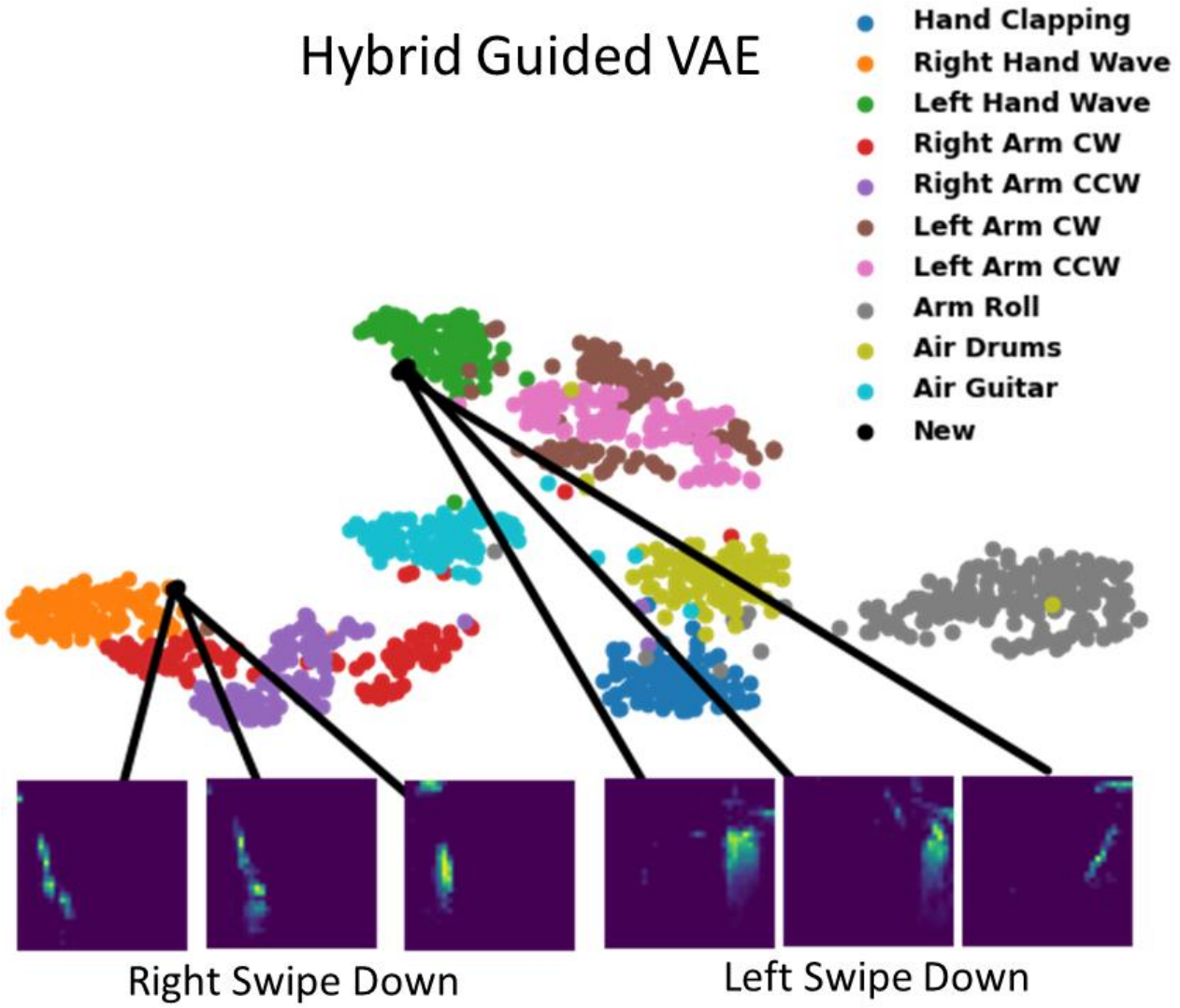
Between two classes



Across “other” uncorrelated factors

Labelling Unlabelled Gestures

- To test generalization of the learned encoder, evaluated how the VAE model performs when provided with new gesture data captured in a new environment
- Recorded gestures belonging to two classes not present in the DVSGesture dataset

