Query Selection based on Latent Space Sampling

Benjamin Killeen

Outline

Motivation: Supervised Learning

- **Big data** facilitates supervised learning.

Motivation: Supervised Learning

- **Big data** facilitates supervised learning.

Figure: ImageNet examples [2].

Motivation: Supervised Learning

- **Big data** facilitates supervised learning.
- Labels are necessary for **specialized** tasks.

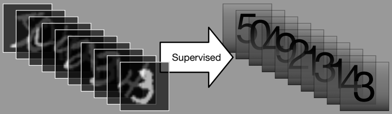


Figure: MNIST classification [1].

Figure: ImageNet examples [2].

Unsupervised Learning

- Understanding the underlying **structure** of data.

Unsupervised Learning

- Understanding the underlying **structure** of data.
- Avoids the high **cost** of labeling.

Unsupervised Learning

- Understanding the underlying **structure** of data.
- Avoids the high **cost** of labeling.

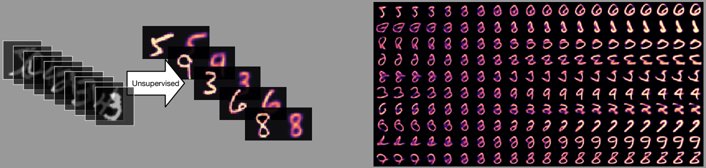


Figure: latent space learned by an auto-encoder.

Semi-supervised Learning

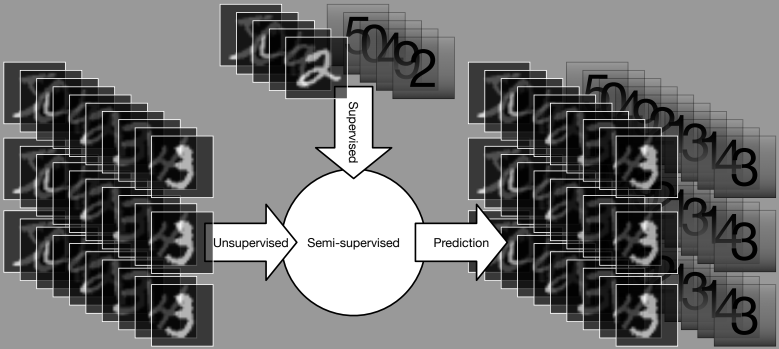


Figure: semi-supervised learning.

Starting from Scratch

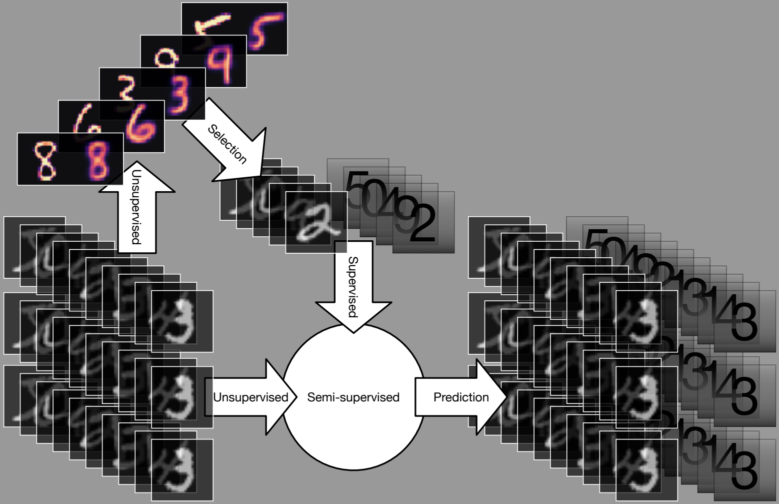


Figure: semi-supervised learning.

Starting from Scratch

- Supervised learning excels where **labels are available**.

Starting from Scratch

- Supervised learning excels where **labels are available**.
- Many domains require **starting from scratch** with absolutely no labels.

Starting from Scratch

- Supervised learning excels where **labels are available**.
- Many domains require **starting from scratch** with absolutely no labels.
- *e.g.* scientific image analysis.

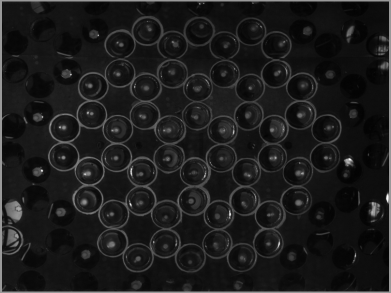


Figure: gyroscope tracking [3].

