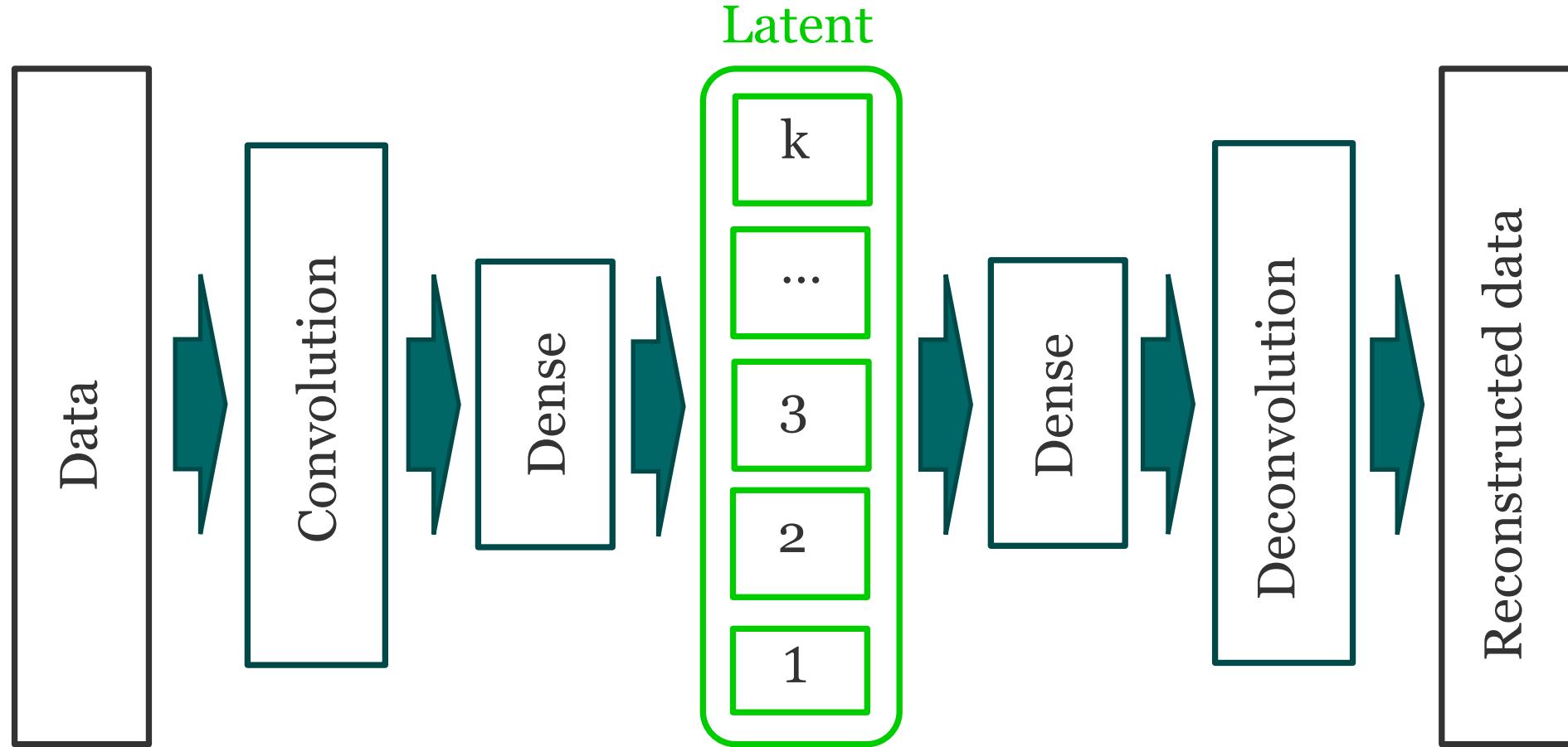


# Lecture 29: Variational Autoencoders

Sergei V. Kalinin

# Autoencoders



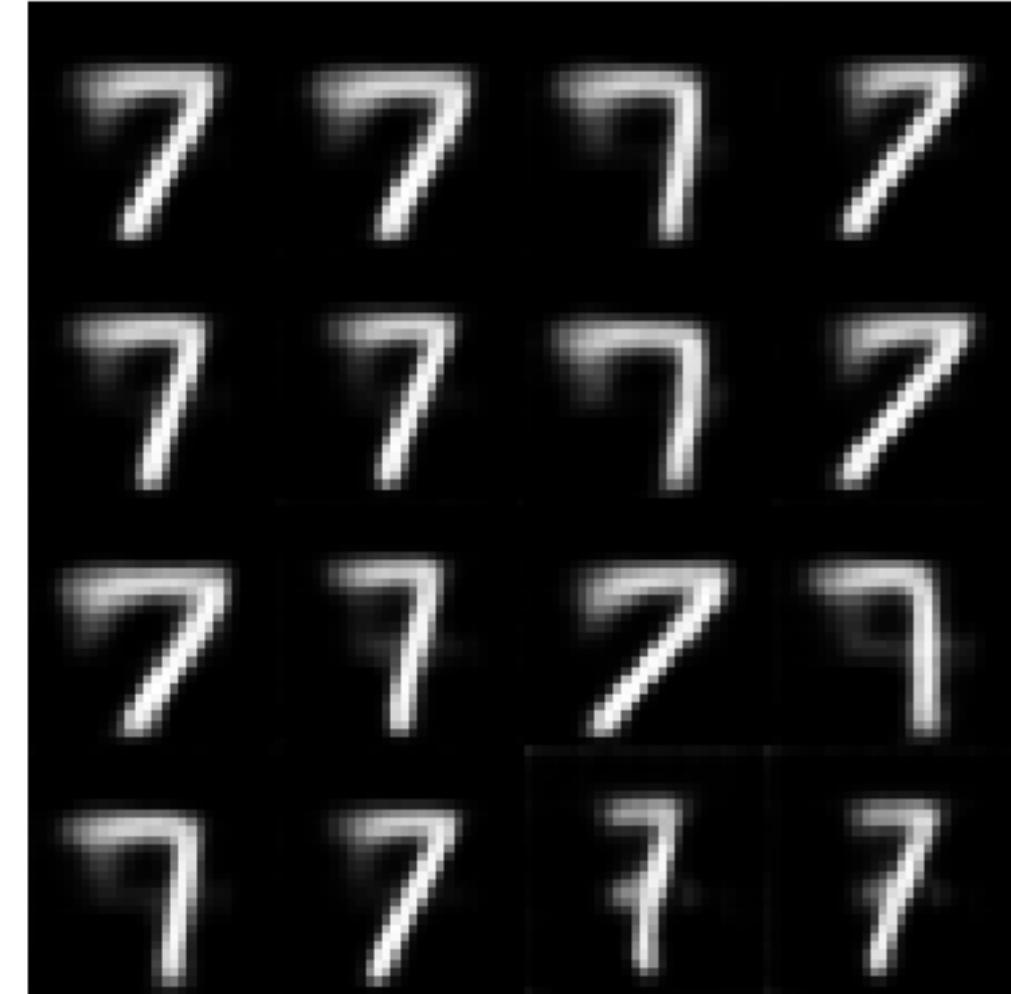
**Loss:** reconstruction loss

# The AE reconstructs data

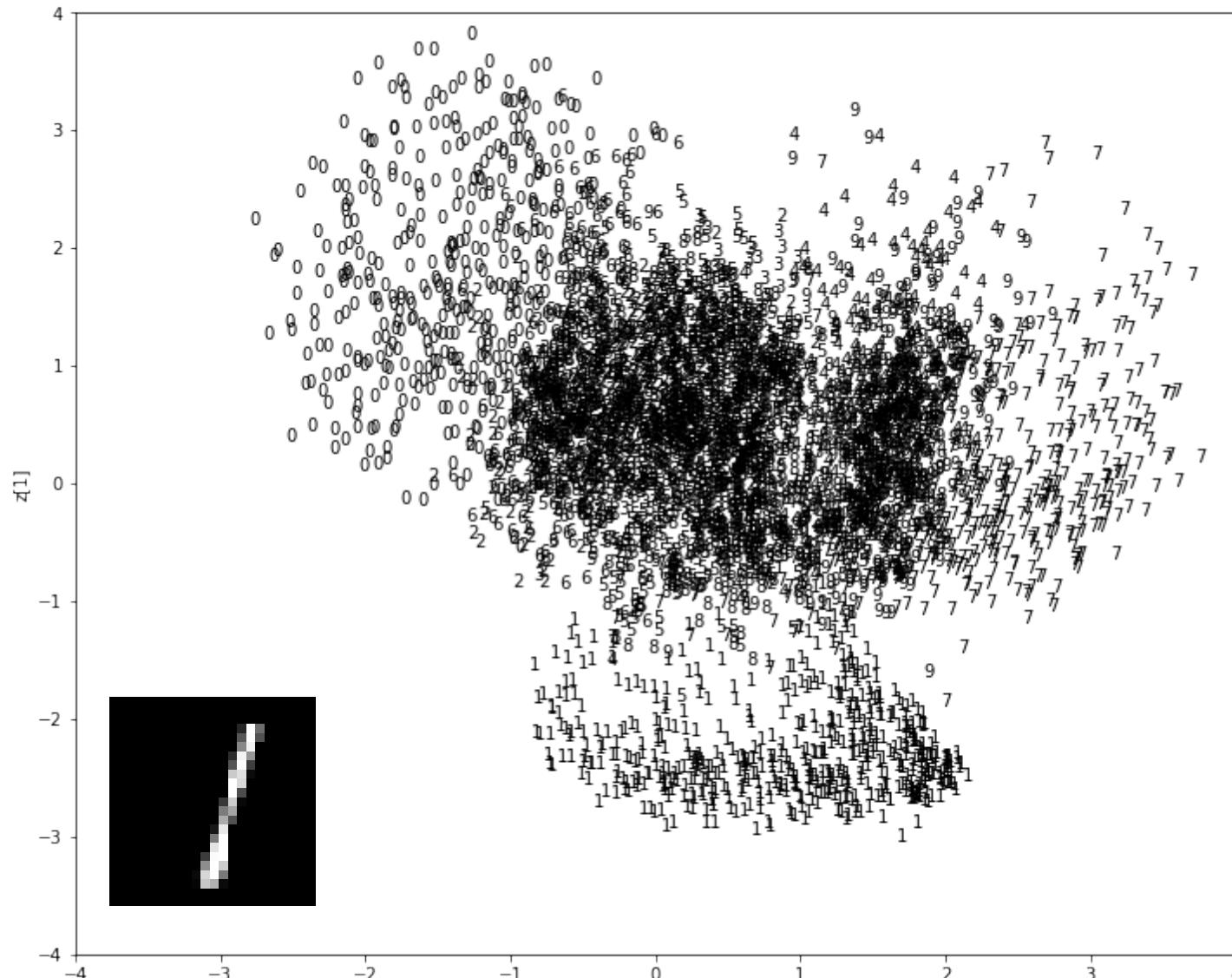
Input data



Decoded data

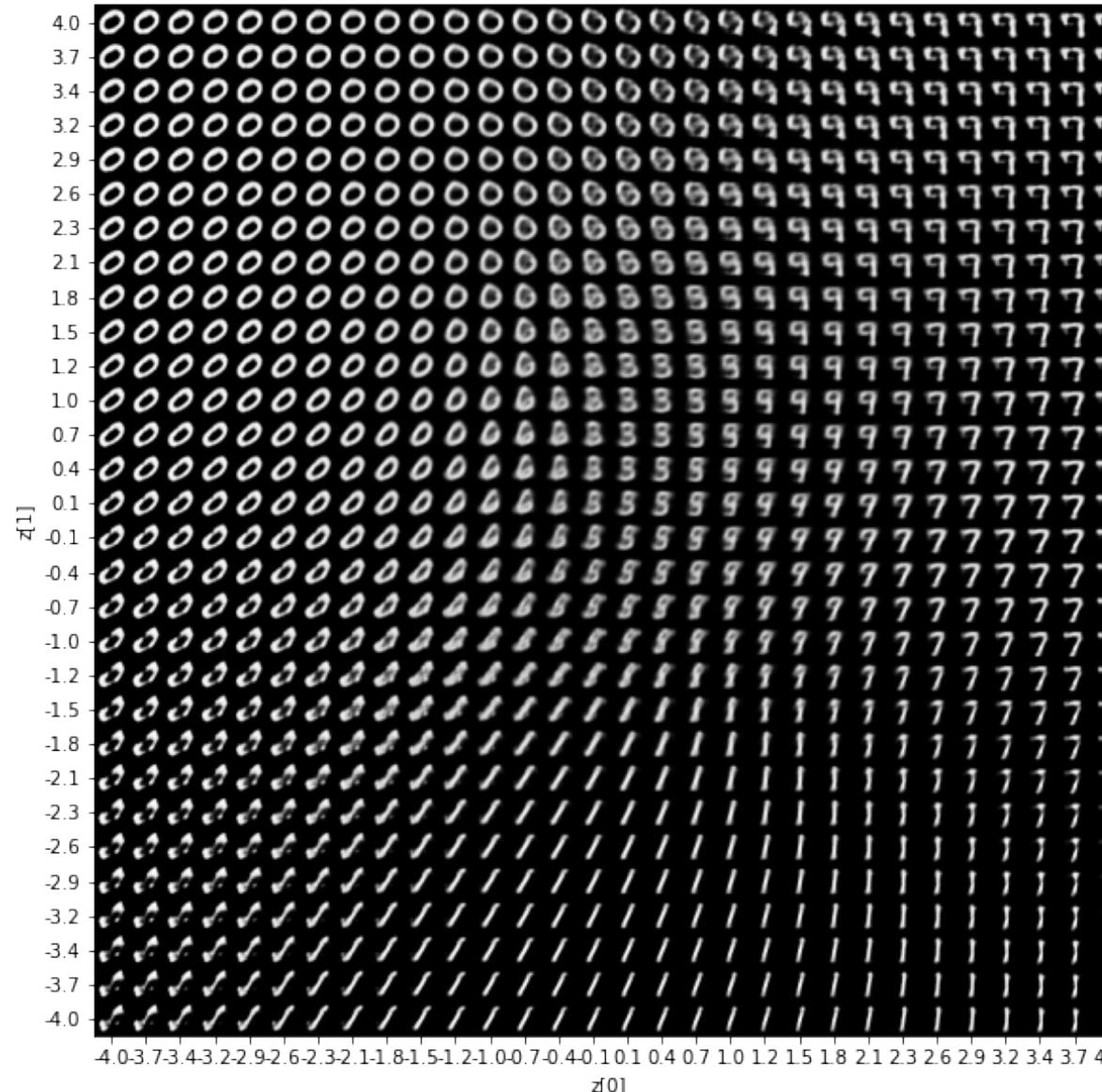


# Encoding: Image → Latent Space



**Latent distribution:** Encoding the data via low dimensional vector

# Decoding: Latent Space → Image



**Latent representation:** Decoding images from uniform grid in latent space

# Image Reconstruction

Test color images (Ground Truth)



Test gray images (Input)



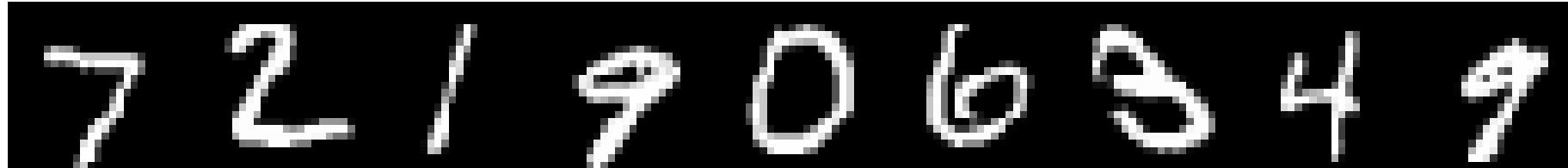
Colorized test images (Predicted)



- **Training:** pairs of the grayscale and color images
- **Application:** new grayscale images (from the same distribution)
- **Concern:** has to be from the same distribution

# Image Denoising

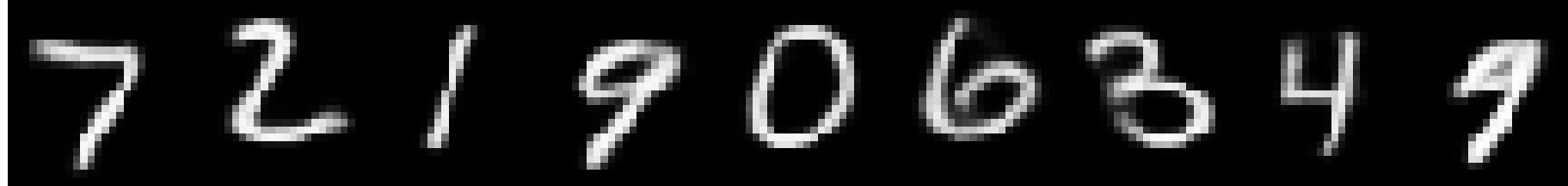
Ground truth



Noisy input



Reconstruction



- **Training:** pairs of the high-noise and low-noise images
- **Application:** new high noise images (from the same distribution)
- **Concern:** has to be from the same distribution