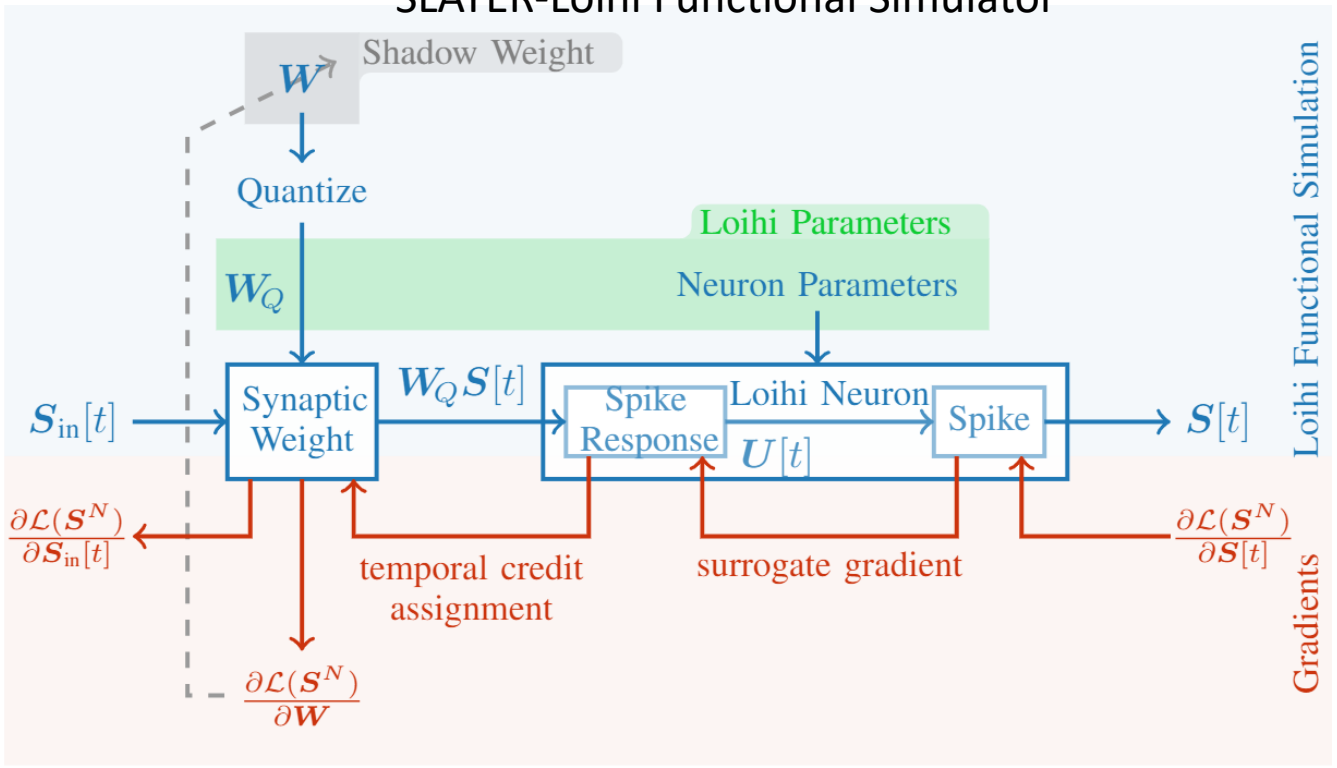


Neuromorphic Implementation

SLAYER Pre-Training

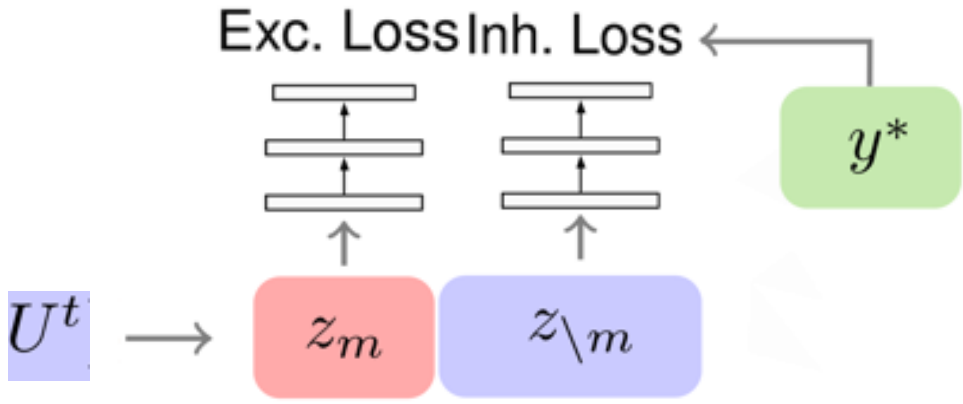
- To train with the same neuron model and quantization as the Loihi SLAYER was used
- SLAYER has a differentiable functional simulator of the Loihi chip for one-to-one mapping of trained networks onto hardware

SLAYER-Loihi Functional Simulator



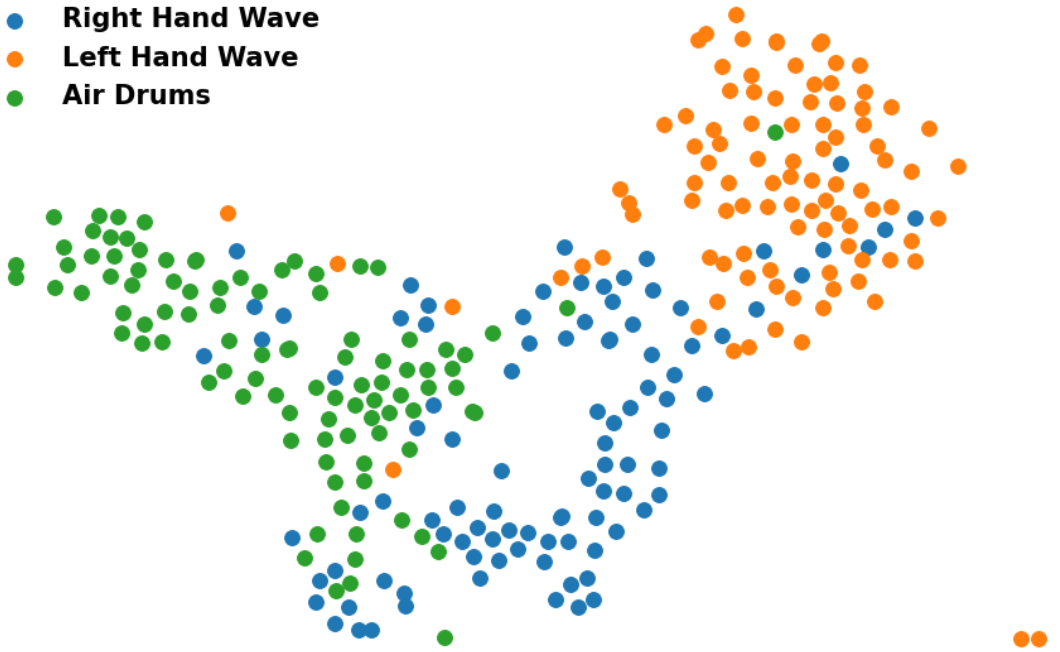
SLAYER Membrane Potential Encoding

- The mean and variance of the network were made spiking
- Uses the quantized membrane potential of the neuron for the latent representation instead of ANN trained full precision values
- The network can be mapped to the Loihi for inference
- Not necessary to map the decoder to the Loihi



Loihi Encoder Inference

- T-SNE of the latent space representation of three gesture classes using the encoder mapped onto the Loihi
- Three classes are separable
- No clear separation with all classes
- Could be due to the low-precision integers used for synaptic weights and membrane potentials



Conclusion and Future Work

Contributions

1. End-to-end trainable event-based SNNs for processing neuromorphic sensor data event-by-event and embedding them in a latent space.
2. A Hybrid Guided-VAE that encodes event-based camera data in a latent space representation of salient features for clustering and pseudo-labeling.
3. A proof-of-concept implementation of the Hybrid Guided-VAE on Intel's Loihi Neuromorphic Research Processor.