Method

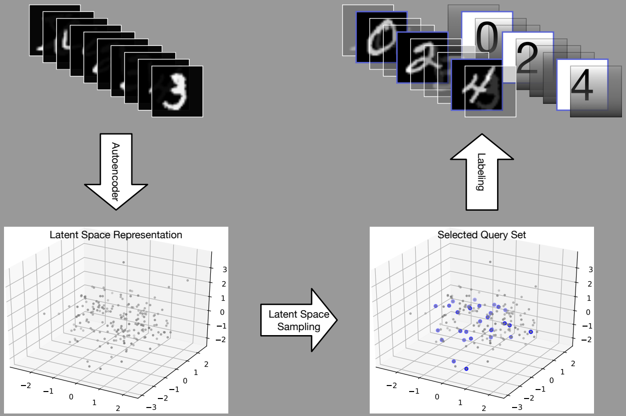


Figure: high-level overview

Method

- **Goal:** explore learned latent space $\mathbb{R}^L$ representations for optimal query selection.

Method

- **Goal:** explore learned latent space $\mathbb{R}^L$ representations for optimal query selection.
- **Constraint:** $L = 2$

Method

- **Goal:** explore learned latent space $\mathbb{R}^L$ representations for optimal query selection.
- **Constraint:** $L = 2$
- **Train** multiple auto-encoders:
  - Convolutional [4]
  - Variational (VAE) [5]
  - "t-SNE style" [6]
  - "Variational t-SNE style"

Method

- **Goal:** explore learned latent space $\mathbb{R}^L$ representations for optimal query selection.
- **Constraint:** $L = 2$
- **Train** multiple auto-encoders:
  - Convolutional [4]
  - Variational (VAE) [5]
  - "t-SNE style" [6]
  - "Variational t-SNE style"
- **Explore** multiple sampling strategies:
  - Distribution based
  - Cluster based (hierarchical)
  - Performance based

Convolutional Auto-encoder

- 3×3 convolution with 64 kernels (3x)
- 2×2 max-pooling
- 3×3 convolution with 32 kernels (3x)
- 2×2 max-pooling
- 3×3 convolution with 32 kernels (3x)
- 2×2 max-pooling
- 1024-node dense layer (2x)
- L-node dense layer, no activation
- 1024-node dense layer (2x)
- 3×3 convolution with 64 kernels (3x)
- 2×2 transpose convolution
- 3×3 convolution with 32 kernels (3x)
- 2×2 transpose convolution
- 3×3 convolution with 32 kernels (3x)
- 2×2 transpose convolution

Architectures

- VAE:
  - $2L$-node dense layer.
  - Normal-distribution sampling layer.
  - loss: binary cross-entropy with KL divergence from normal $\mathrm{KL}(Z||\mathcal{N})$.

Architectures

- VAE:
  - $2L$-node dense layer.
  - Normal-distribution sampling layer.
  - loss: binary cross-entropy with KL divergence from normal $\mathrm{KL}(Z||\mathcal{N})$.
- "t-SNE style"
  - loss: binary cross, entropy with added $\mathrm{KL}(P||Q)$ for batches of 64

Architectures

- VAE:
  - $2L$-node dense layer.
  - Normal-distribution sampling layer.
  - loss: binary cross-entropy with KL divergence from normal $\mathrm{KL}(Z||\mathcal{N})$.
- "t-SNE style"
  - loss: binary cross, entropy with added $\mathrm{KL}(P||Q)$ for batches of 64
- "Variational t-SNE style":
  - $2L$-node dense layer.
  - Normal-distribution sampling layer.
  - loss: combination of binary cross-entropy, VAE-like $\mathrm{KL}(Z||\mathcal{N})$, and t-SNE style $\mathrm{KL}(P||Q)$.

Sampling

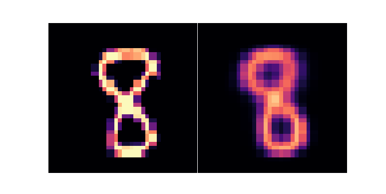
Algorithm 1 Select examples from the encoding Z according to a distribution f.