# Variational Autoencoders



Diederik P. Kingma

Other names ▾

 FOLLOW

Research Scientist, [Google Brain](#)  
Verified email at google.com - [Homepage](#)

Machine Learning Deep Learning Neural Networks Generative Models Variational Inference

TITLE	CITED BY	YEAR
<a href="#">Adam: A Method for Stochastic Optimization</a> DP Kingma, J Ba Proceedings of the 3rd International Conference on Learning Representations ...	141306	2014
<a href="#">Auto-Encoding Variational Bayes</a> DP Kingma, M Welling arXiv preprint arXiv:1312.6114	26540	2013
<a href="#">Semi-Supervised Learning with Deep Generative Models</a> DP Kingma, S Mohamed, DJ Rezende, M Welling Advances in Neural Information Processing Systems, 3581-3589	2946	2014

- Variational Autoencoder (VAE): uses “reparameterization trick” to sample from the latent space
- Can be used for same tasks as AE
- Have a much better-behaved latent space: **disentanglement of the representations**

# VAE Training

Latent manifold → Image space

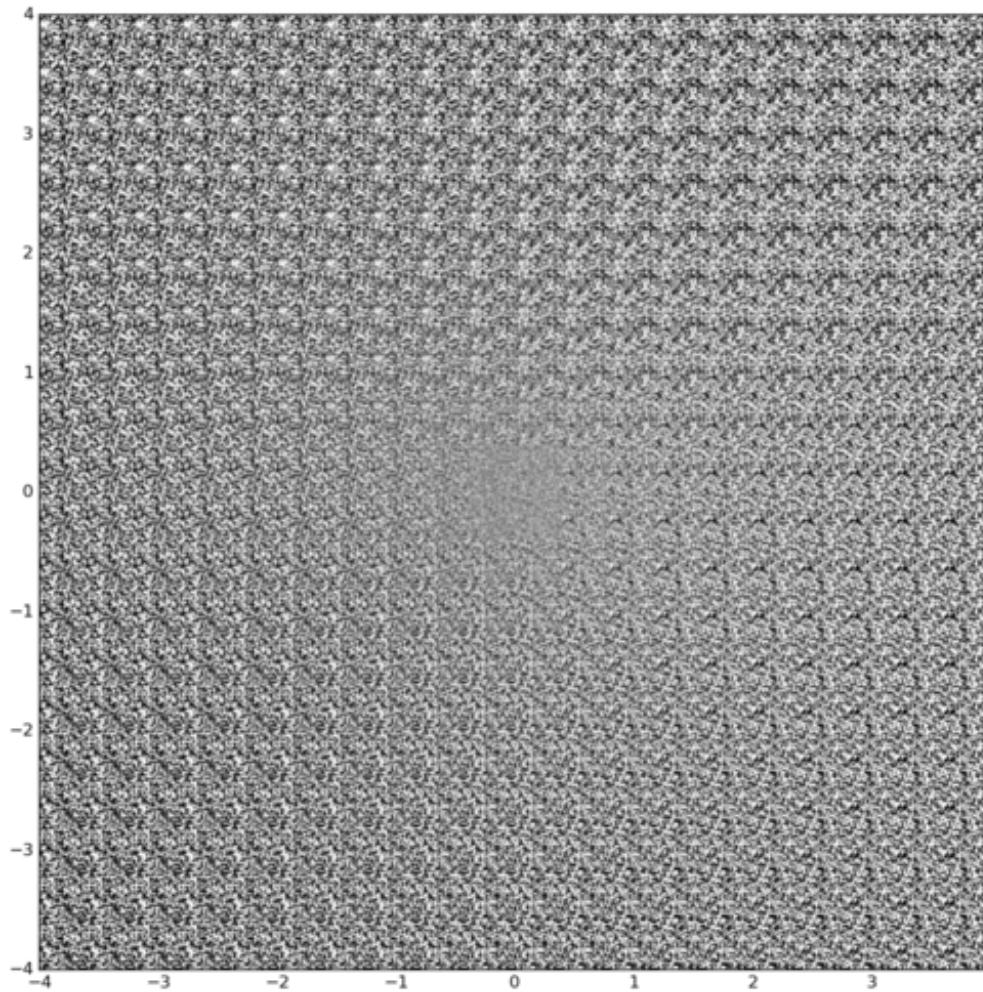
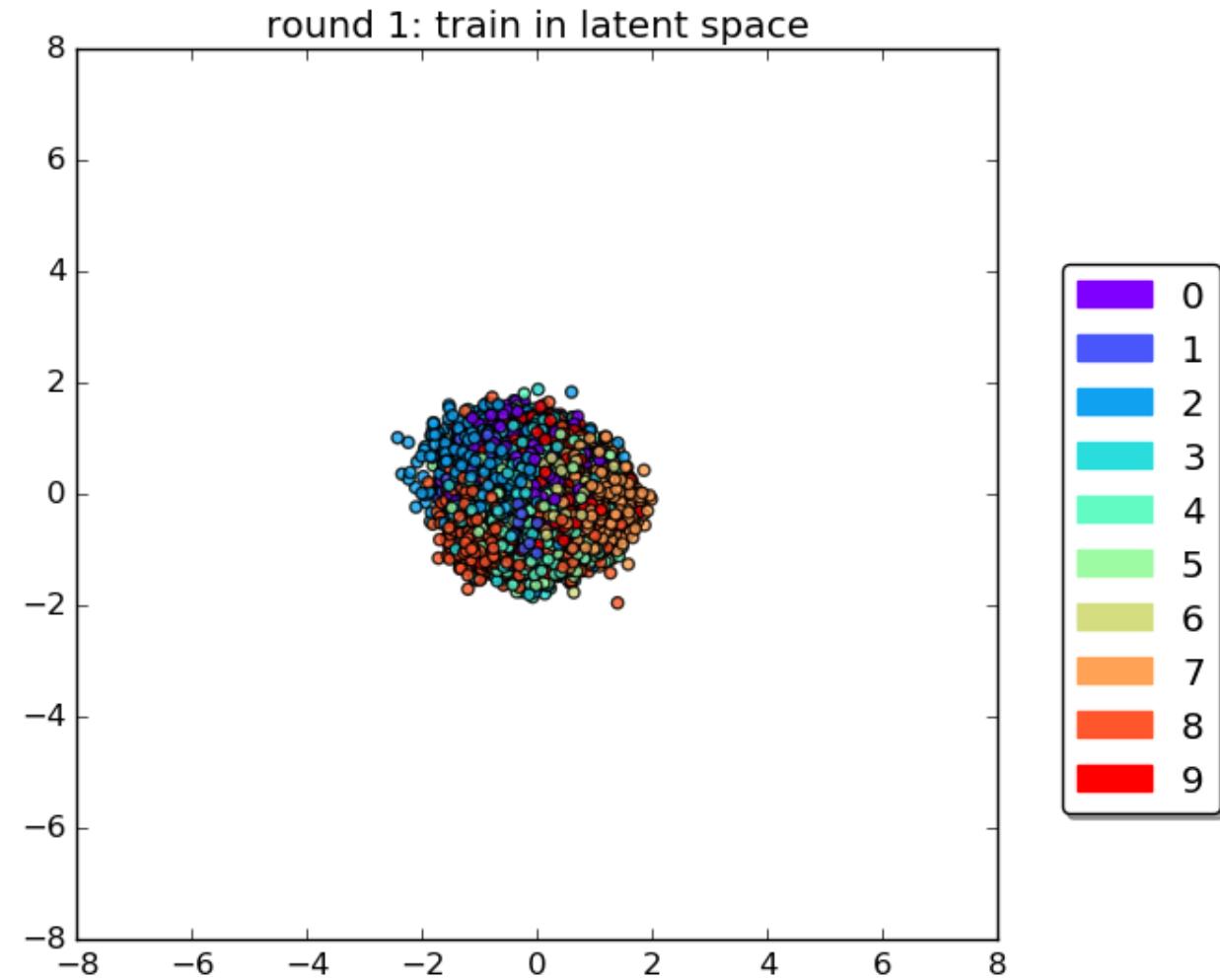
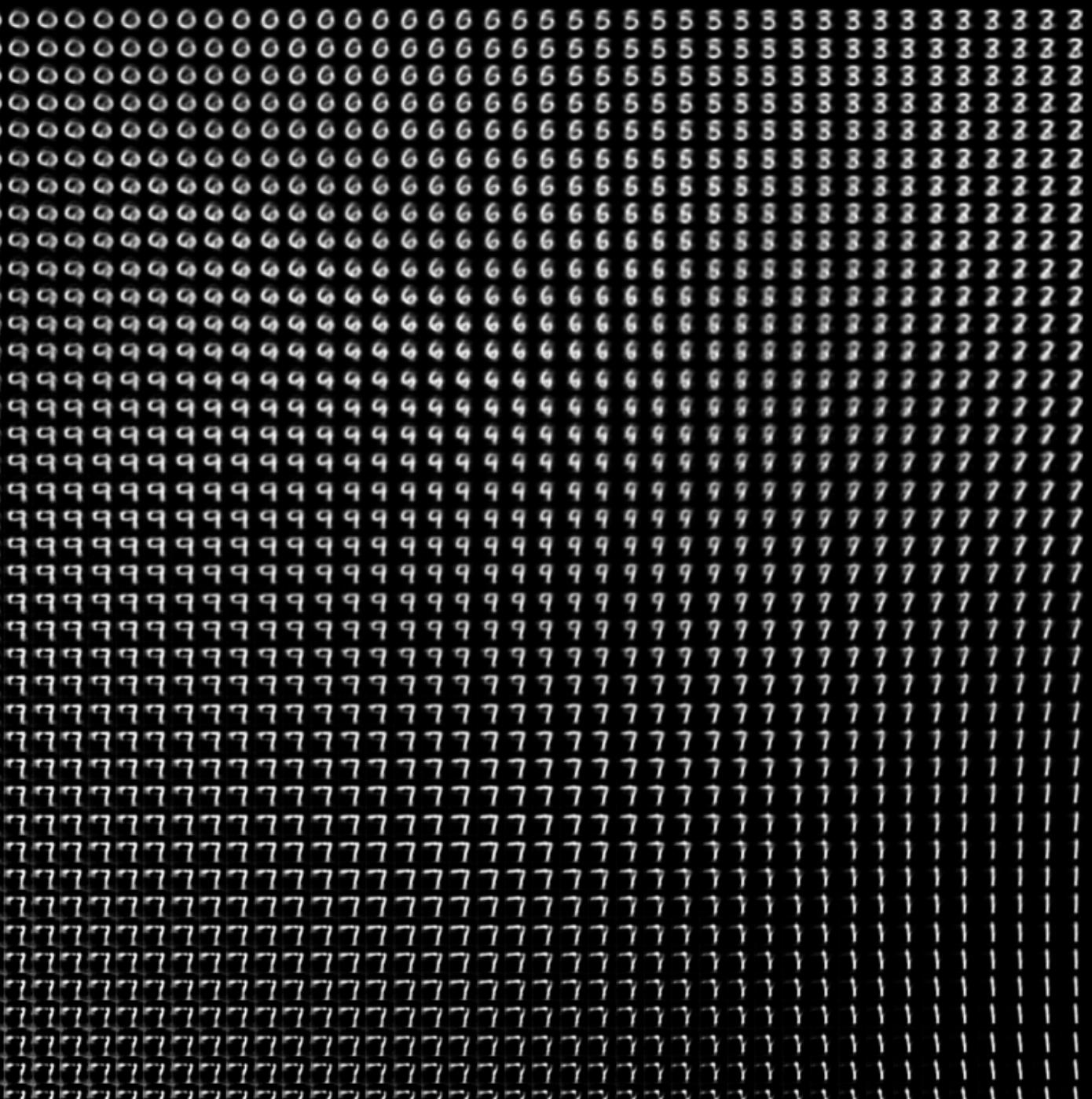


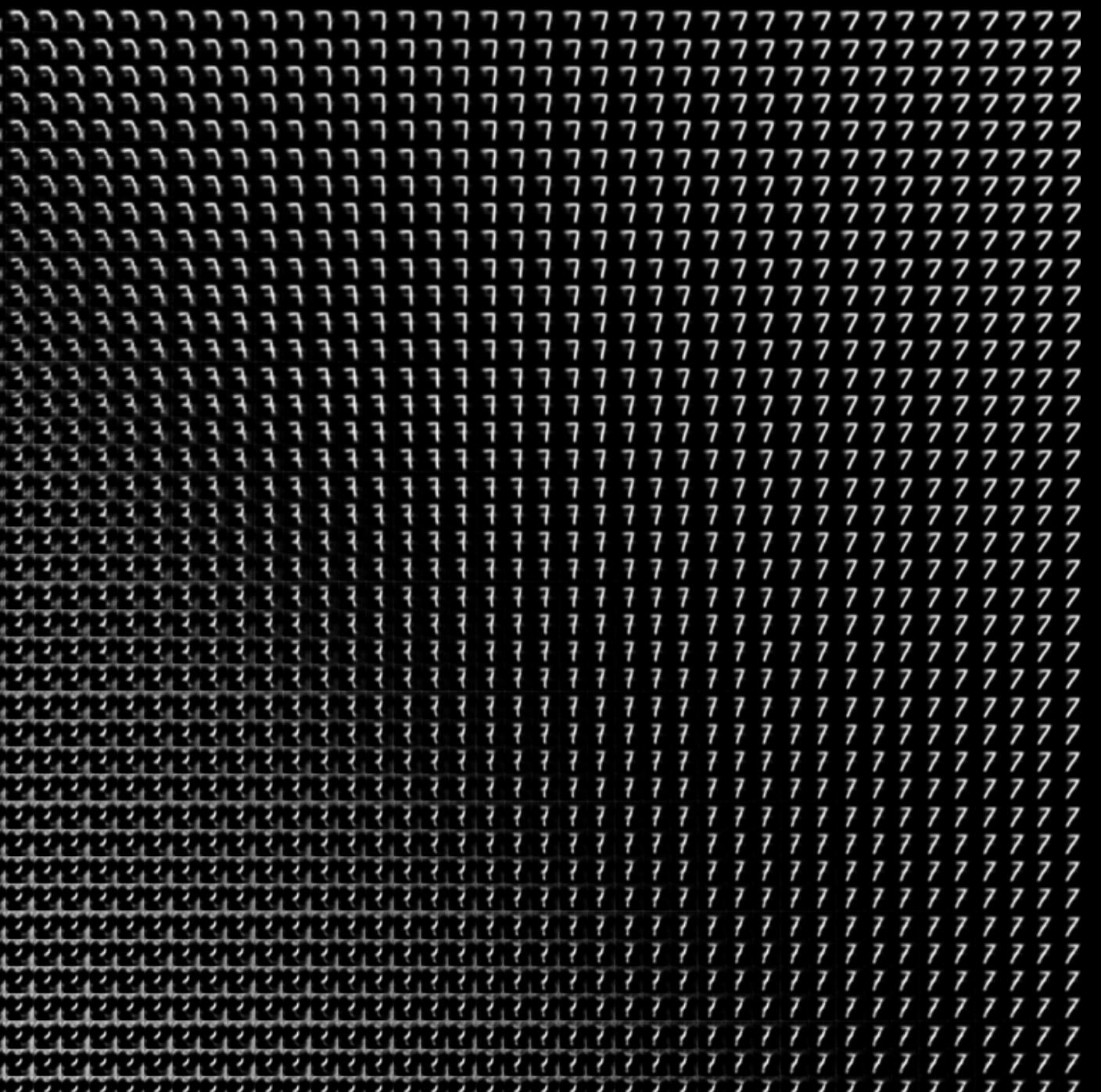
Image space → Latent space



# Autoencoder latent representation



# Autoencoder latent representation (digit 7)



# VAE latent representation

The image displays a large grid of binary digits, specifically zeros and ones, arranged in a repeating pattern. The pattern consists of diagonal bands of ones and zeros, creating a visual effect similar to a barcode or a technical diagram. The ones are represented by the digit '1' and the zeros by the digit '0'. The pattern repeats across the entire grid, with each band being approximately 8 units wide.