Future Work

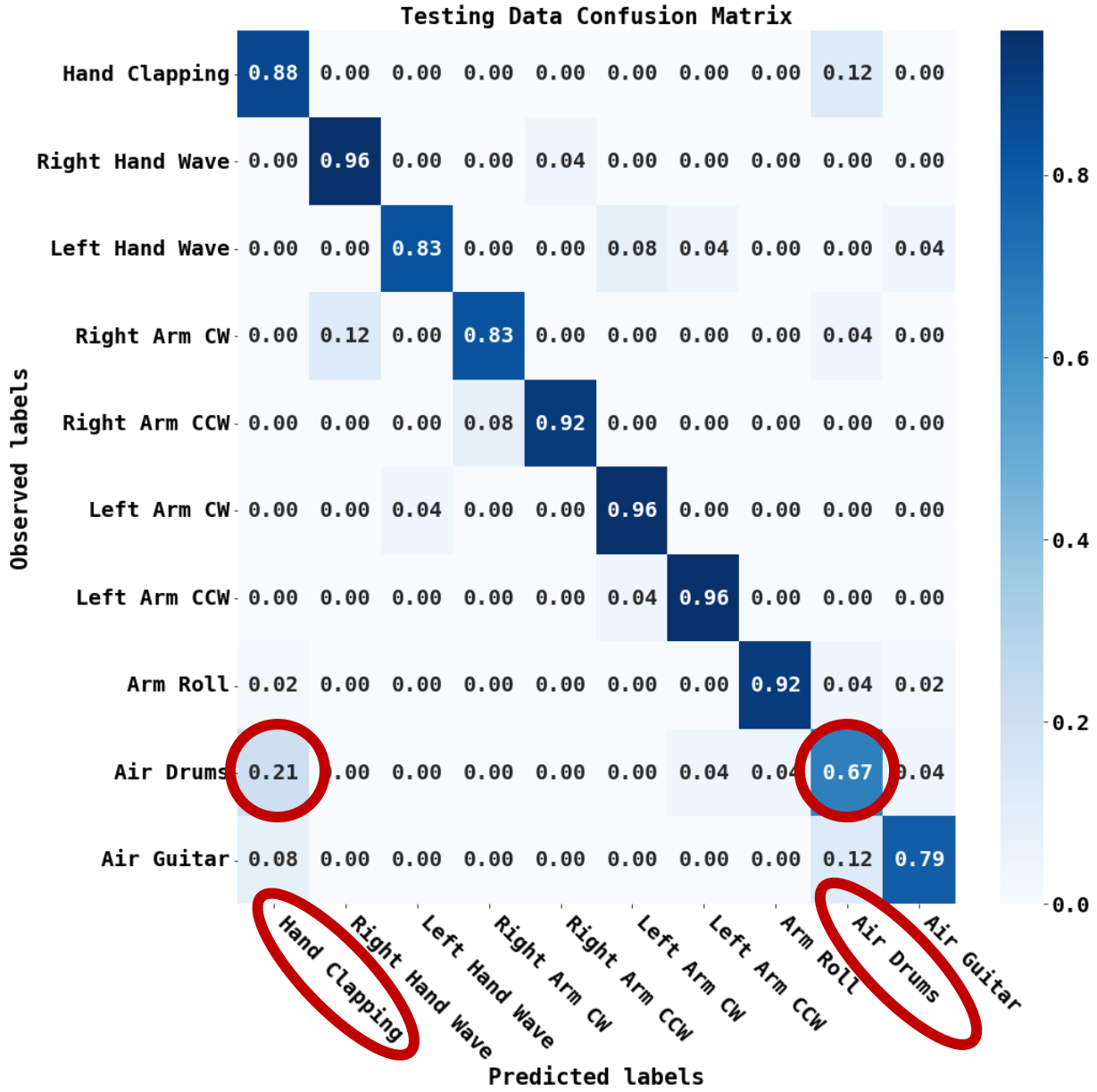
- Address limitations of neuromorphic hardware model
- Improve disentanglement of classes
- Add online learning of unlabeled data with the SNN encoder on hardware
- Create demonstration of self-supervised online learning
- Try method with other types of sensor data such as EMG

