Caltech

Dance Generation Using Discrete Latent Variables

Yoojin Chung, Alveera Khan, Daniel Kyme, Brandon Quach, Ruoyun Zheng

Introduction

Typical video generation techniques involve large amounts of training data and computation. We examine the benefits of reducing the data size by training on a discrete latent space in regards to the generation of dancing figures, a task typically involving large amounts of preprocessing (i.e. pose estimation) and data.
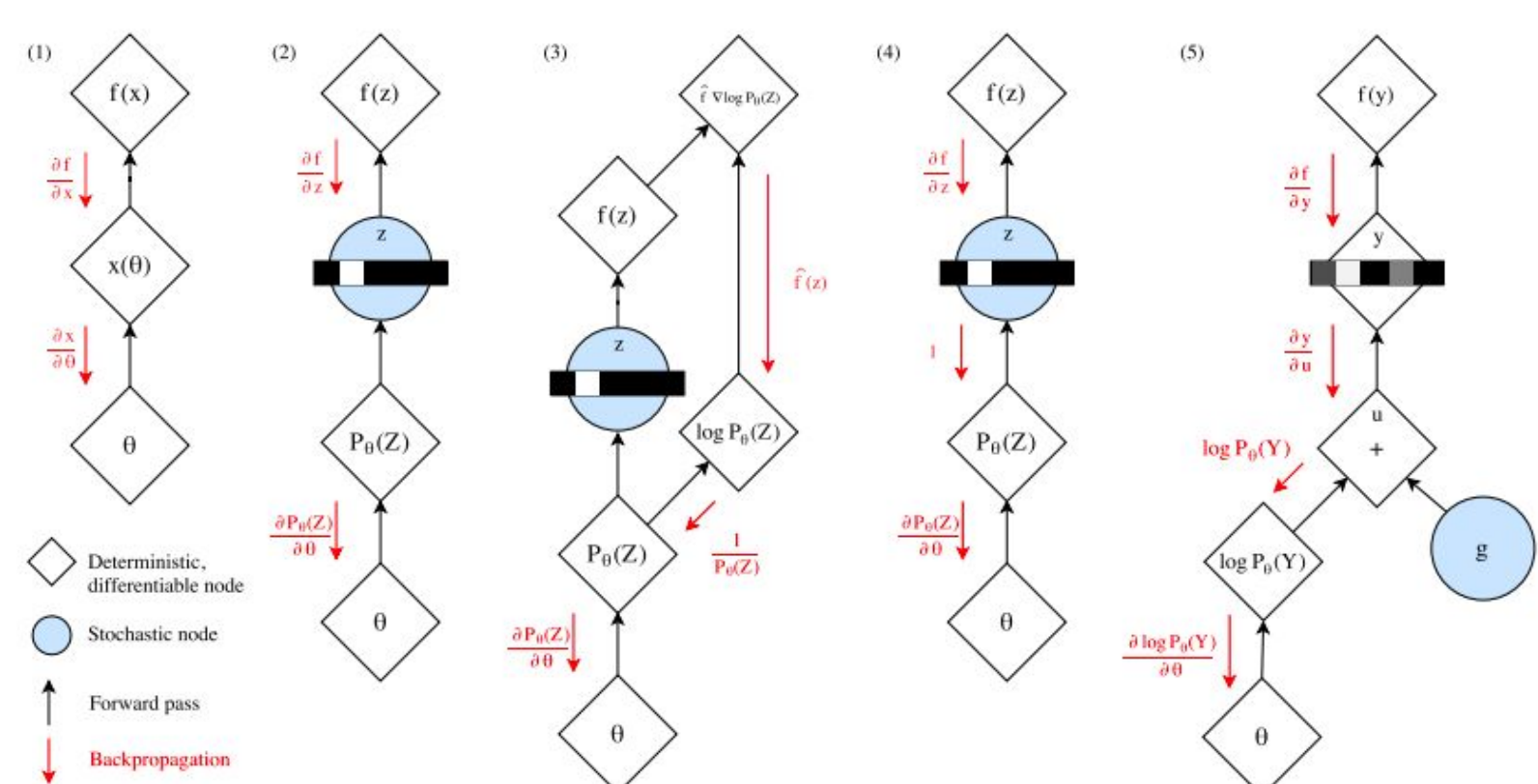
Approach

The typical method involves the use of a VAE + RNN combination (sampling on approximate posterior). Our approach was to instead just use a VAE, getting rid of the computationally intensive RNN. Thus, generation of a sequence was done by continuously sampling on the prior.