# VAE latent representation (digit 7)

The image consists of a large grid of black digits on a white background. The digits are arranged in rows and columns. Each row contains a repeating sequence of the digits 1 through 9. The first few rows are as follows:  
Row 1: 1 2 3 4 5 6 7 8 9  
Row 2: 1 2 3 4 5 6 7 8 9  
Row 3: 1 2 3 4 5 6 7 8 9  
Row 4: 1 2 3 4 5 6 7 8 9  
Row 5: 1 2 3 4 5 6 7 8 9  
Row 6: 1 2 3 4 5 6 7 8 9  
Row 7: 1 2 3 4 5 6 7 8 9  
Row 8: 1 2 3 4 5 6 7 8 9  
Row 9: 1 2 3 4 5 6 7 8 9  
Row 10: 1 2 3 4 5 6 7 8 9  
Row 11: 1 2 3 4 5 6 7 8 9  
Row 12: 1 2 3 4 5 6 7 8 9  
Row 13: 1 2 3 4 5 6 7 8 9  
Row 14: 1 2 3 4 5 6 7 8 9  
Row 15: 1 2 3 4 5 6 7 8 9  
Row 16: 1 2 3 4 5 6 7 8 9  
Row 17: 1 2 3 4 5 6 7 8 9  
Row 18: 1 2 3 4 5 6 7 8 9  
Row 19: 1 2 3 4 5 6 7 8 9  
Row 20: 1 2 3 4 5 6 7 8 9  
Row 21: 1 2 3 4 5 6 7 8 9  
Row 22: 1 2 3 4 5 6 7 8 9  
Row 23: 1 2 3 4 5 6 7 8 9  
Row 24: 1 2 3 4 5 6 7 8 9  
Row 25: 1 2 3 4 5 6 7 8 9  
Row 26: 1 2 3 4 5 6 7 8 9  
Row 27: 1 2 3 4 5 6 7 8 9  
Row 28: 1 2 3 4 5 6 7 8 9  
Row 29: 1 2 3 4 5 6 7 8 9  
Row 30: 1 2 3 4 5 6 7 8 9  
Row 31: 1 2 3 4 5 6 7 8 9  
Row 32: 1 2 3 4 5 6 7 8 9  
Row 33: 1 2 3 4 5 6 7 8 9  
Row 34: 1 2 3 4 5 6 7 8 9  
Row 35: 1 2 3 4 5 6 7 8 9  
Row 36: 1 2 3 4 5 6 7 8 9  
Row 37: 1 2 3 4 5 6 7 8 9  
Row 38: 1 2 3 4 5 6 7 8 9  
Row 39: 1 2 3 4 5 6 7 8 9  
Row 40: 1 2 3 4 5 6 7 8 9  
Row 41: 1 2 3 4 5 6 7 8 9  
Row 42: 1 2 3 4 5 6 7 8 9  
Row 43: 1 2 3 4 5 6 7 8 9  
Row 44: 1 2 3 4 5 6 7 8 9  
Row 45: 1 2 3 4 5 6 7 8 9  
Row 46: 1 2 3 4 5 6 7 8 9  
Row 47: 1 2 3 4 5 6 7 8 9  
Row 48: 1 2 3 4 5 6 7 8 9  
Row 49: 1 2 3 4 5 6 7 8 9  
Row 50: 1 2 3 4 5 6 7 8 9  
Row 51: 1 2 3 4 5 6 7 8 9  
Row 52: 1 2 3 4 5 6 7 8 9  
Row 53: 1 2 3 4 5 6 7 8 9  
Row 54: 1 2 3 4 5 6 7 8 9  
Row 55: 1 2 3 4 5 6 7 8 9  
Row 56: 1 2 3 4 5 6 7 8 9  
Row 57: 1 2 3 4 5 6 7 8 9  
Row 58: 1 2 3 4 5 6 7 8 9  
Row 59: 1 2 3 4 5 6 7 8 9  
Row 60: 1 2 3 4 5 6 7 8 9  
Row 61: 1 2 3 4 5 6 7 8 9  
Row 62: 1 2 3 4 5 6 7 8 9  
Row 63: 1 2 3 4 5 6 7 8 9  
Row 64: 1 2 3 4 5 6 7 8 9  
Row 65: 1 2 3 4 5 6 7 8 9  
Row 66: 1 2 3 4 5 6 7 8 9  
Row 67: 1 2 3 4 5 6 7 8 9  
Row 68: 1 2 3 4 5 6 7 8 9  
Row 69: 1 2 3 4 5 6 7 8 9  
Row 70: 1 2 3 4 5 6 7 8 9  
Row 71: 1 2 3 4 5 6 7 8 9  
Row 72: 1 2 3 4 5 6 7 8 9  
Row 73: 1 2 3 4 5 6 7 8 9  
Row 74: 1 2 3 4 5 6 7 8 9  
Row 75: 1 2 3 4 5 6 7 8 9  
Row 76: 1 2 3 4 5 6 7 8 9  
Row 77: 1 2 3 4 5 6 7 8 9  
Row 78: 1 2 3 4 5 6 7 8 9  
Row 79: 1 2 3 4 5 6 7 8 9  
Row 80: 1 2 3 4 5 6 7 8 9  
Row 81: 1 2 3 4 5 6 7 8 9  
Row 82: 1 2 3 4 5 6 7 8 9  
Row 83: 1 2 3 4 5 6 7 8 9  
Row 84: 1 2 3 4 5 6 7 8 9  
Row 85: 1 2 3 4 5 6 7 8 9  
Row 86: 1 2 3 4 5 6 7 8 9  
Row 87: 1 2 3 4 5 6 7 8 9  
Row 88: 1 2 3 4 5 6 7 8 9  
Row 89: 1 2 3 4 5 6 7 8 9  
Row 90: 1 2 3 4 5 6 7 8 9  
Row 91: 1 2 3 4 5 6 7 8 9  
Row 92: 1 2 3 4 5 6 7 8 9  
Row 93: 1 2 3 4 5 6 7 8 9  
Row 94: 1 2 3 4 5 6 7 8 9  
Row 95: 1 2 3 4 5 6 7 8 9  
Row 96: 1 2 3 4 5 6 7 8 9  
Row 97: 1 2 3 4 5 6 7 8 9  
Row 98: 1 2 3 4 5 6 7 8 9  
Row 99: 1 2 3 4 5 6 7 8 9  
Row 100: 1 2 3 4 5 6 7 8 9

# VAE latent representation (digit 8)

The image consists of a large grid of small, faint, light gray numbers. The numbers are arranged in a 10x10 pattern, with each number being a single digit. The digits are very small and light, making them difficult to read individually. The grid is set against a white background.

# Word Embeddings

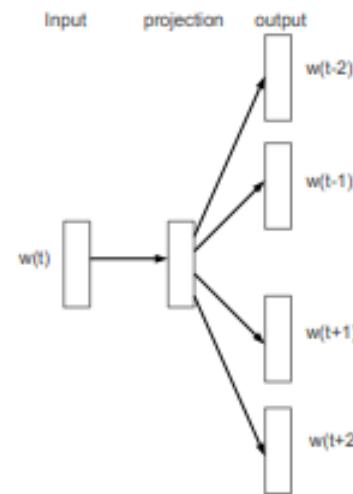
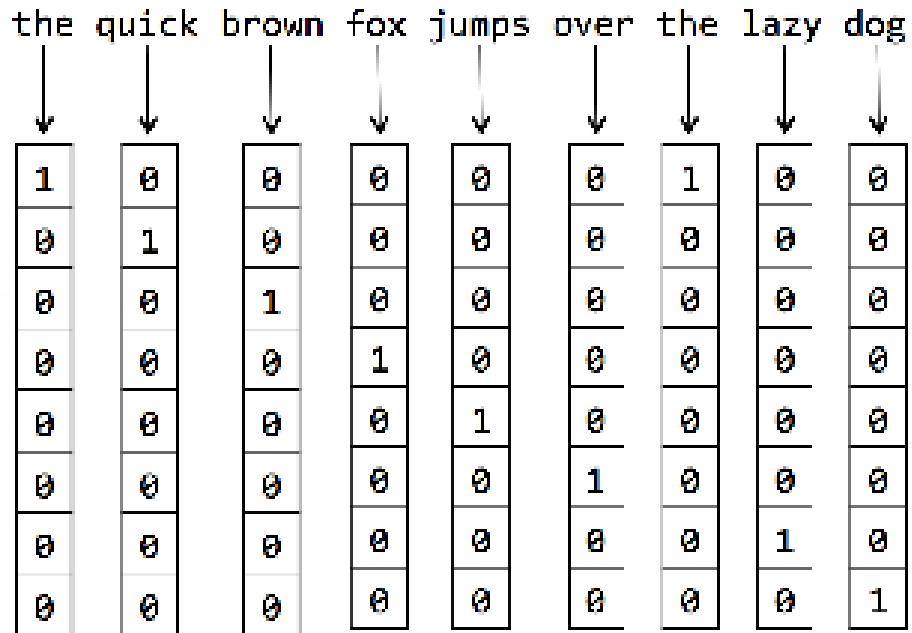
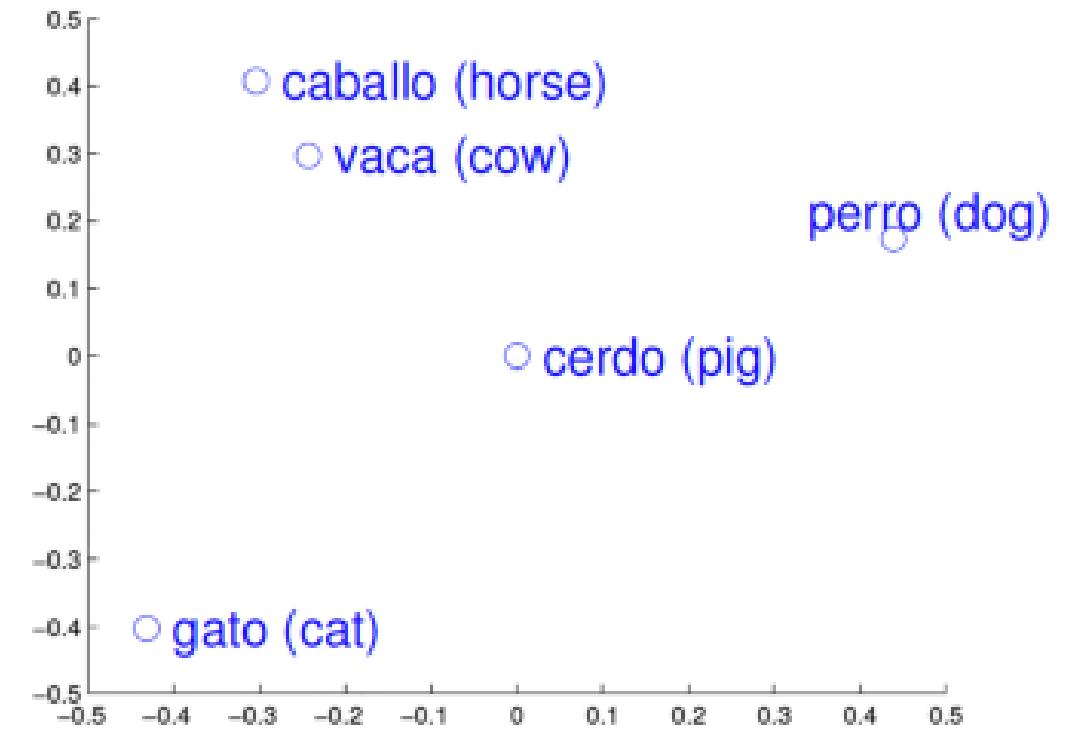
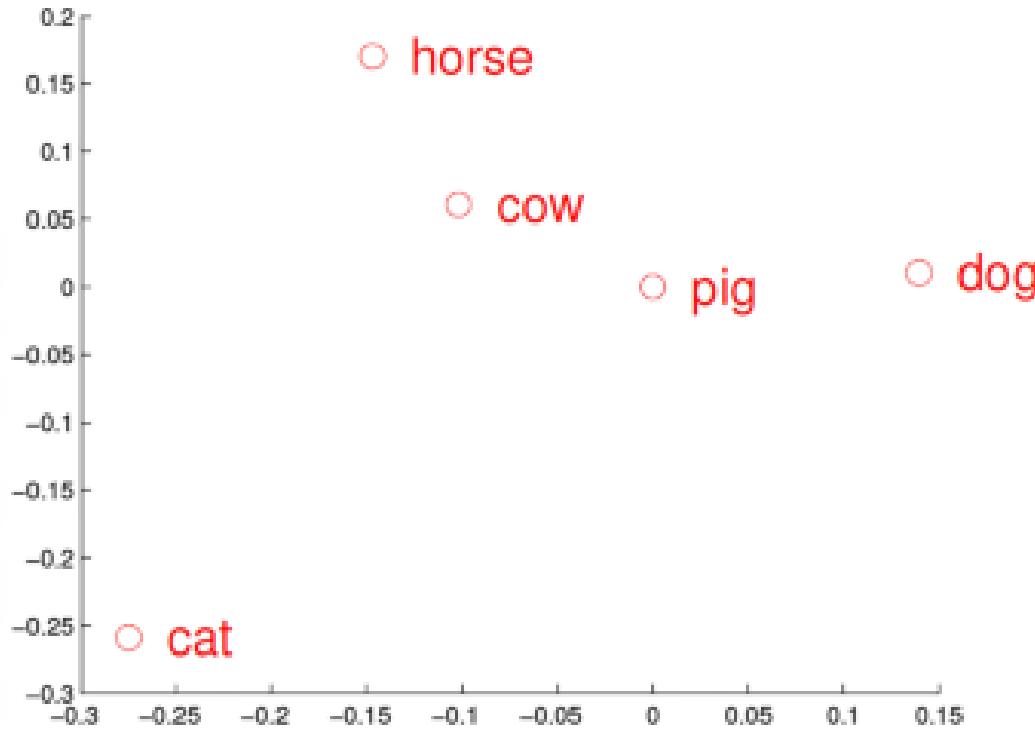


Figure 1: The Skip-gram model architecture. The training objective is to learn word vector representations that are good at predicting the nearby words.

<https://medium.com/geekculture/word-embeddings-in-ai-10a9e430cb59>

[https://proceedings.neurips.cc/paper\\_files/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf)

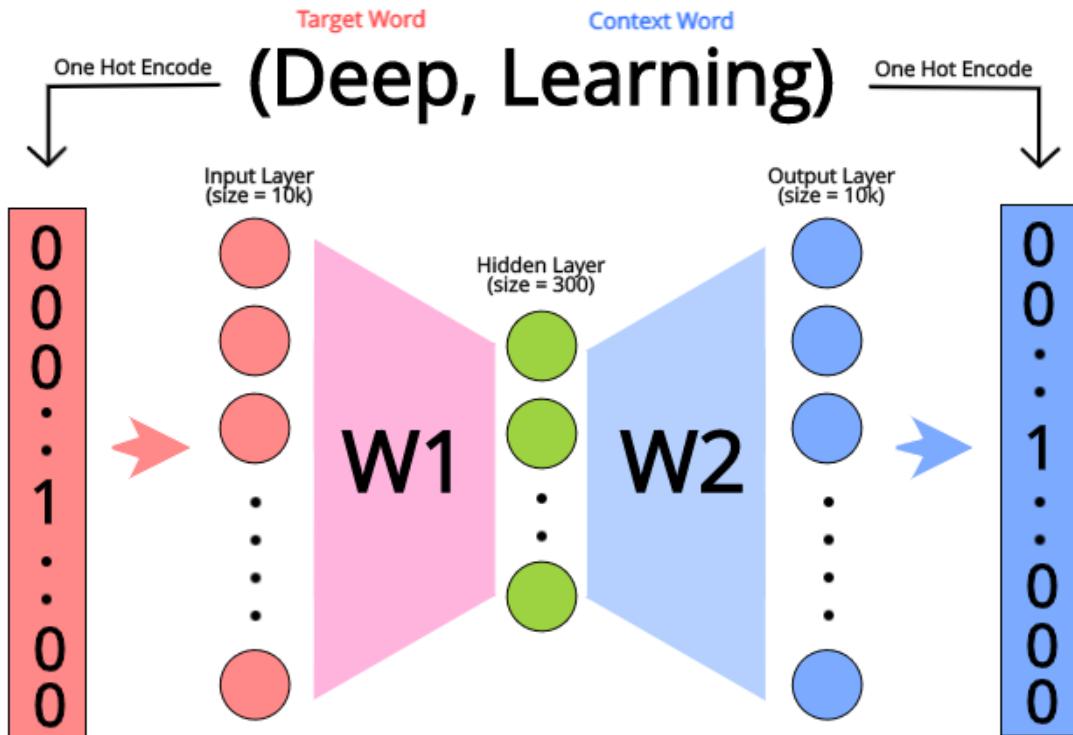
# Word Embeddings



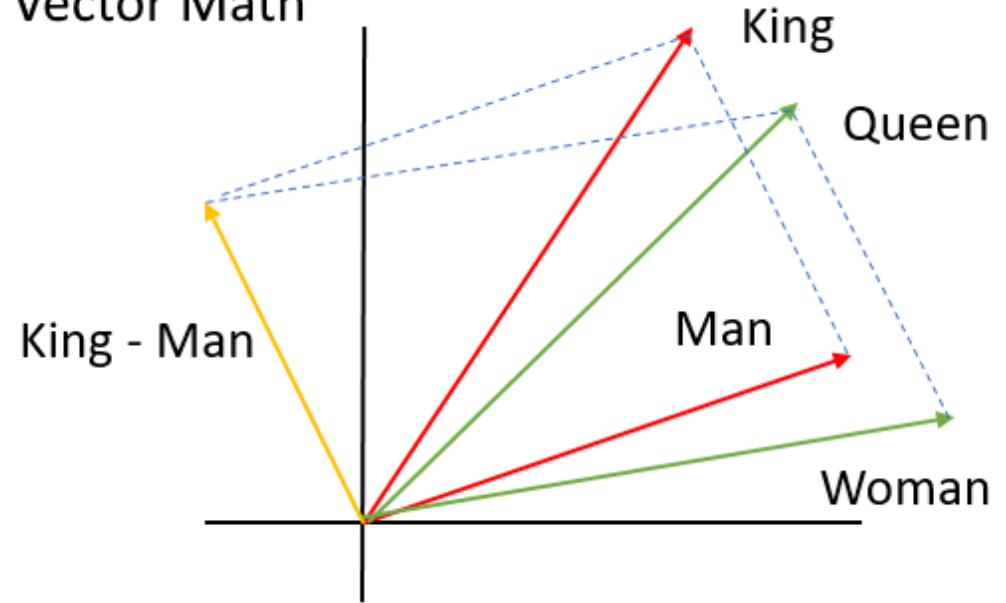
<https://medium.com/geekculture/word-embeddings-in-ai-10a9e430cb59>