Thank You

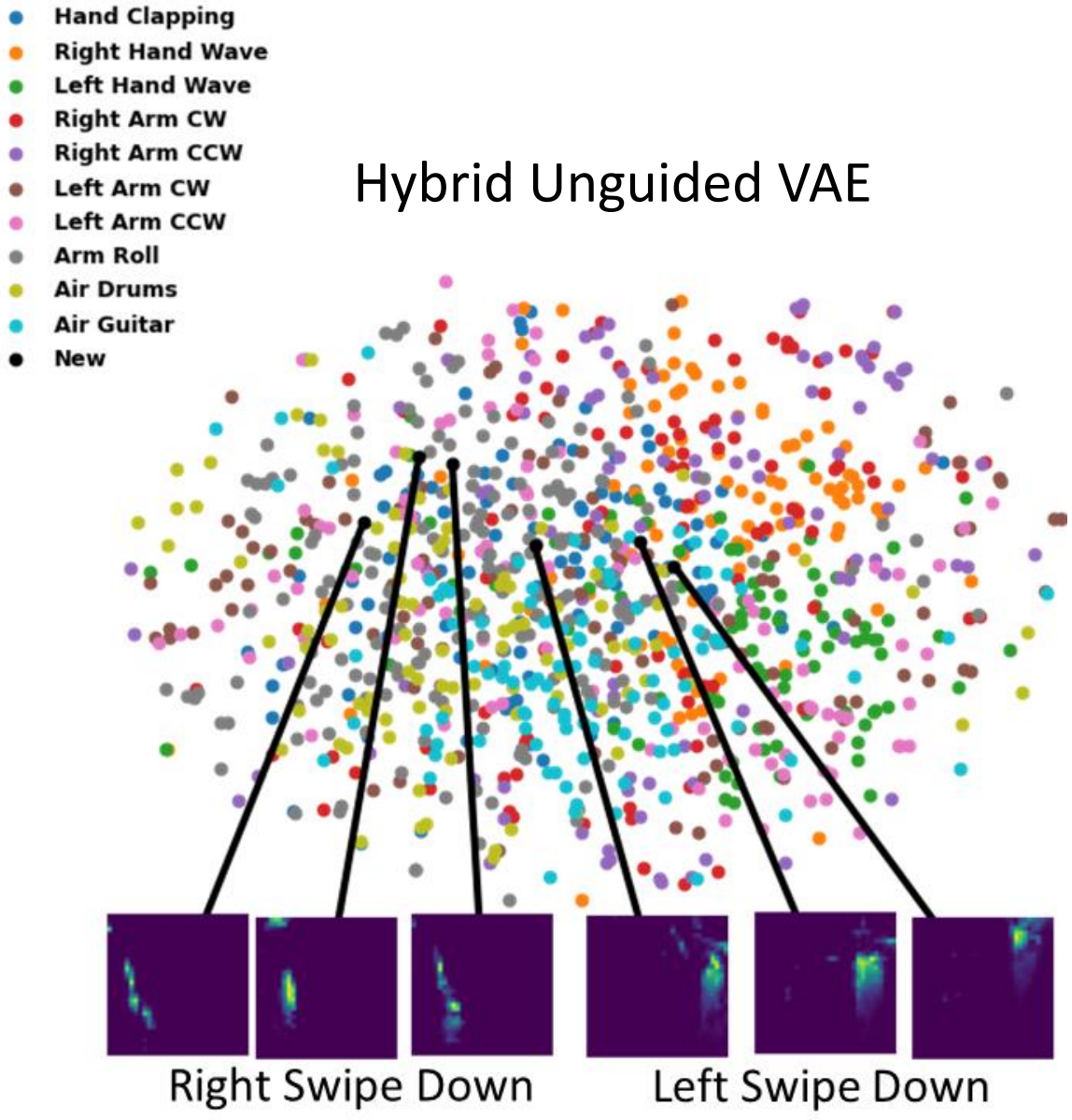
Questions?