Require: encoding $Z \in \mathbb{R}^L$, probability distribution $f : \mathbb{R}^L \to \mathbb{R}$, distance metric $d : \mathbb{R}^{2 \times L} \to \mathbb{R}$, distance threshold $t \in \mathbb{R}$, $n \in \mathbb{N}$.

```
Q ← {}
while |Q| < n do
    Draw z ∼ f
    for z_i ∈ Z do                    ▷ the ith row of Z
        if d(z, z_i) < t then
            Q ← Q ∪ {i}
        end if
    end for
end while
return Q
```

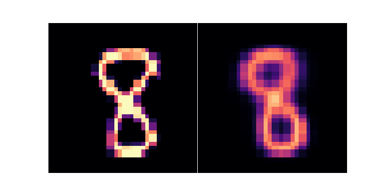
Results

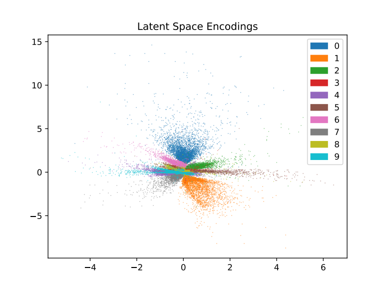
- Reconstructions



Results

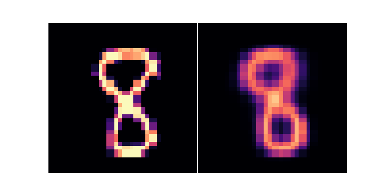
- Reconstructions
- Encodings

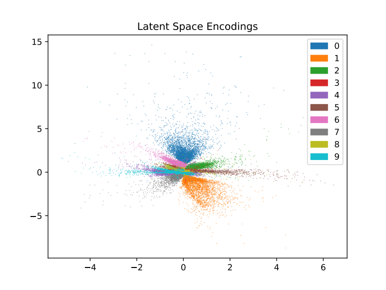


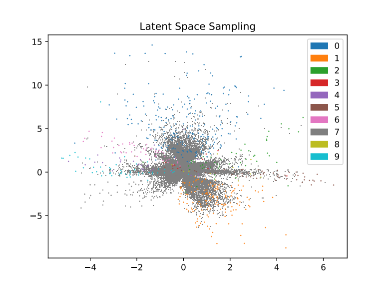


Results

- Reconstructions
- Encodings
- Samplings

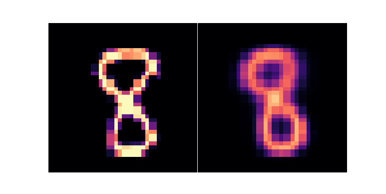


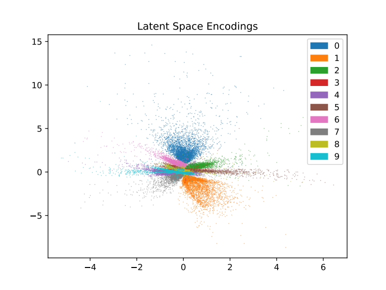


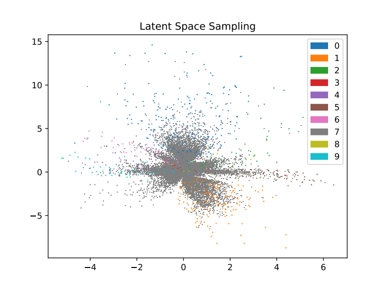


Results

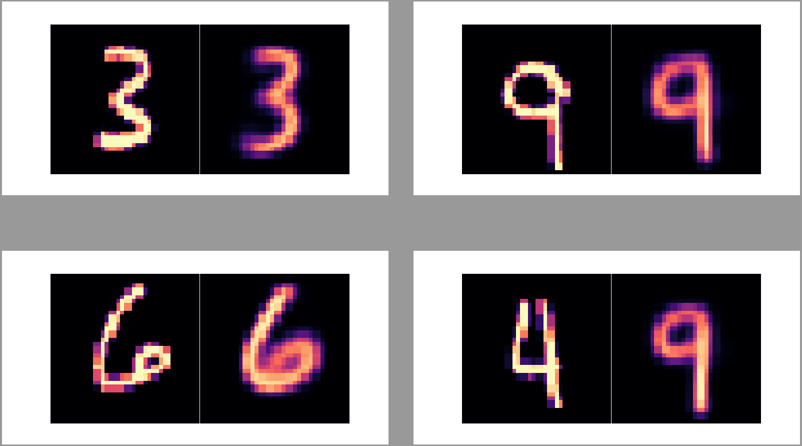
- Reconstructions
- Encodings
- Samplings