Model:

t_1 t_2 t_3 t_4 t_5 t_6 Picture Frames (120X208)

l_1 l_2 l_3 l_4 l_5 l_6 Latent Variables (80X240)
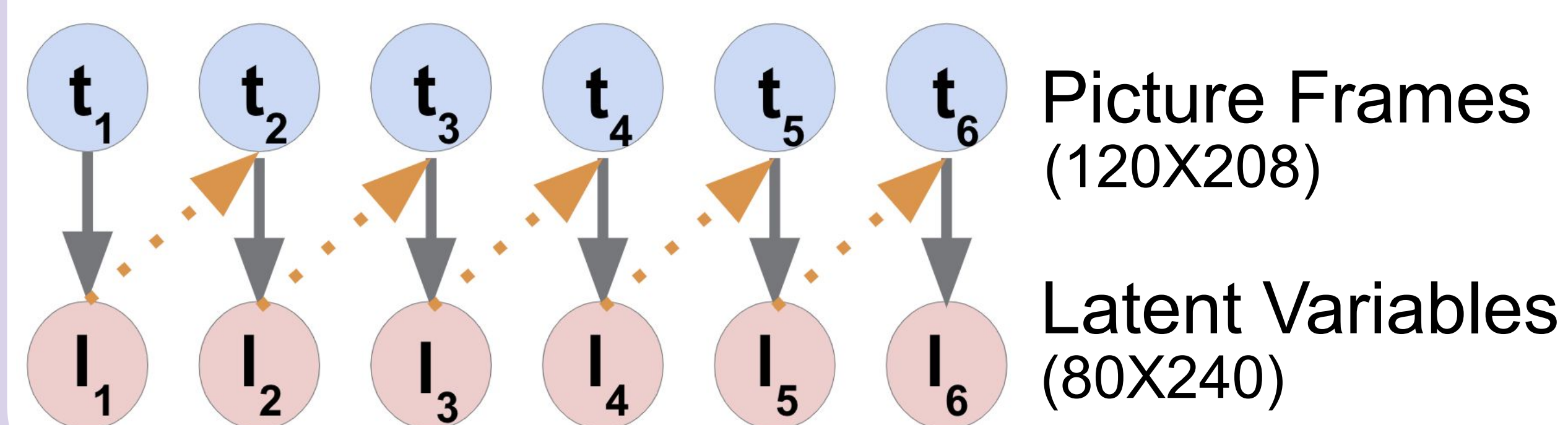
Gumbel Softmax:

Output

Gumbel VAE

Fig 1. Seed image (frame 0)

Fig 2. Generation (frame 1)
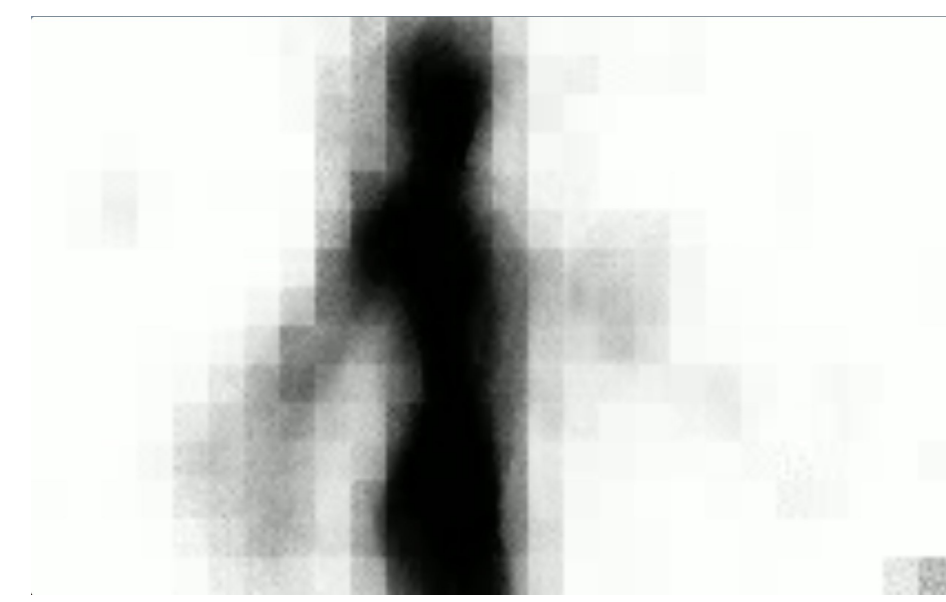
Fig 3. Evolution (frame 2)

Fig 4. Final Evolution (frame 16)