# Word Vectors

Skip Gram Architecture

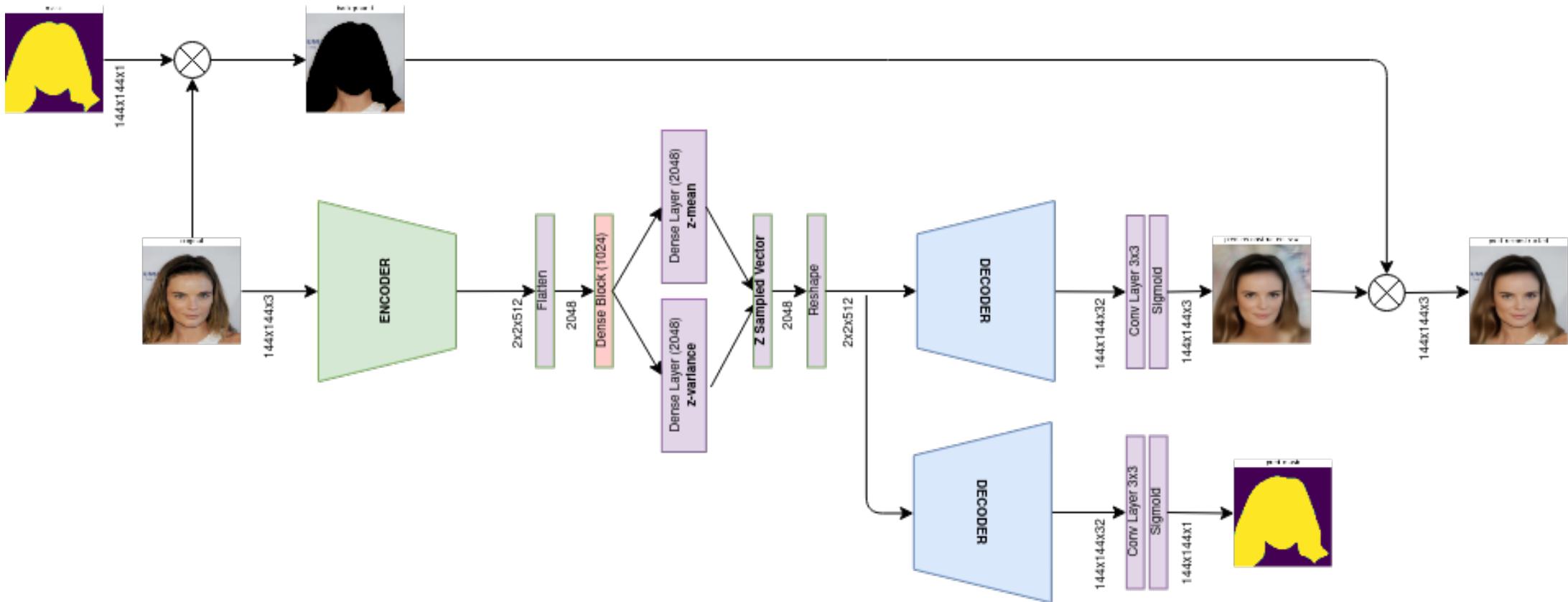


Vector Math



<https://medium.com/analytics-vidhya/word-embeddings-in-nlp-word2vec-glove-fasttext-24d4d4286a73>

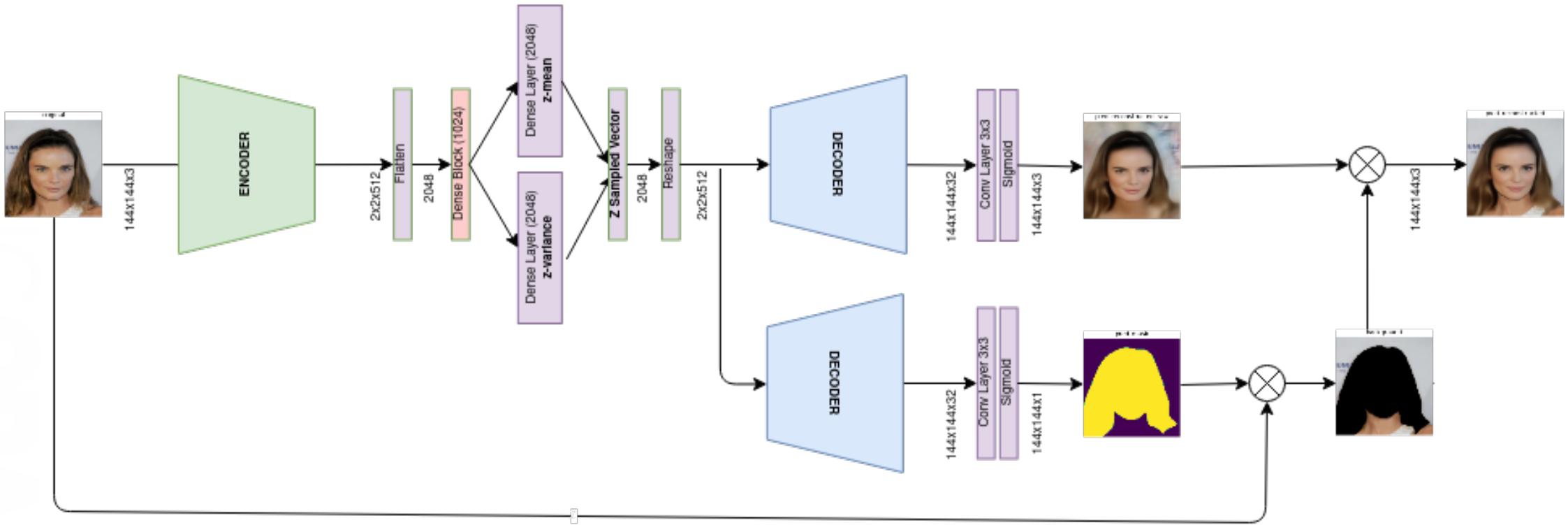
# Making Deep Fakes with VAE



During training, the labels of the face masks are used too, which replaces the background of the reconstructed image such that the loss function is applied only over the face pixels.

<https://rtoledo.me/post/2021-05-31-edit-face-attributes-using-vae/edit-face-attributes-using-vae/>

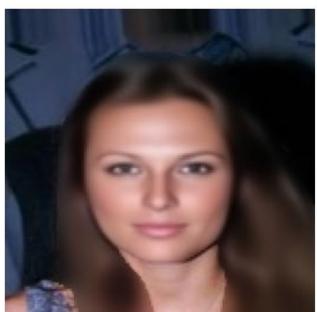
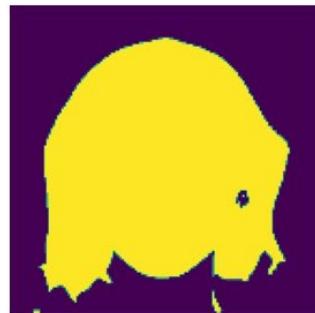
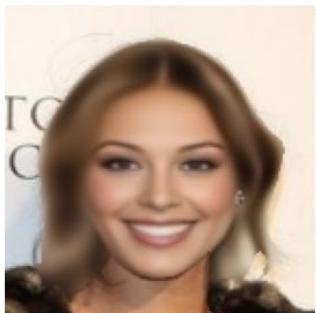
# Making Deep Fakes with VAE



In the prediction mode, the background replacement is done by the predicted mask itself, not requiring any extra input but a sample image.

<https://rtoledo.me/post/2021-05-31-edit-face-attributes-using-vae/edit-face-attributes-using-vae/>

# Reconstruction



<https://rtoledo.me/post/2021-05-31-edit-face-attributes-using-vae/edit-face-attributes-using-vae/>

# Changing Attributes



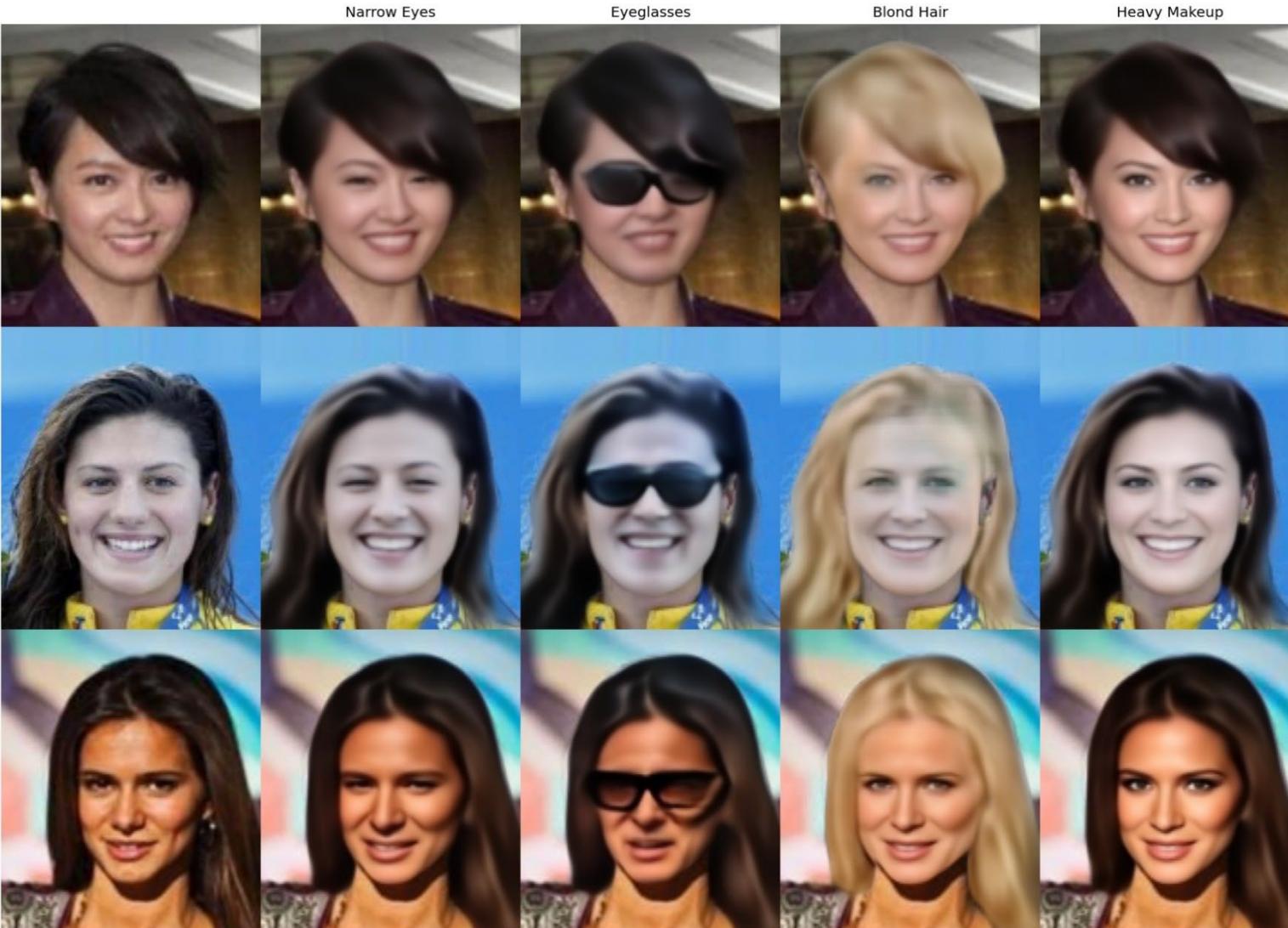
<https://rtoledo.me/post/2021-05-31-edit-face-attributes-using-vae/edit-face-attributes-using-vae/>

# Changing Attributes



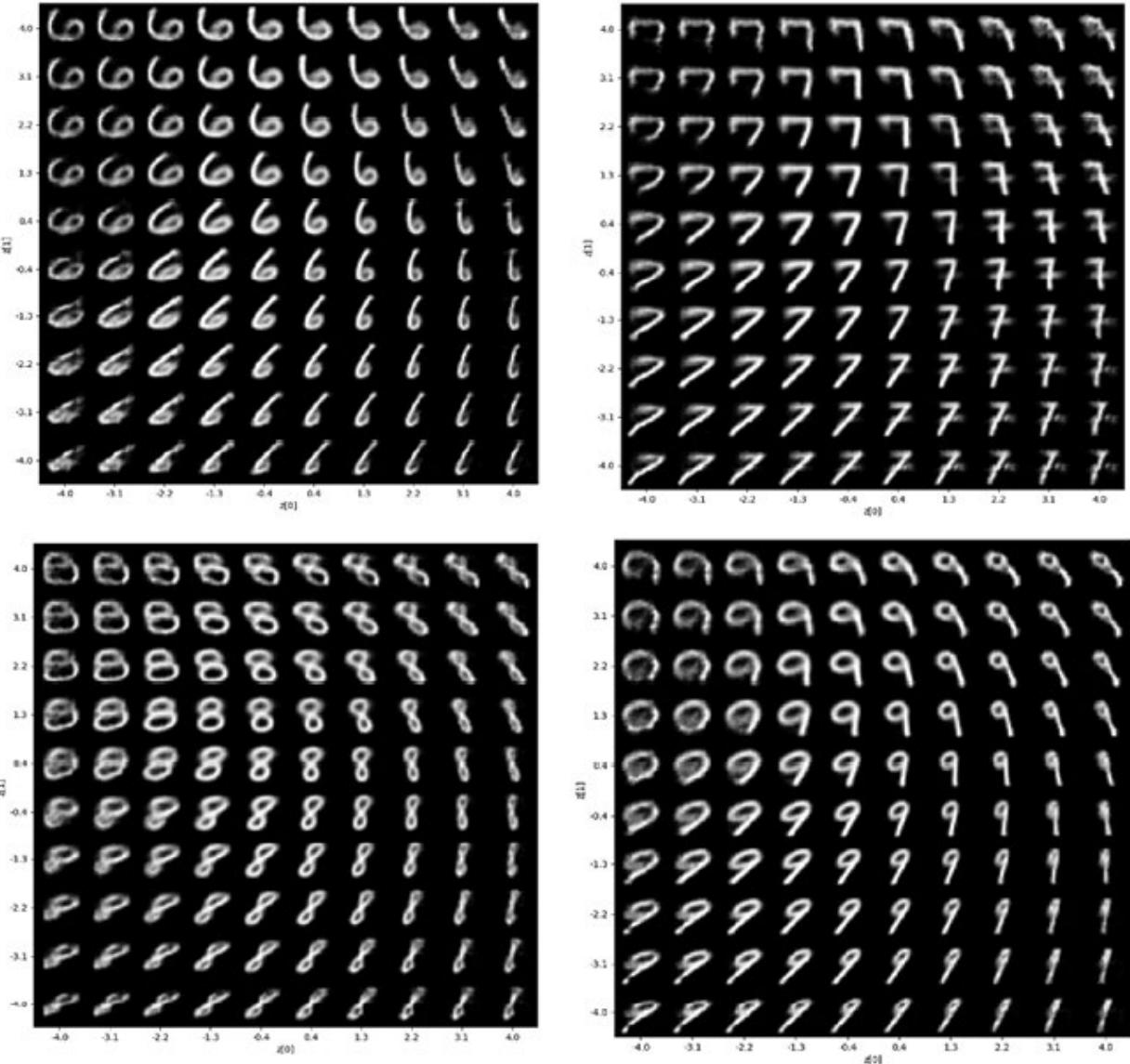
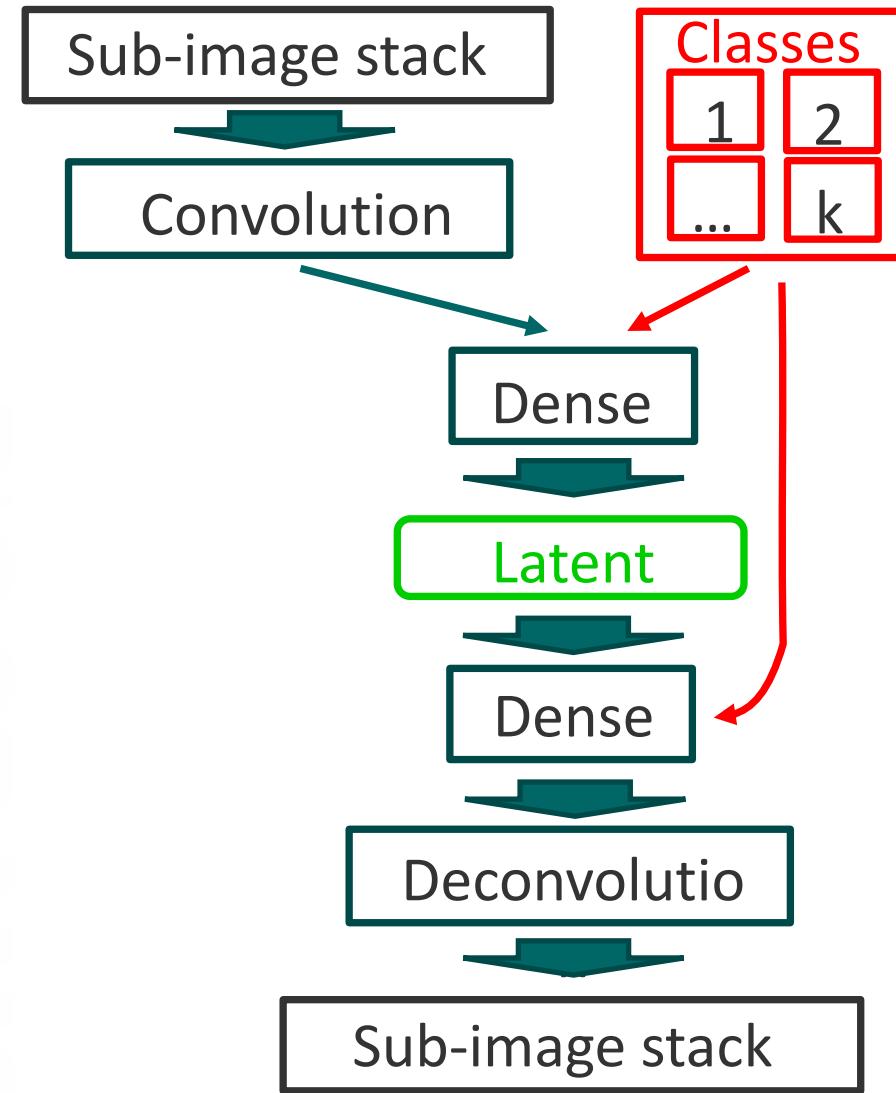
<https://rtoledo.me/post/2021-05-31-edit-face-attributes-using-vae/edit-face-attributes-using-vae/>