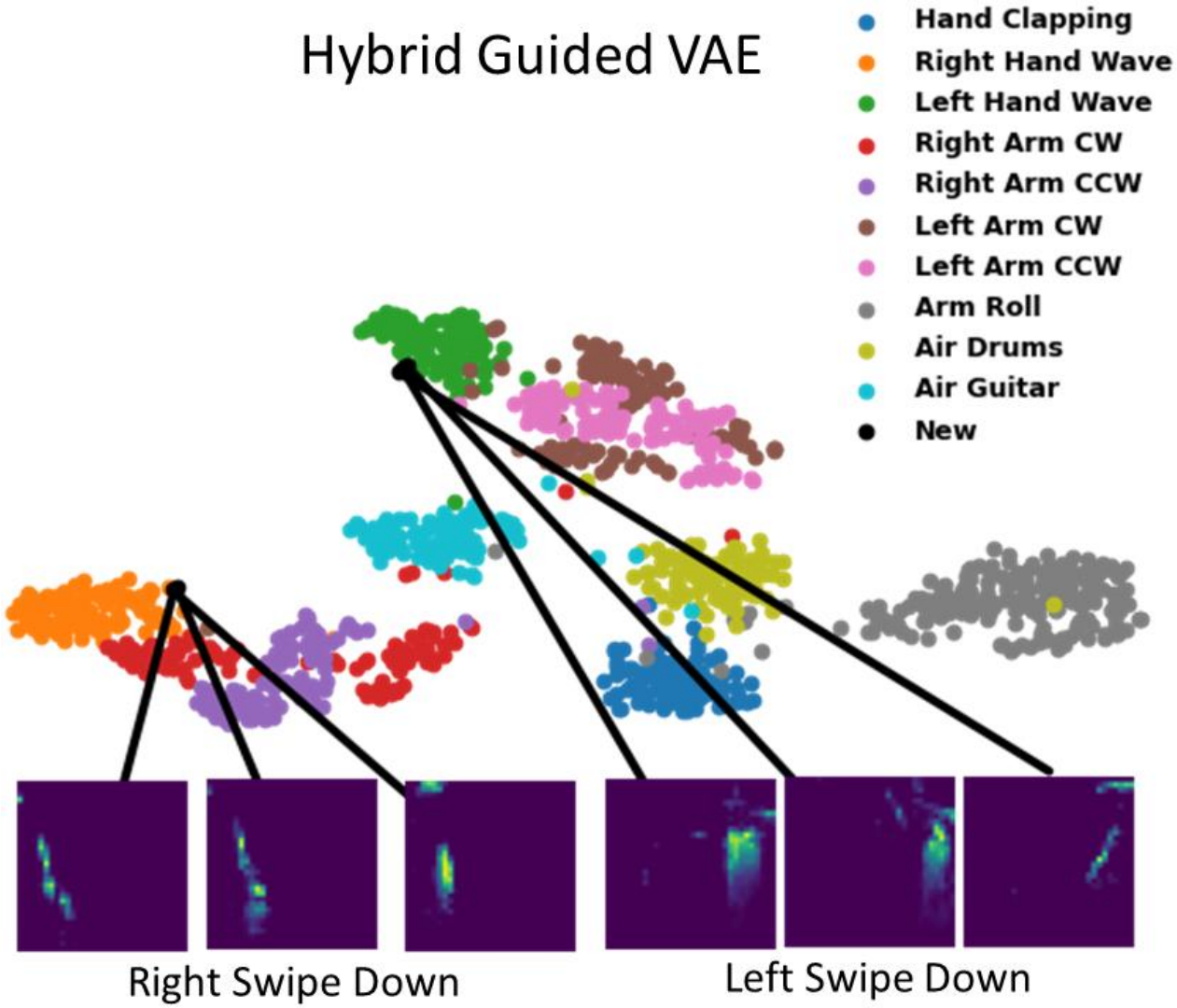
Latent Space Confusion Matrix

- Certain very similar gestures are confused, such as Right and Wave and Right Arm Clockwise or Air Drums with Hand Clapping