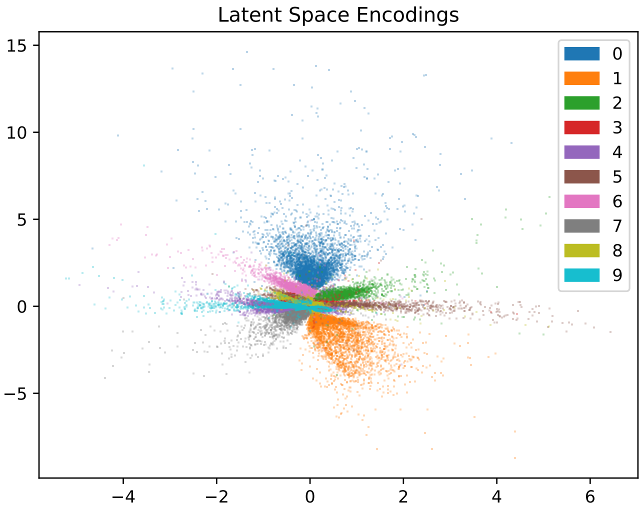




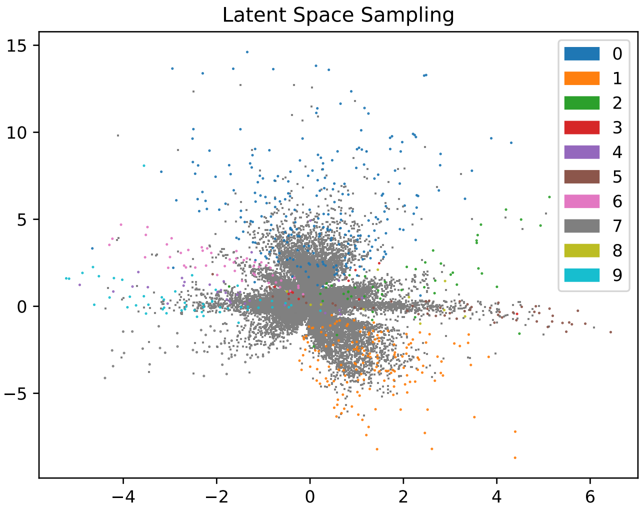
Convolutional Auto-encoder



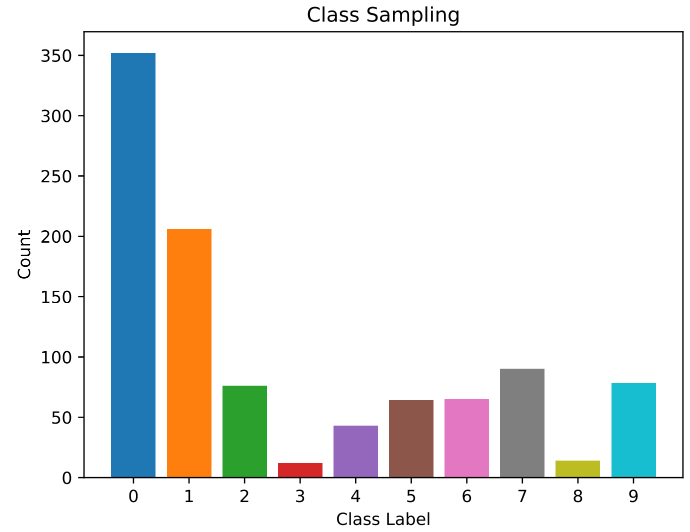
Convolutional Auto-encoder



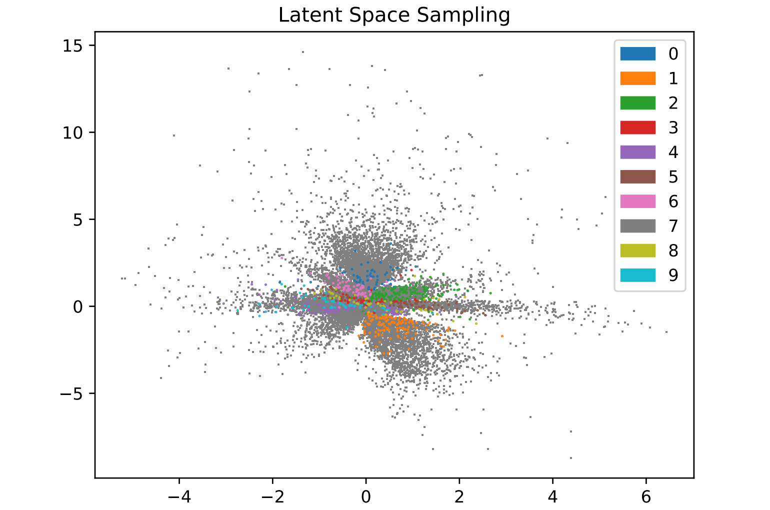
Uniform Spatial Sampling



Uniform Spatial Sampling



Normal Spatial Sampling

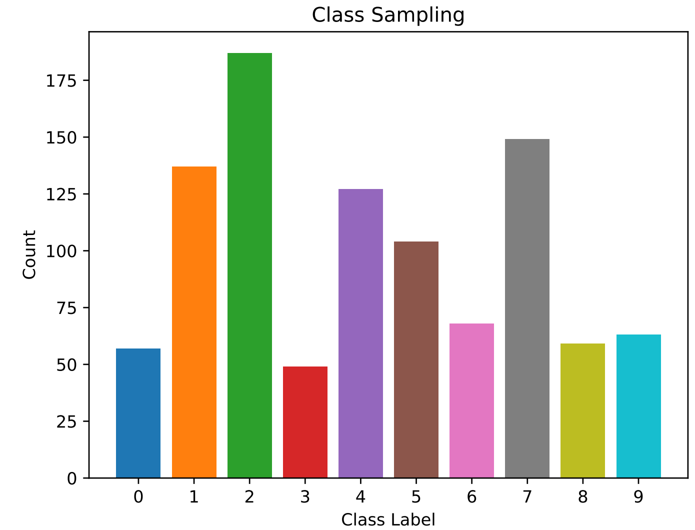