# Changing Attributes



<https://rtoledo.me/post/2021-05-31-edit-face-attributes-using-vae/edit-face-attributes-using-vae/>

# Conditional VAE

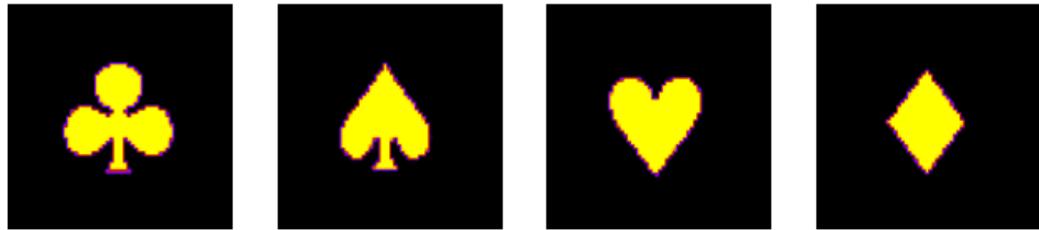


Note the trends in the latent representation for each digit: **disentanglement of the representations**

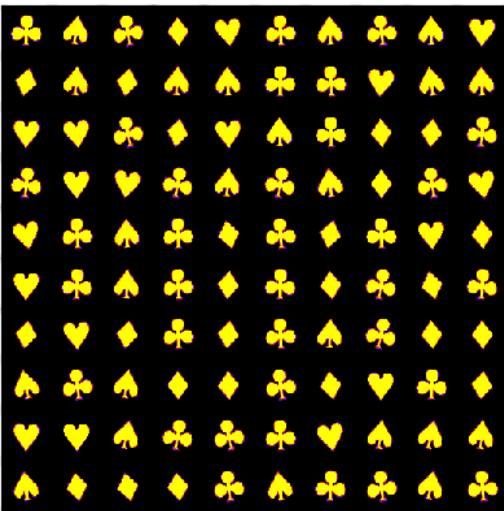
# VAE on Cards

Introduce the **cards** data set:

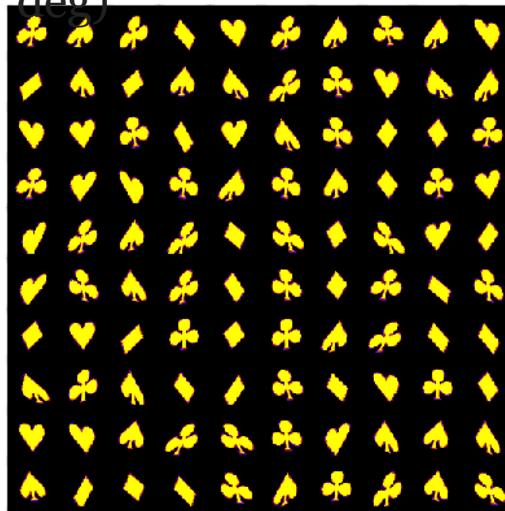
- Classical 4 hands (diamonds, clubs, pikes, hearts)
- Interesting similarities (pires and hearts)
- And invariances on affine transforms (e.g. diamonds)



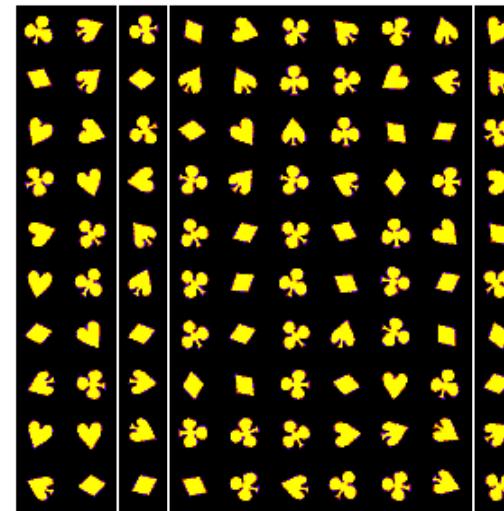
**Cards 1:** Low R (12 deg) and low S (1 deg)



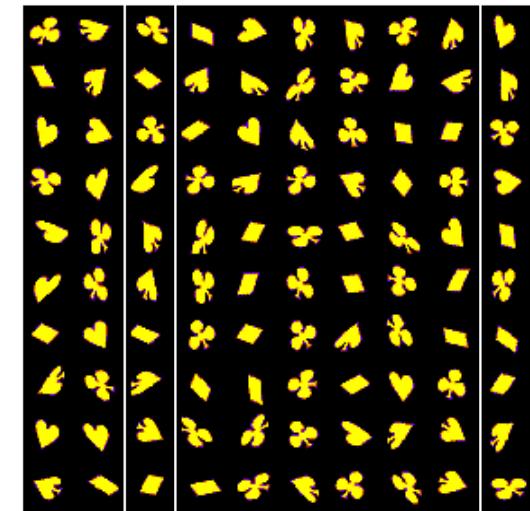
**Cards 2:** Low R (12 deg) and high S (20 deg)



**Cards 3:** High R (120 deg) and Low S (1 deg)



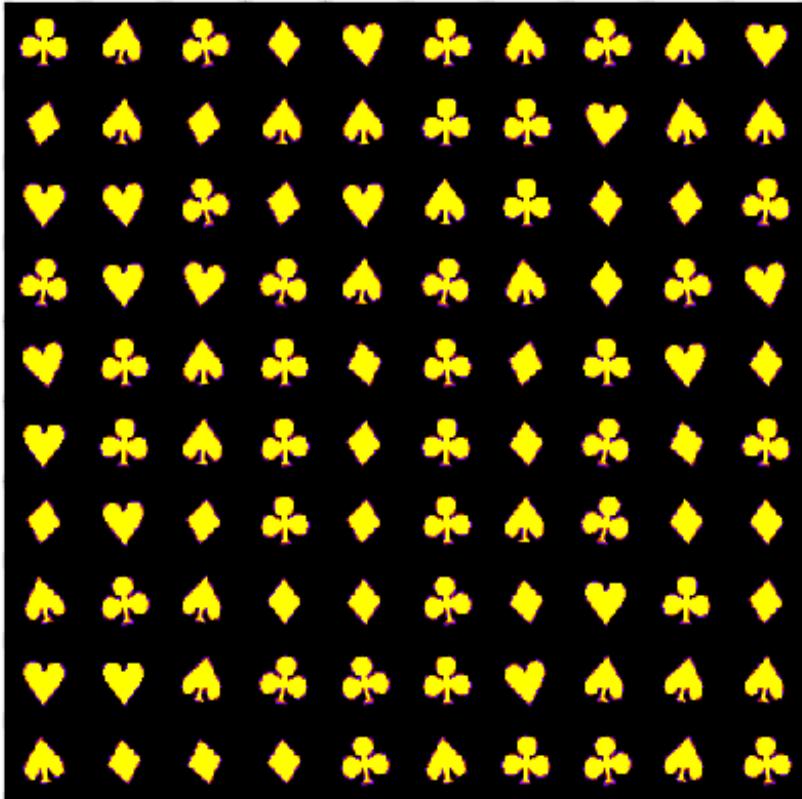
**Cards 4:** High R (120 deg) and high S (20 deg)



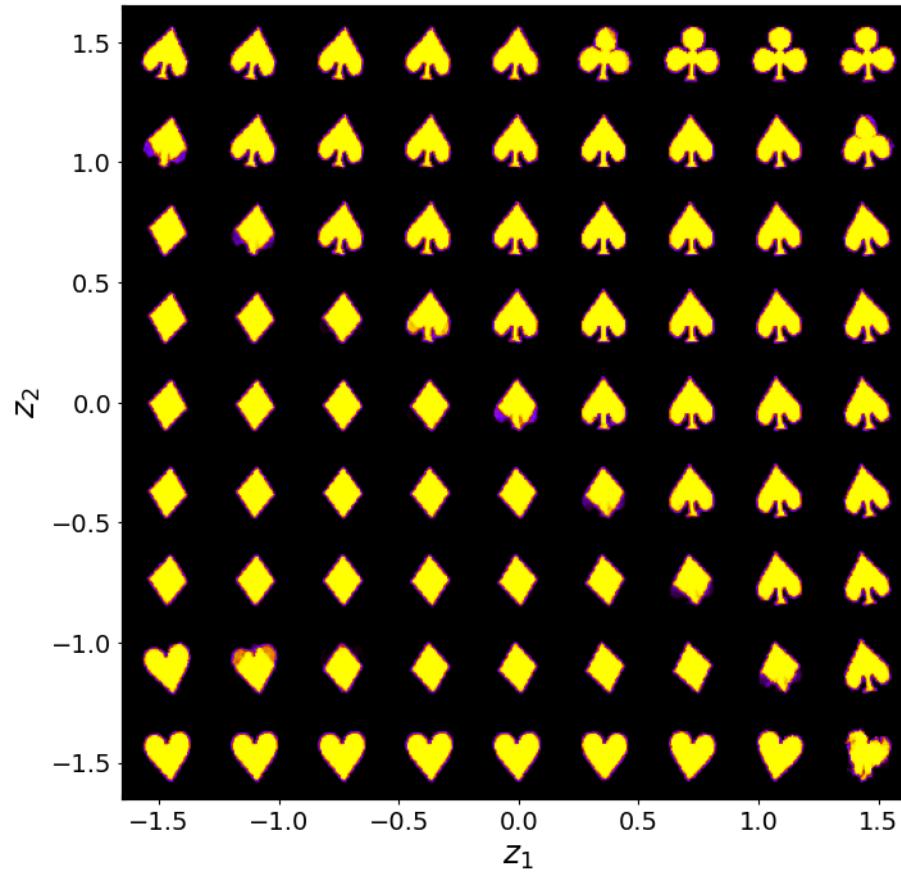
- Shear, rotations, and translations are **known** factors of variability (or traits) in data
- Can VAE disentangle representations and **discover** these factors of variability

# VAE on Cards

Example of data

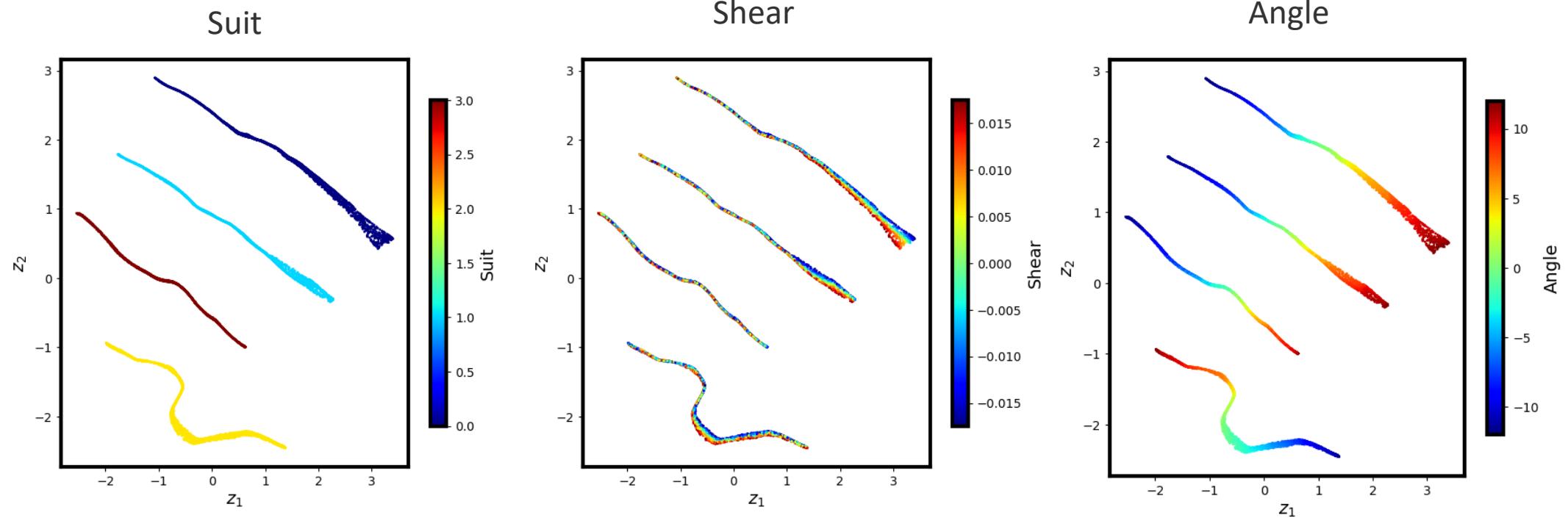


Latent representation



Cards 1: Low rotation (12 deg) and low shear (1 deg)

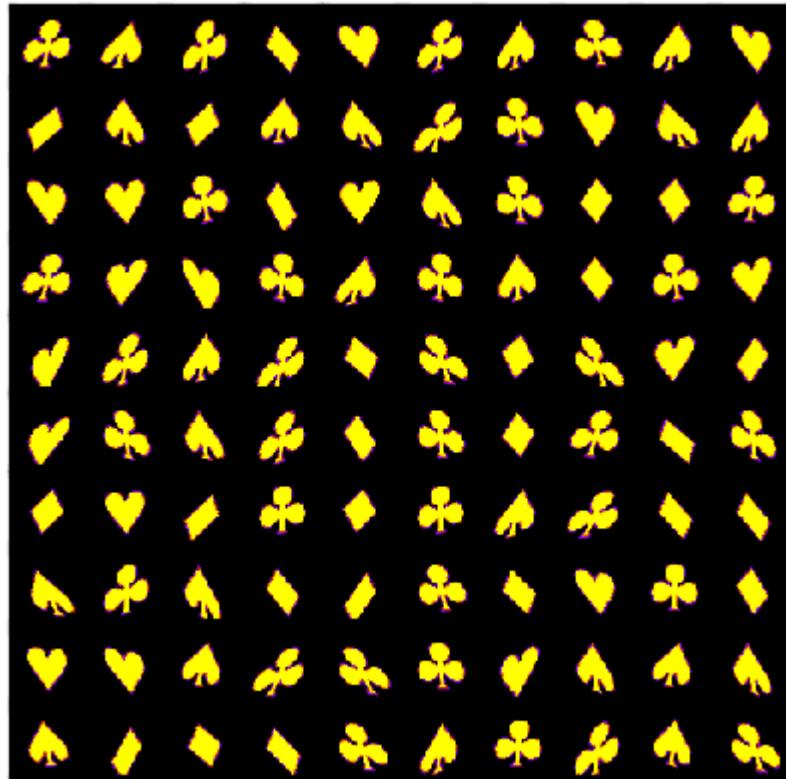
# VAE on Cards



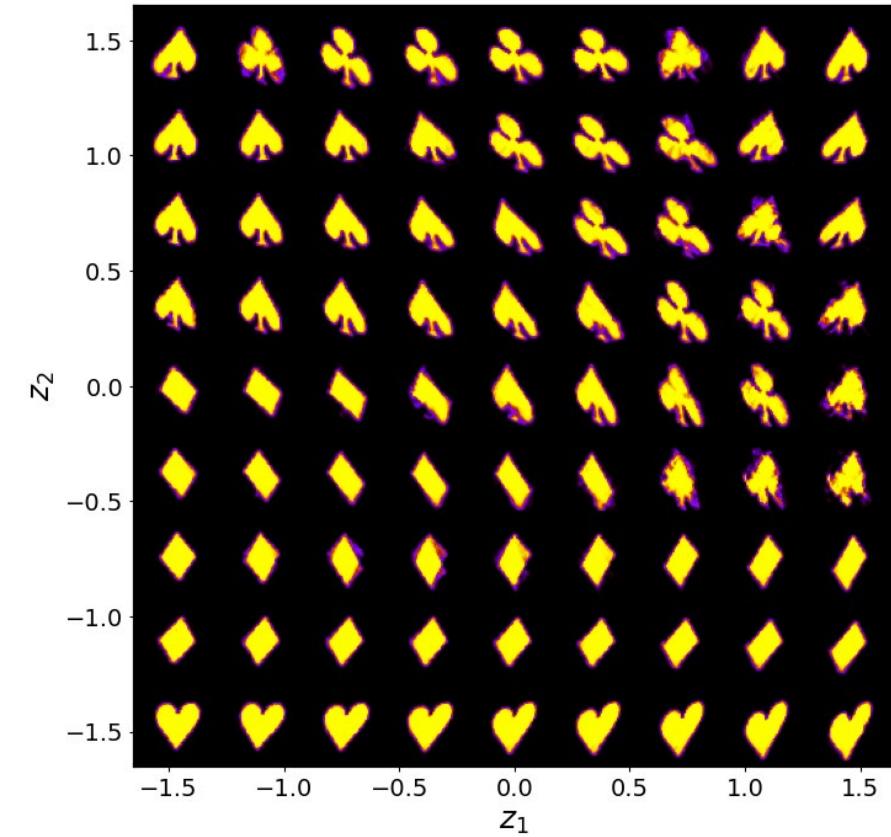
**Cards 1:** Low rotation (12 deg) and low shear (1 deg)