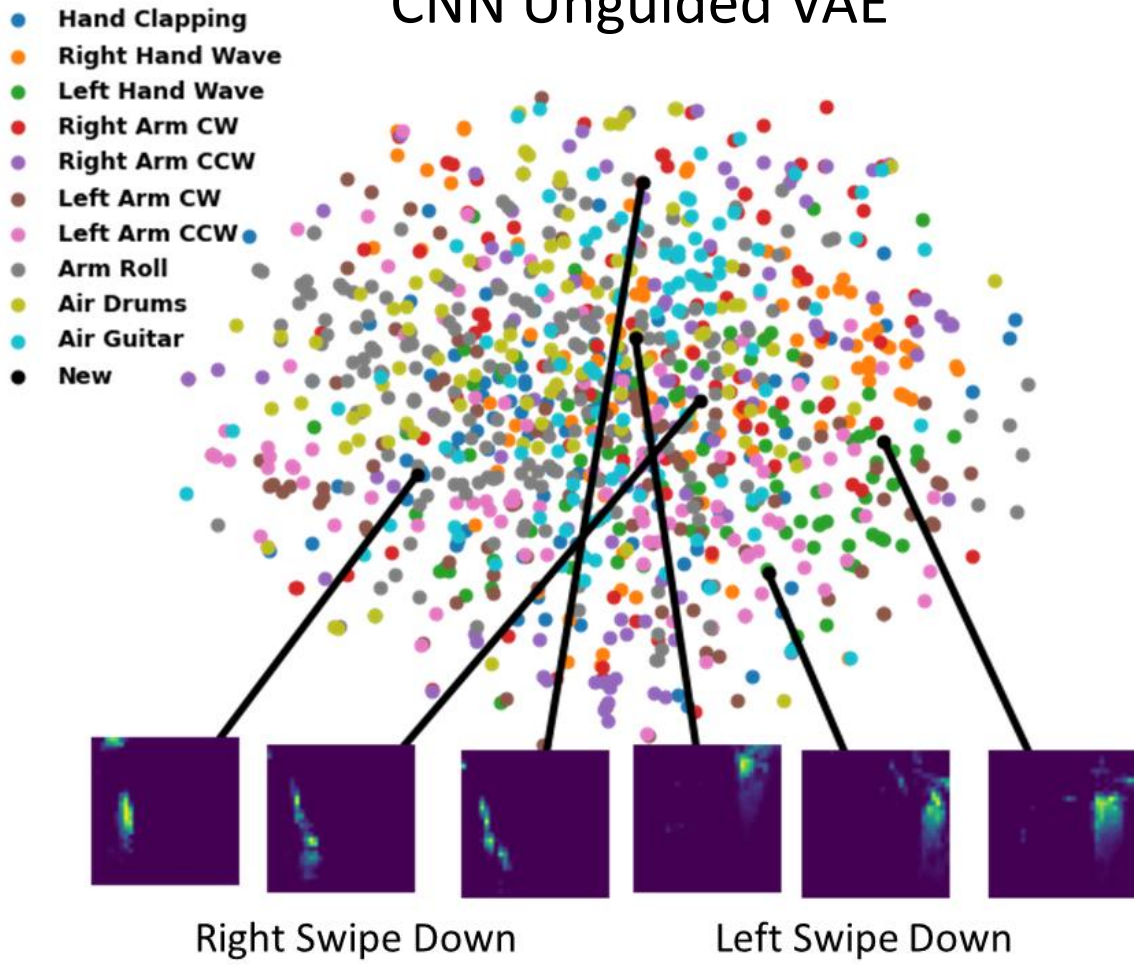


Ablation Study

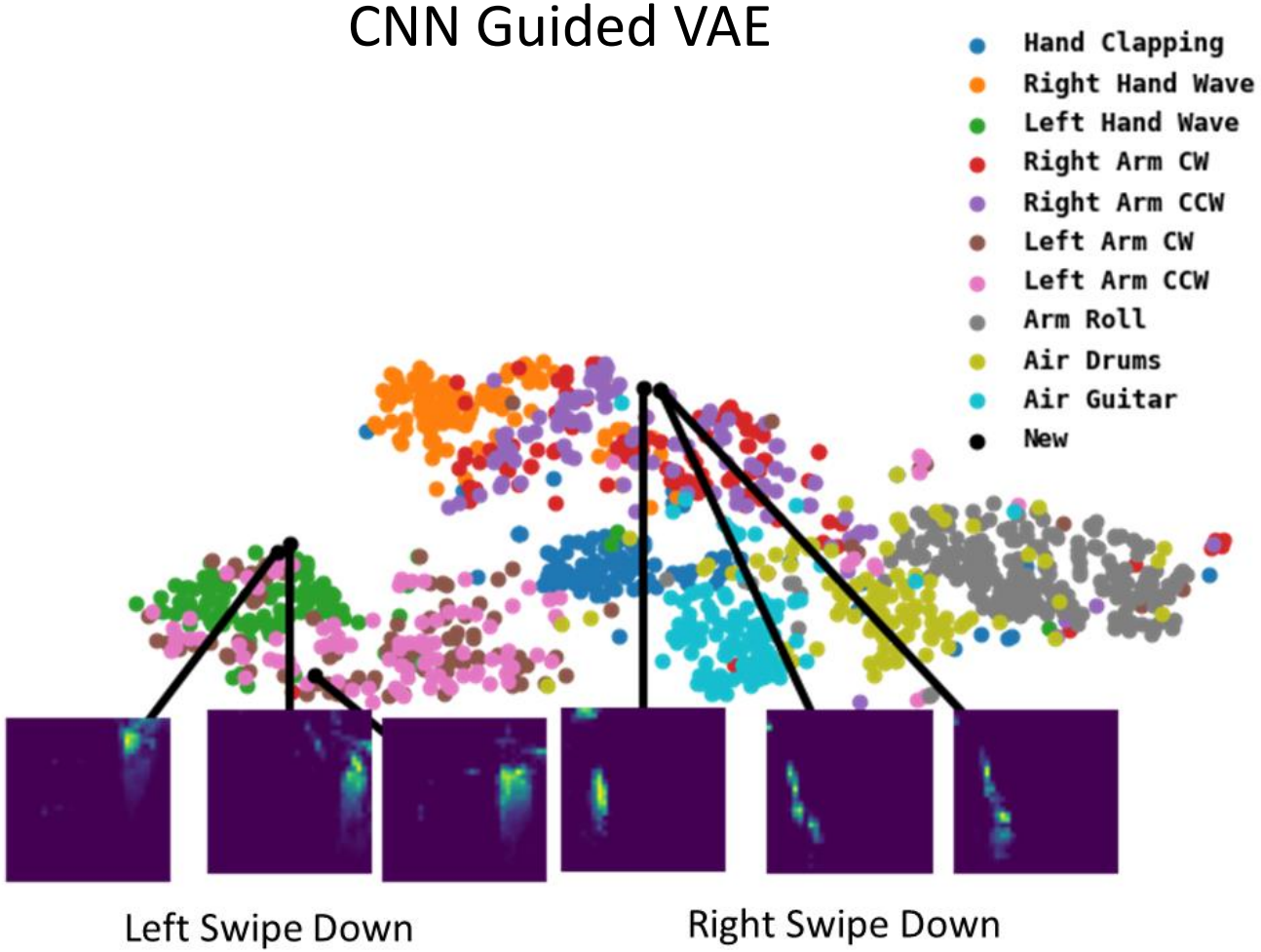




CNN Unguided VAE



CNN Guided VAE



VAE Comparison to Classifier Output

- T-SNE visualization of the features learned by the convolutional layers of DECOLLE and SLAYER models
- Features learned by the models do not clearly disentangle classes
- New gestures are not clearly clustered