Normal Spatial Sampling

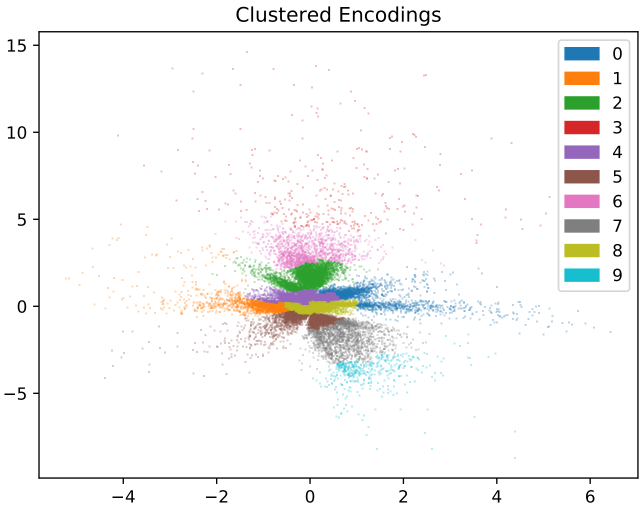
Query Selection based on Latent Space Sampling

Benjamin Killeen

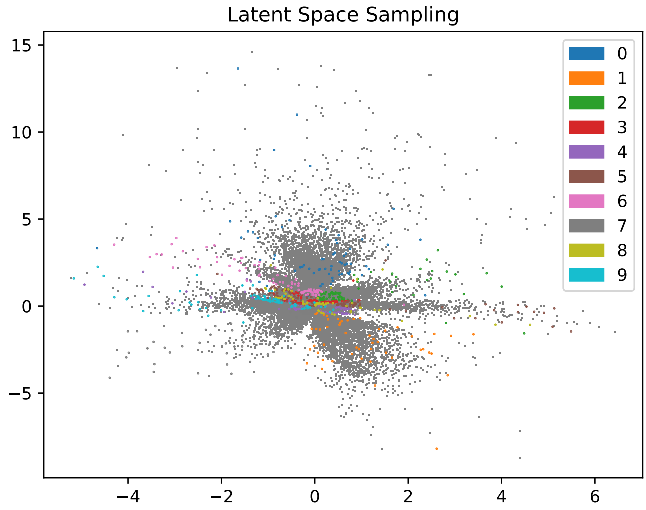
- Motivation
- Method
- Results
  - Convolutional AE
  - Variational AE
  - t-SNE Style AE
  - Variational t-SNE Style AE
  - Reconstruction Error Sampling
- Discussion
- Future Work
- References