# VAE on Cards

Example of data

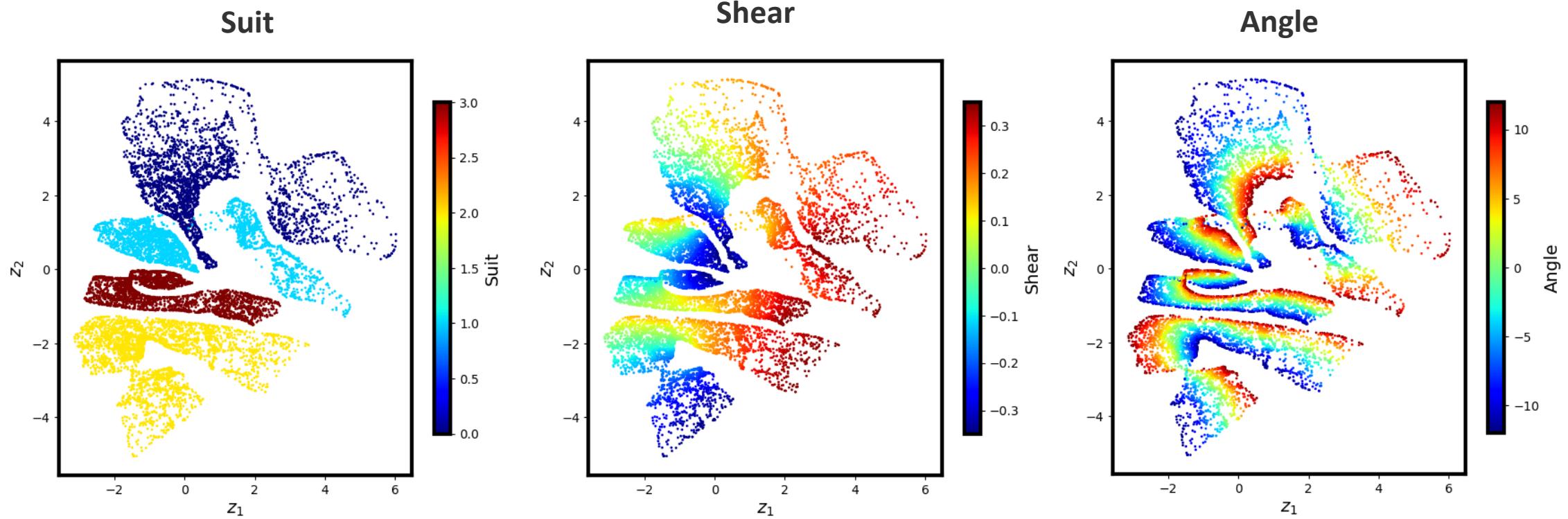


Latent representation



Cards 2: Low rotation (12 deg) and high shear (20 deg)

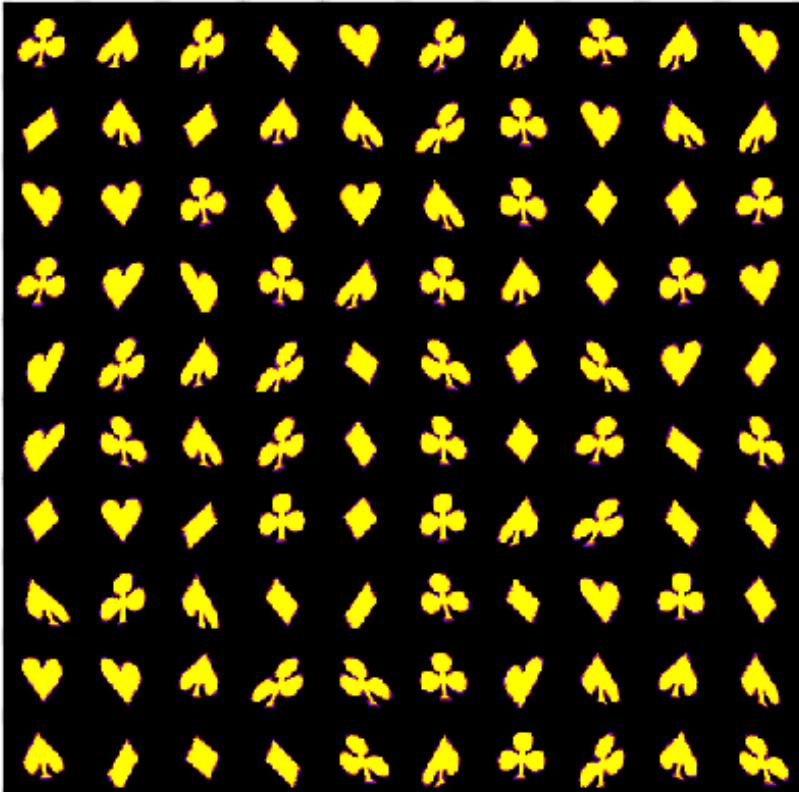
# VAE on Cards



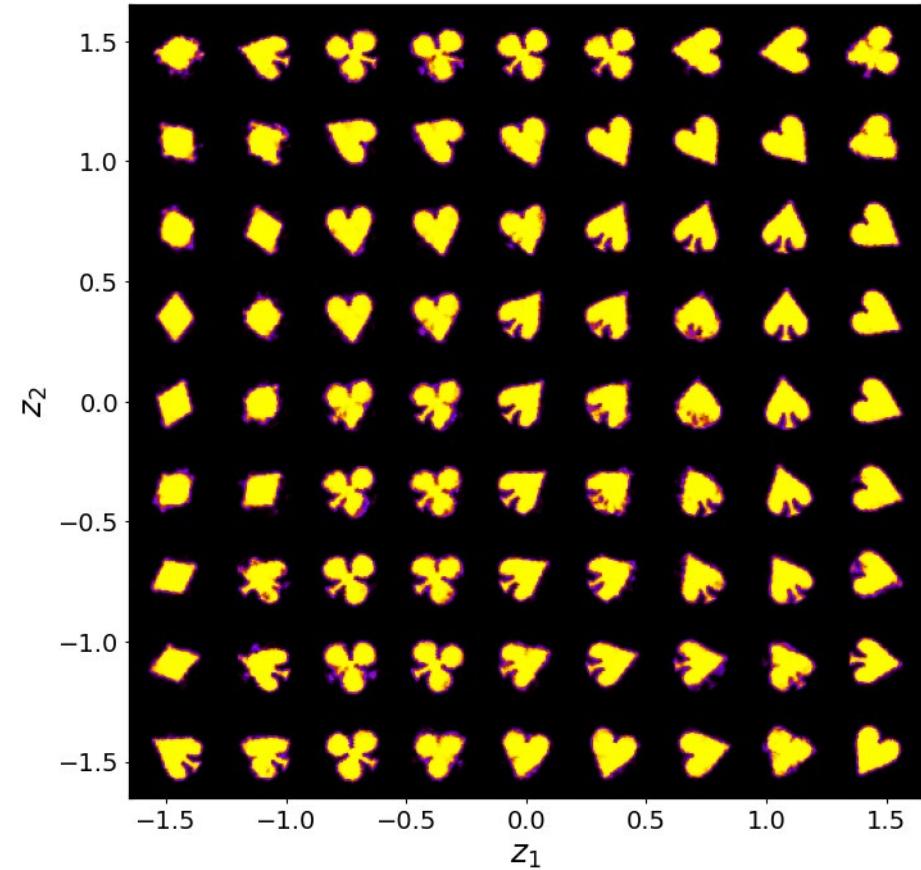
**Cards 2:** Low rotation (12 deg) and high shear (20 deg)

# VAE on Cards

Example of data

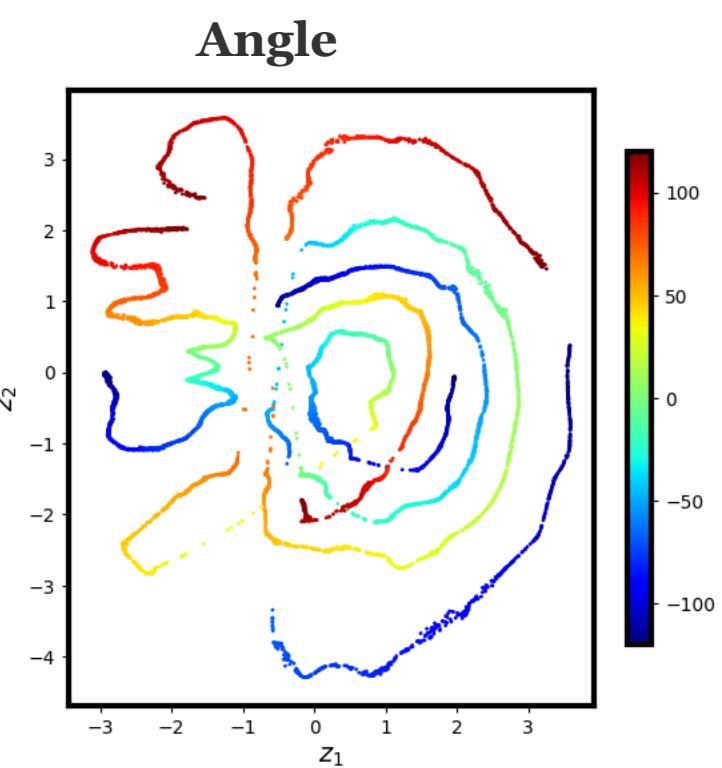
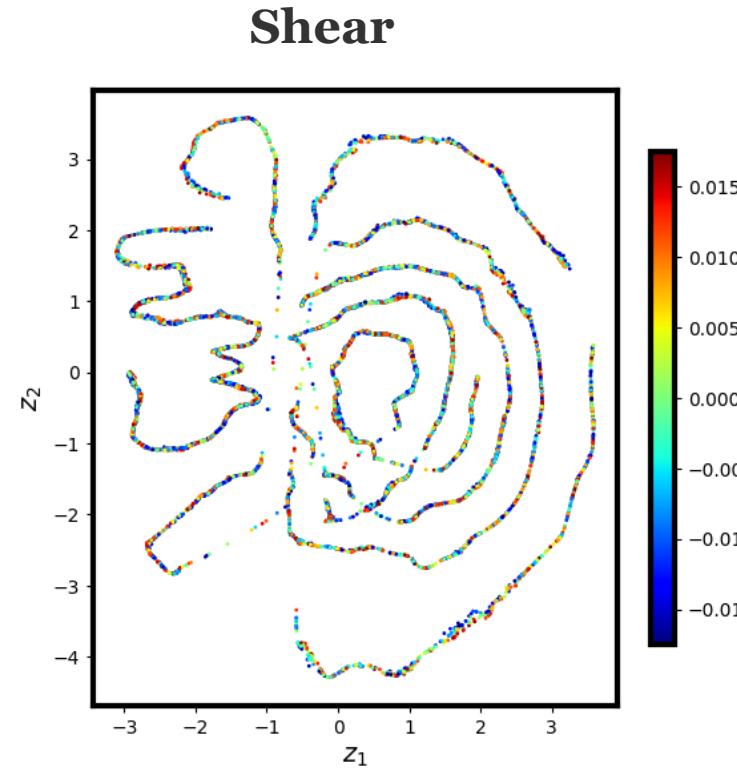
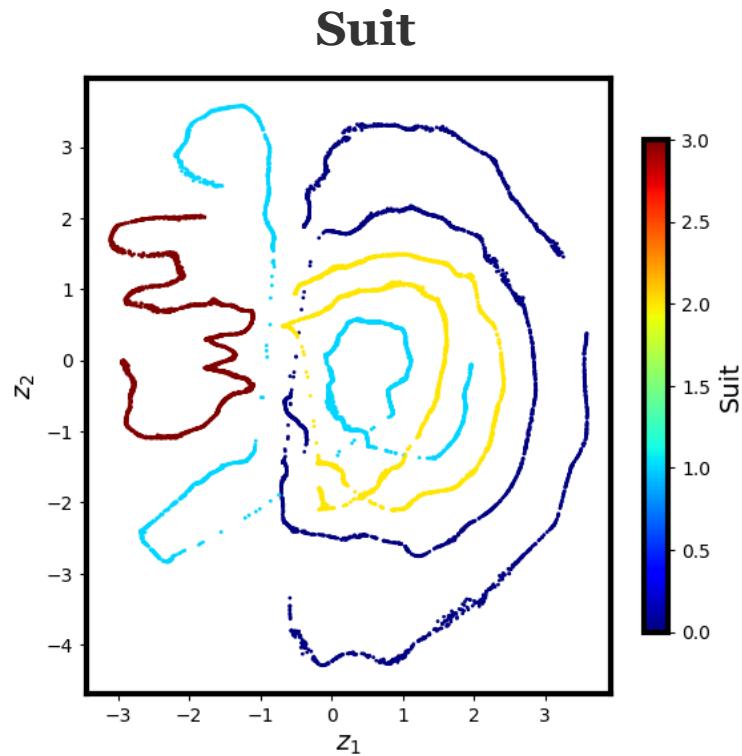


Latent representation



Cards 3: High rotation (120 deg) and low shear (1 deg)

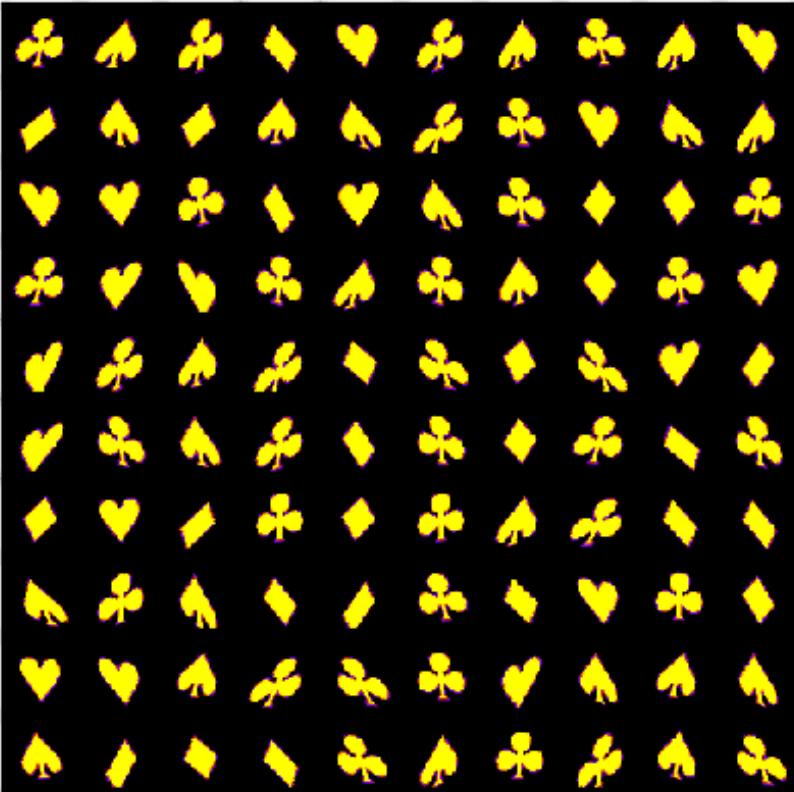
# VAE on Cards



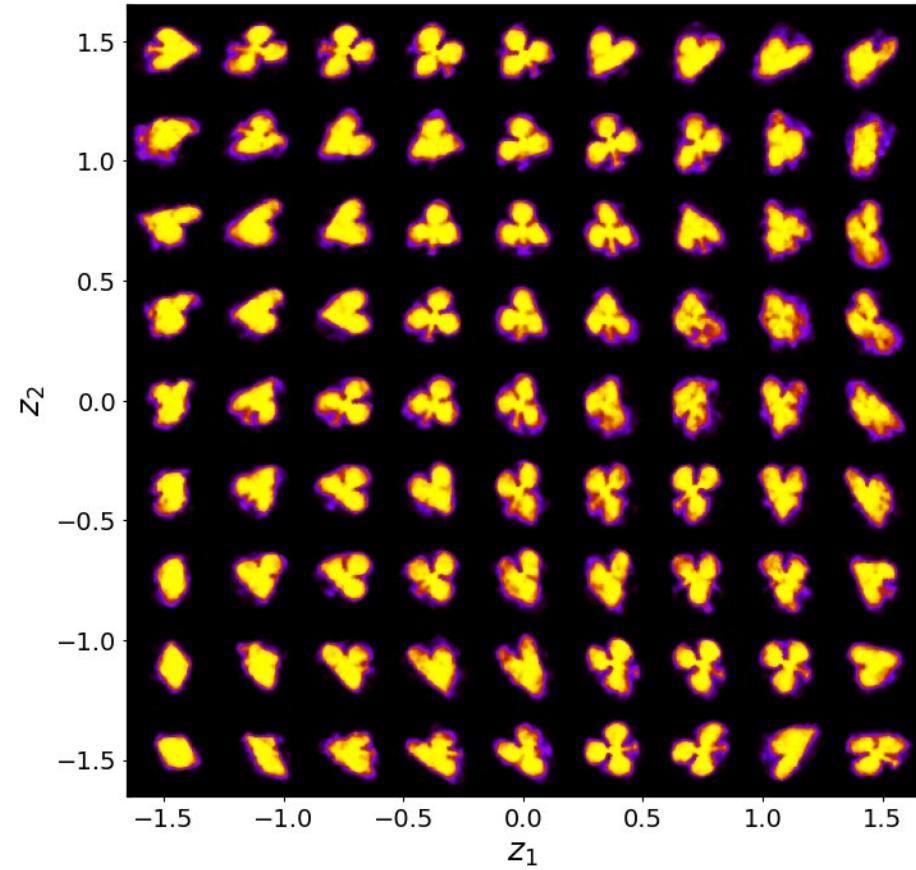
**Cards 3:** High rotation (120 deg) and low shear (1 deg)