Loss for Gumbel VAE

GAN

Losses for GAN
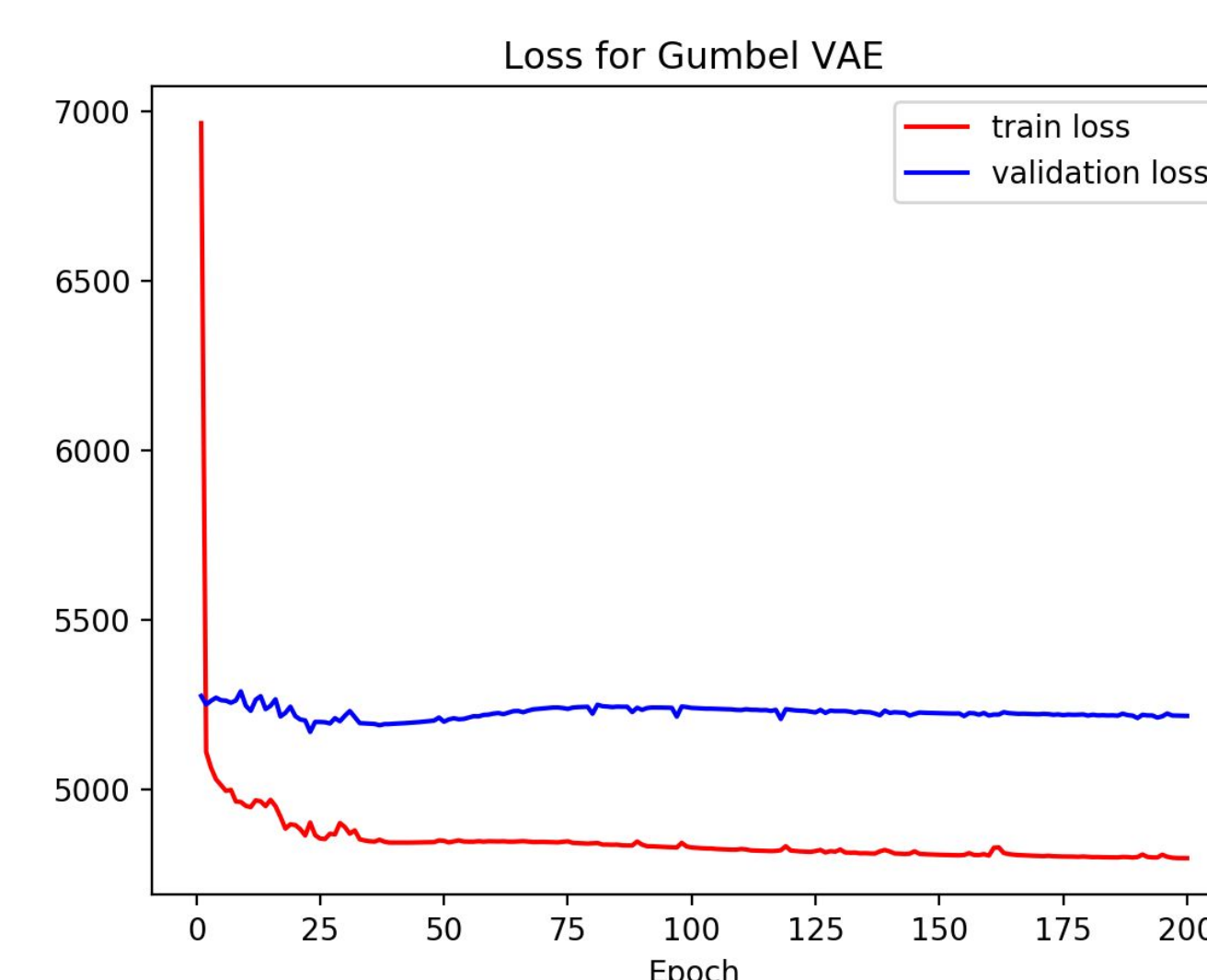
Accuracy for GAN
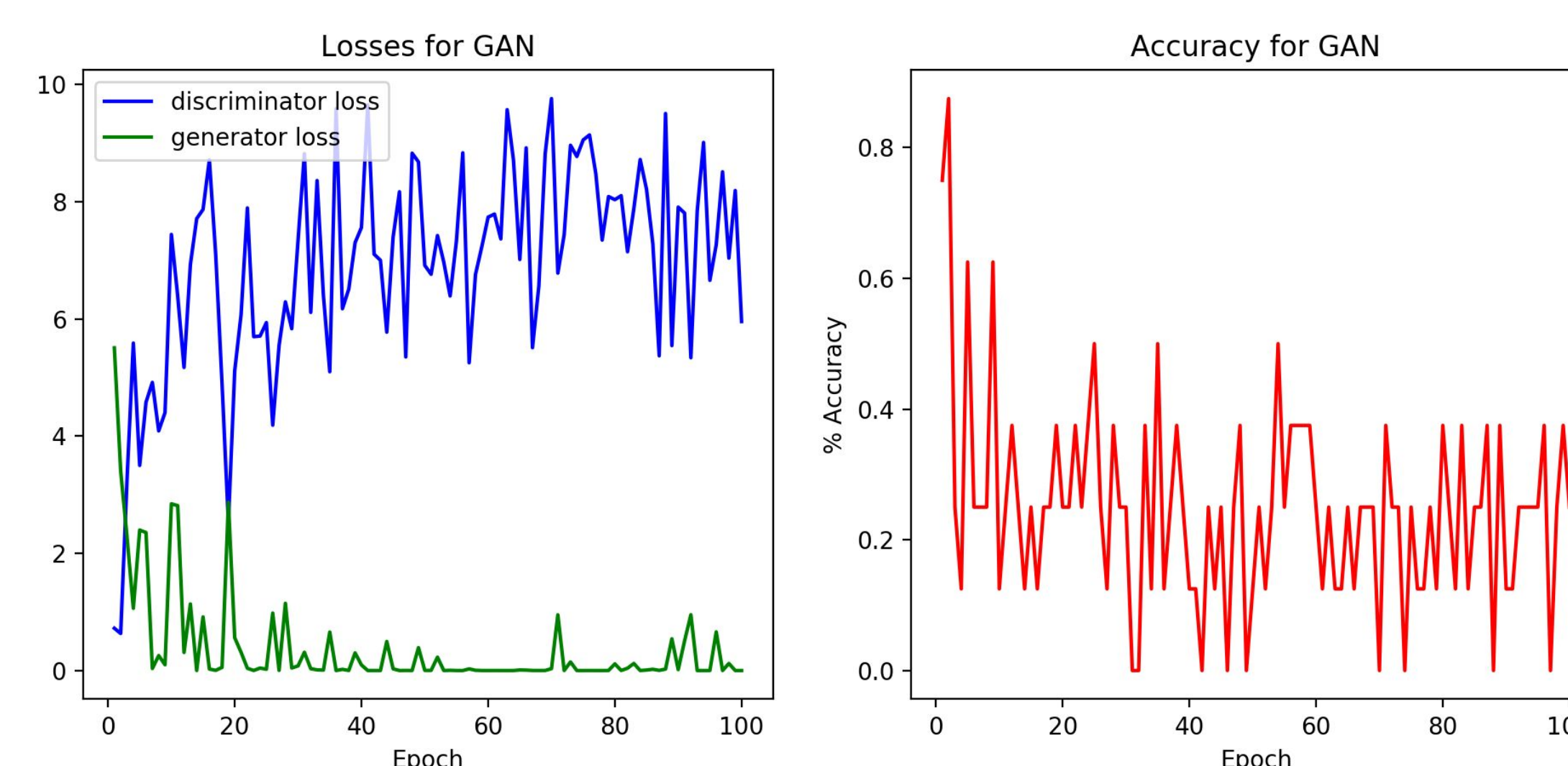
Related Work & Applications