# VAE on Cards

Example of data

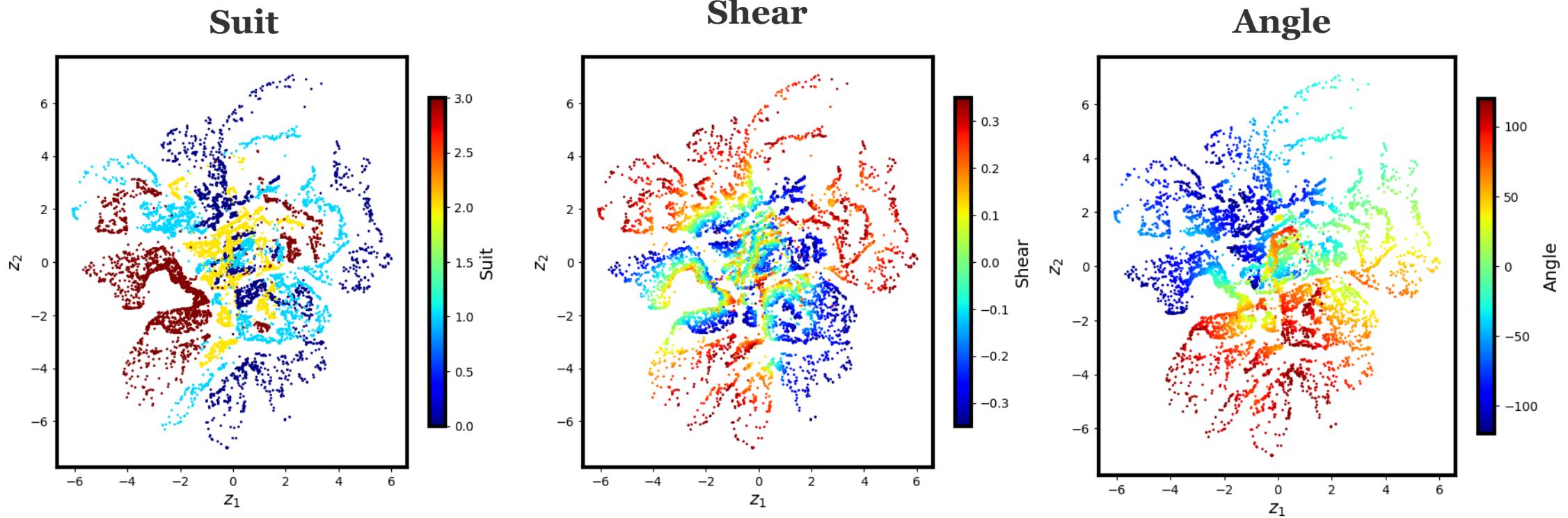


Latent representation



Cards 4: High rotation (120 deg) and high shear (20 deg)

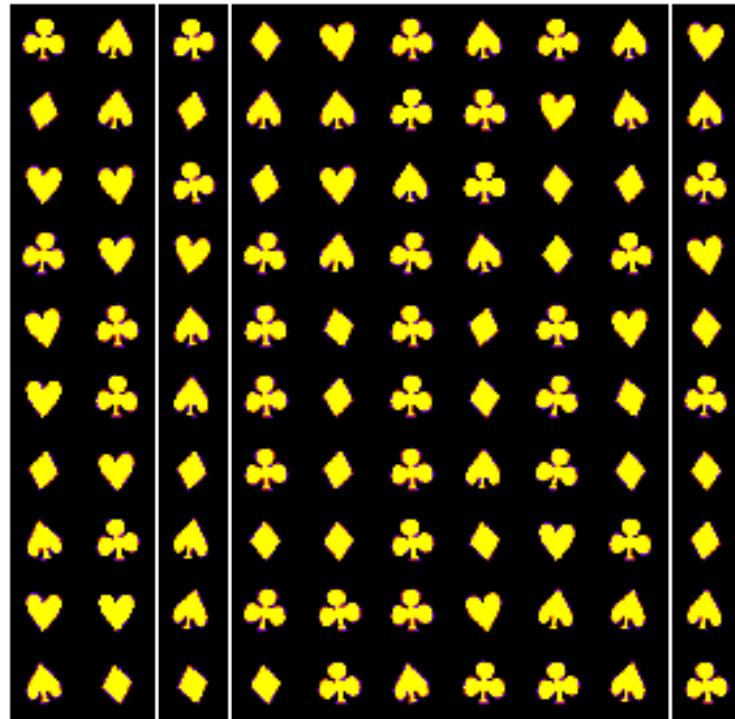
# VAE on Cards



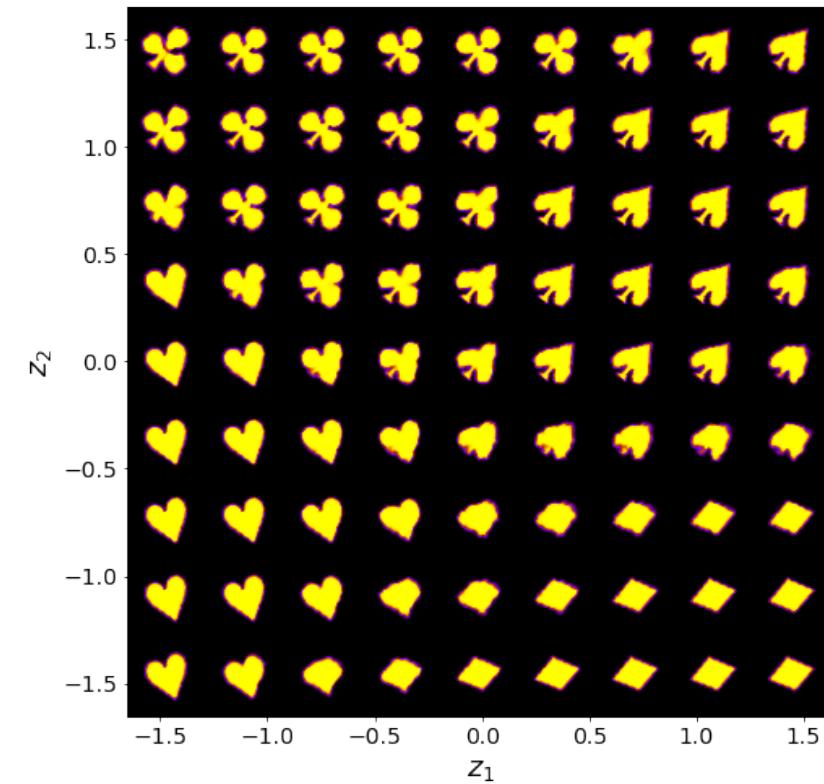
**Cards 4:** High rotation (120 deg) and high shear (20 deg)

# rVAE on Cards

Example of data

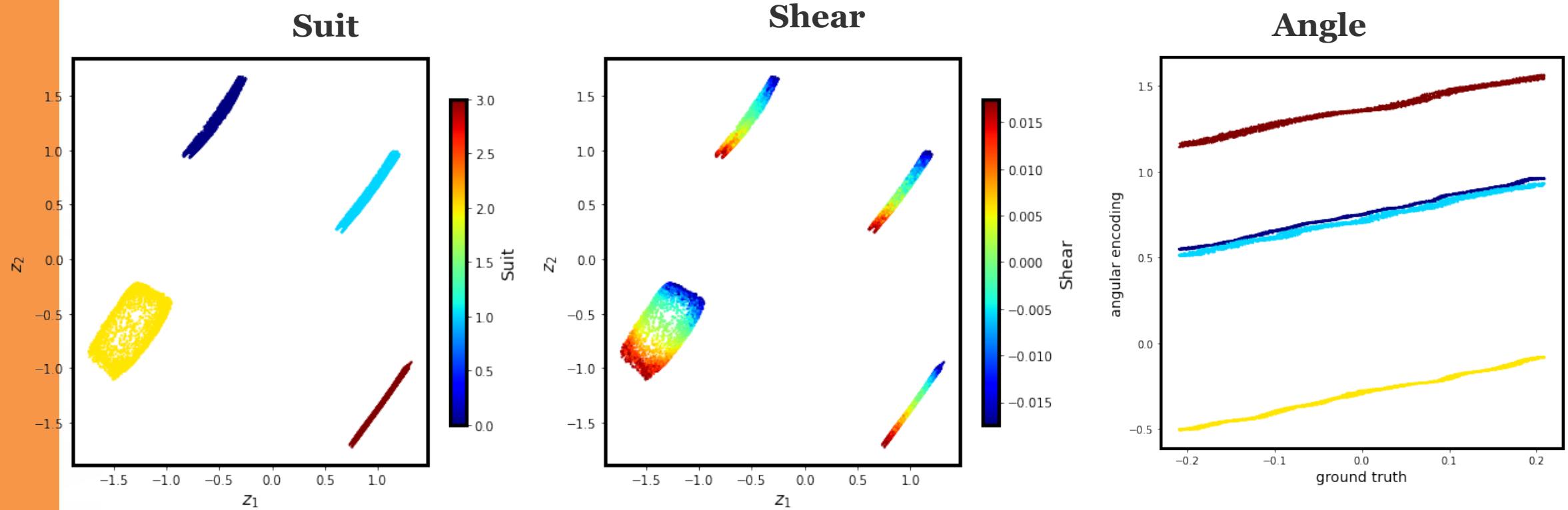


Latent representation



Cards 1: Low rotation (12 deg) and low shear (1 deg)

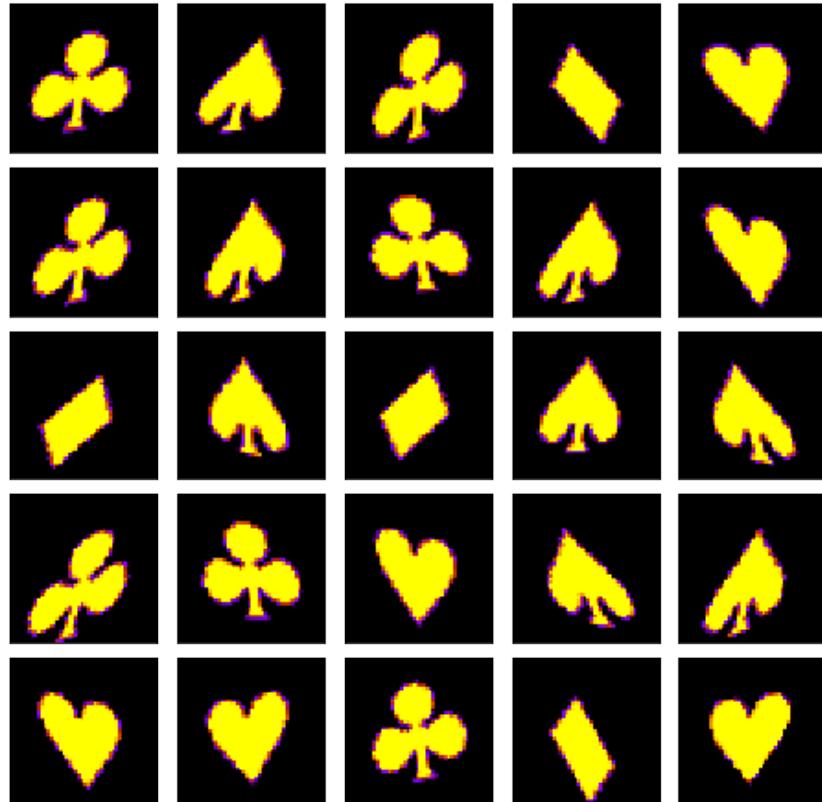
# rVAE on Cards



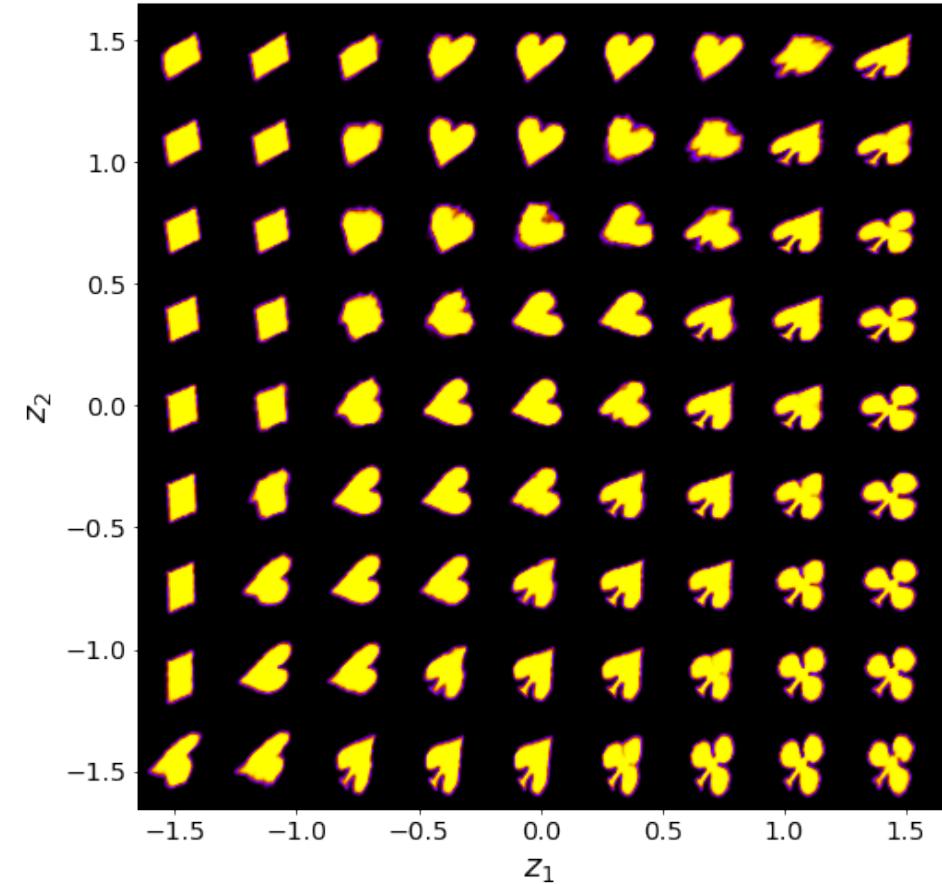
**Cards 1:** Low rotation (12 deg) and low shear (1 deg)

# rVAE on Cards

Example of data

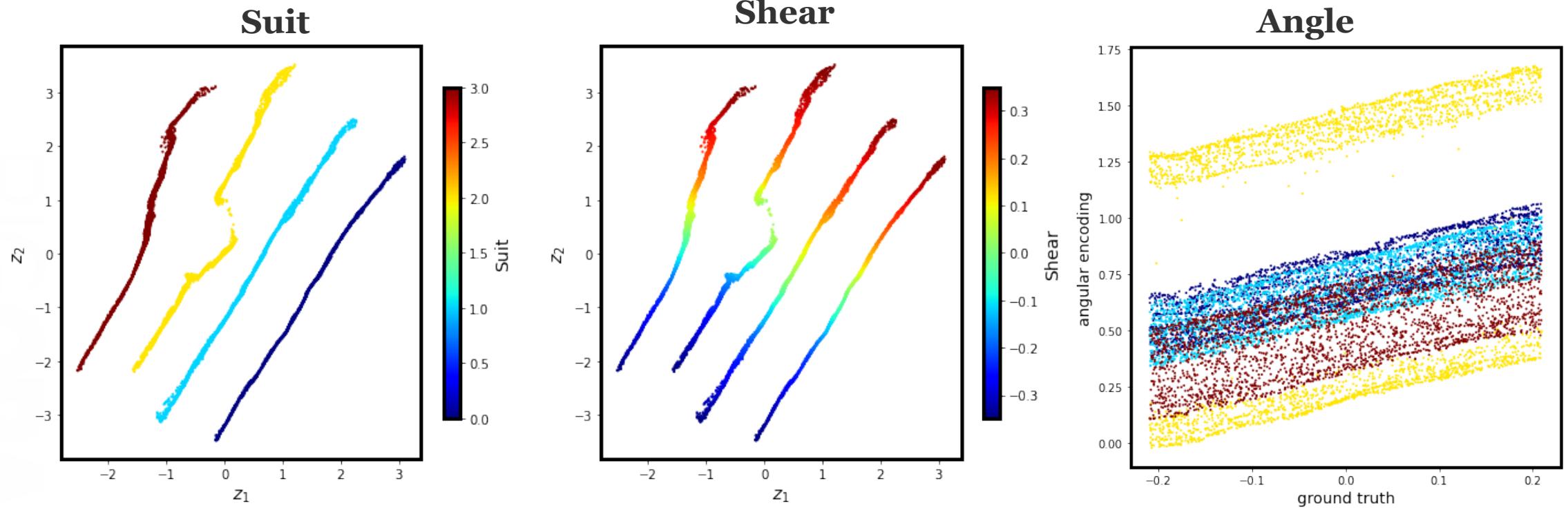


Latent representation



Cards 2: Low rotation (12 deg) and high shear (20 deg)

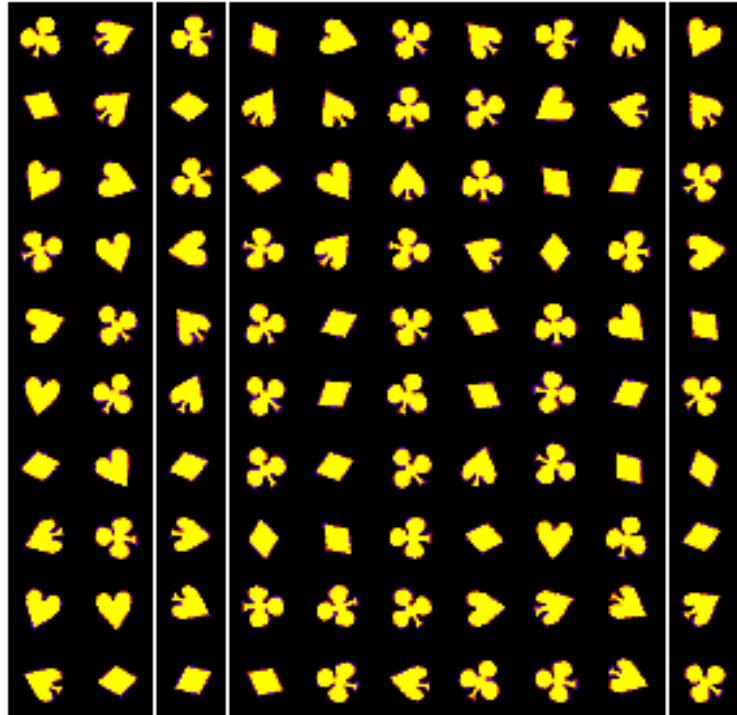
# rVAE on Cards



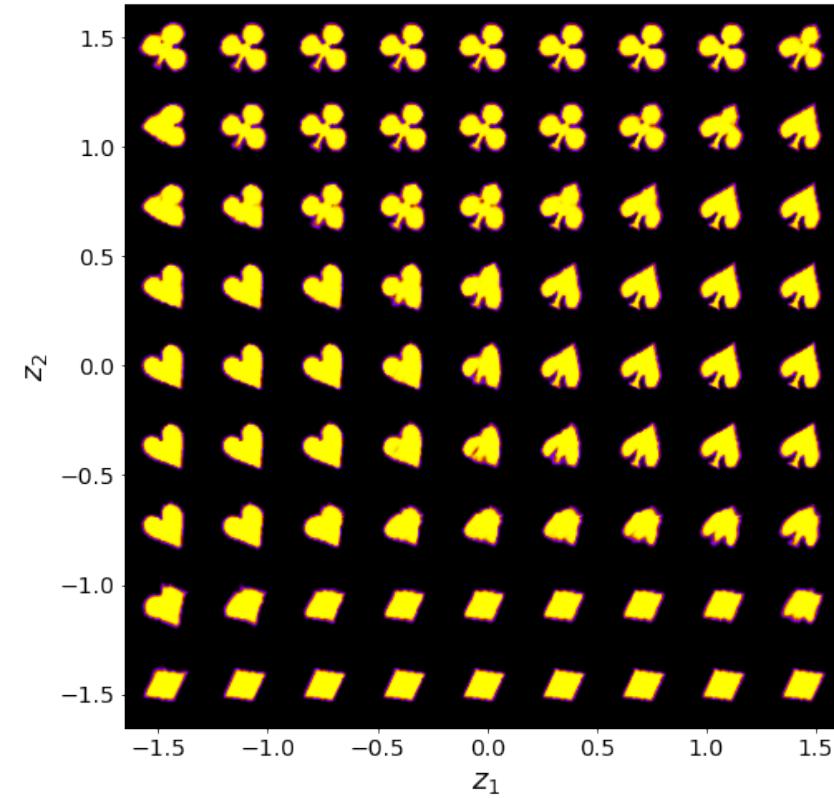
**Cards 2:** Low rotation (12 deg) and high shear (20 deg)

# rVAE on Cards

Example of data

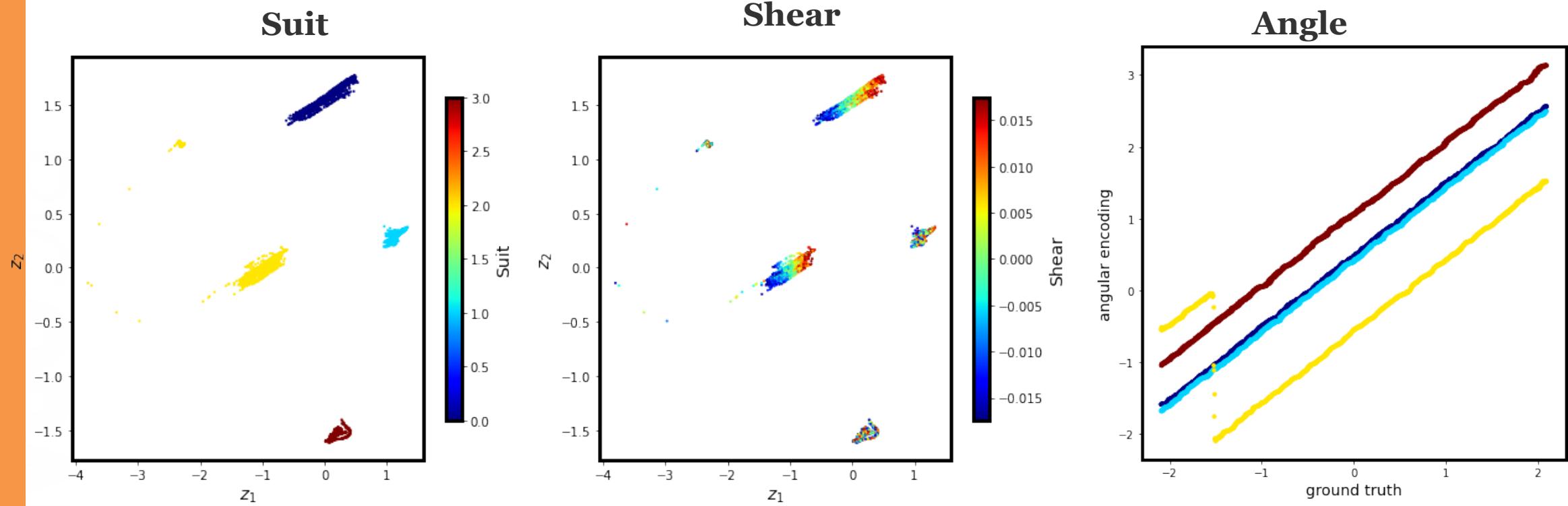


Latent representation



Cards 3: High rotation (120 deg) and low shear (1 deg)

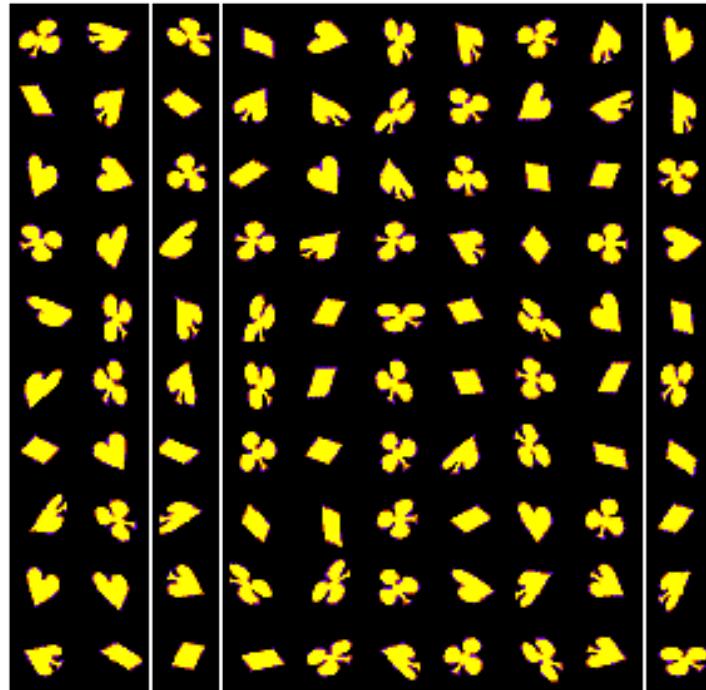
# rVAE on Cards



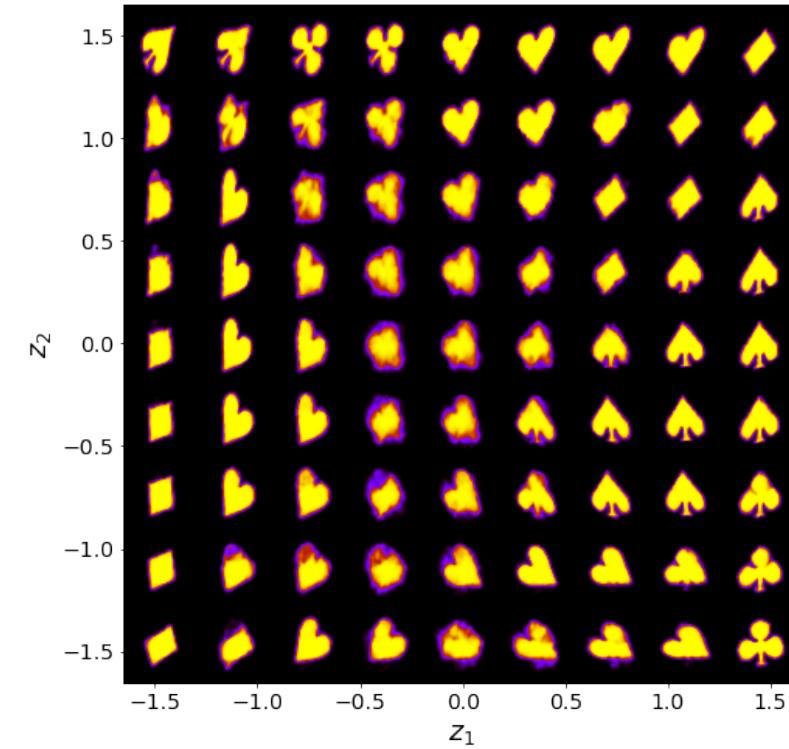
**Cards 3:** High rotation (120 deg) and low shear (1 deg)

# rVAE on Cards

Example of data

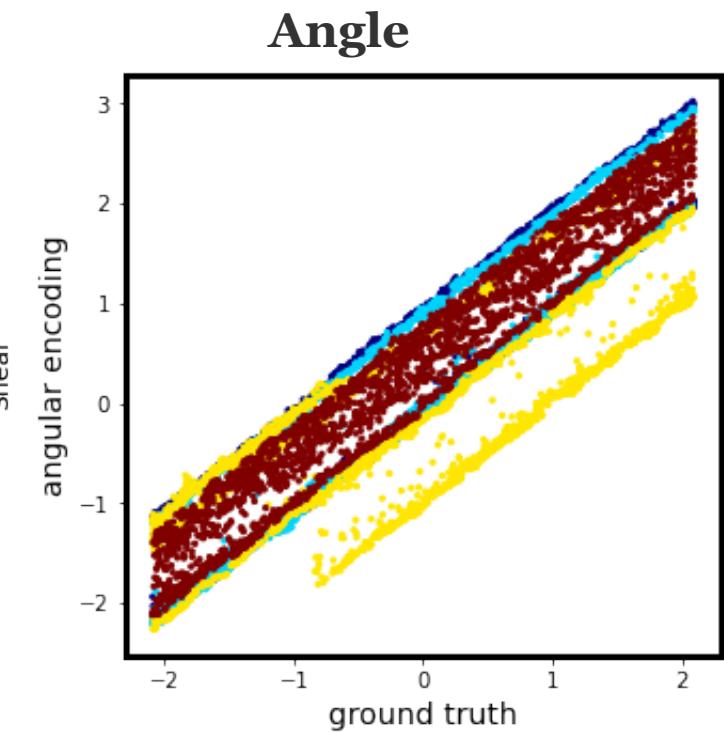
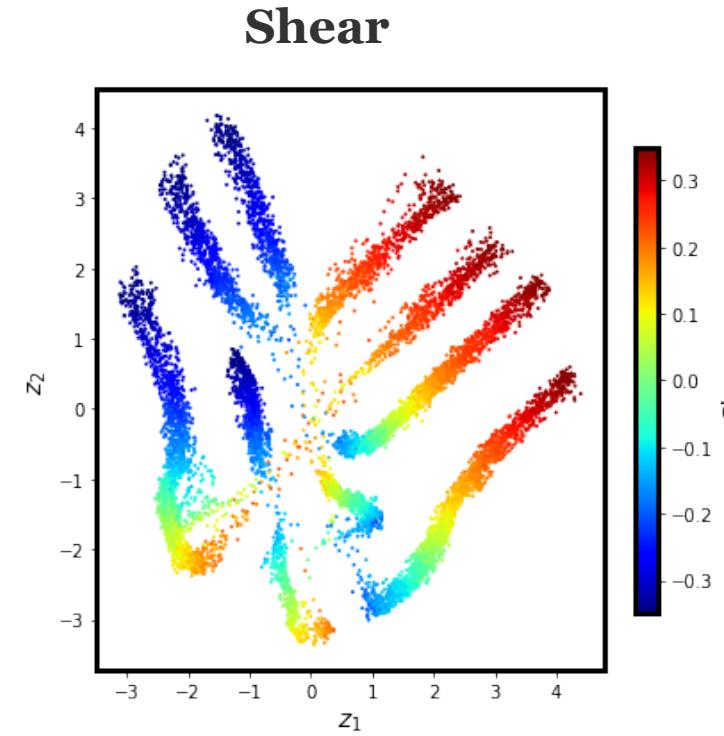
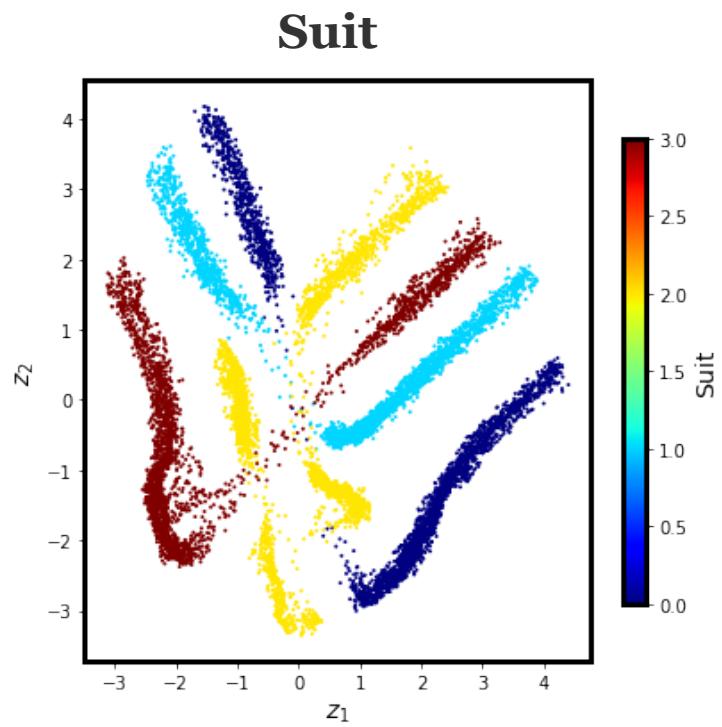


Latent representation



Cards 4: High rotation (120 deg) and high shear (20 deg)

# rVAE on Cards



**Cards 4:** High rotation (120 deg) and high shear (20 deg)