In conjunction with producing our discrete VAE model using the Gumbel-Softmax distribution, we tested different deep generative models, such as Generative Adversarial Networks (GANs). GANs produced video output despite the small sample size and training time given to the model, but was not further pursued due to the intense amount of data and GPU power required to train.

Related models include Variational Recurrent Neural Networks (VRNN) and Long Short Term Memory Autoencoders (LSTM). VRNNs go a step further than VAEs and introduce a third layer to represent the sequences and can be effective in learning sequential data. LSTM Autoencoders can learn the dynamics of temporal ordering of sequences and use a "memory" to recall information across long sequences.

Specific to dancing, the model provides a better understanding of human movement and has the potential to produce new ways to move and thus dance. Outside of dancing, the idea behind the model has applications in taking sequential, repeated data and finding new approaches to fit into an overall sequence.

Conclusion

Through a discrete VAE model that uses a Gumbel-Softmax distribution, we were able to generate a sequence of human dance moves. This gave insights on how to improve the current state of generative human poses and potentially utilize the inherently discrete parts of the human body.

Future works will include training for longer periods of time with more data points, as well as a larger attempt to use GANs.

References

1. Generative Choreography using Deep Learning: arXiv:1605.06921 [cs.AI]
2. Everybody Dance Now: arXiv:1808.07371v1 [cs.GR]
3. Categorical Reparameterization with Gumbel-Softmax: arXiv:1611.01144 [stat.ML]
4. Neural Variational Inference and Learning in Belief Networks: arXiv:1402.0030 [cs.LG]
5. DanceNet: GitHub [https://github.com/jsn5/dancenet/]