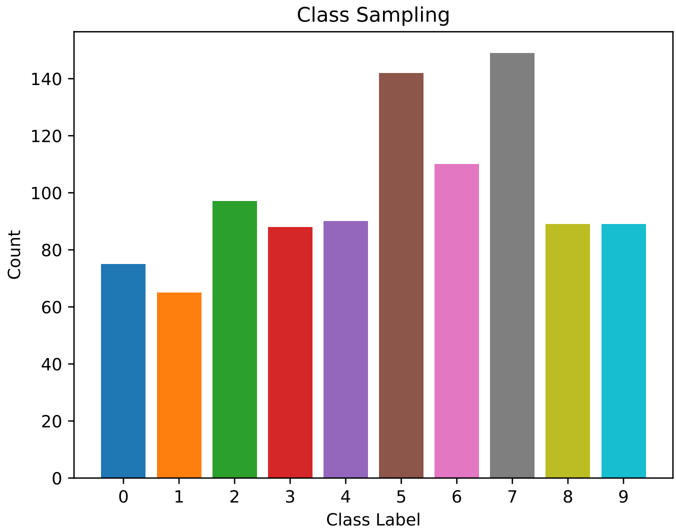
Clustering



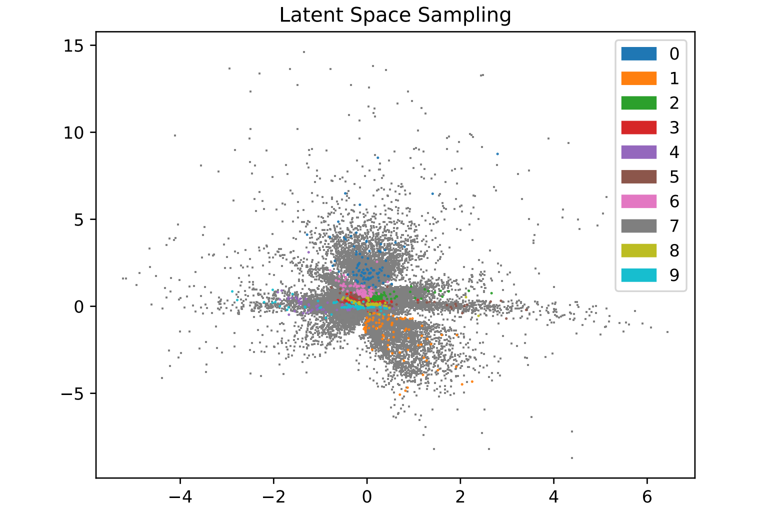
Cluster Sampling



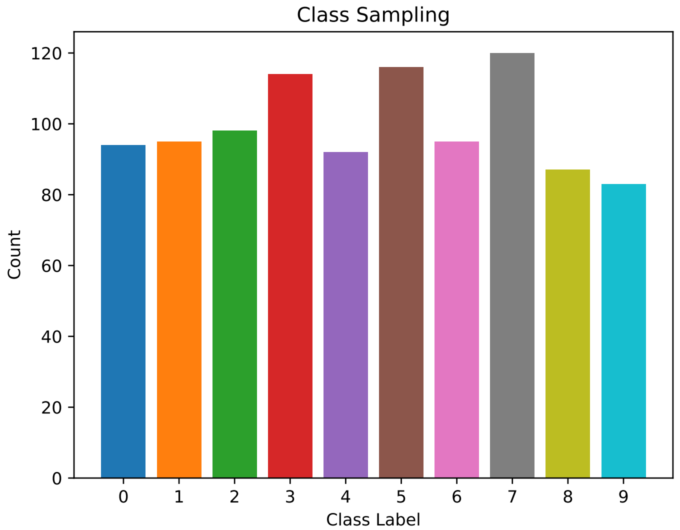
Cluster Sampling



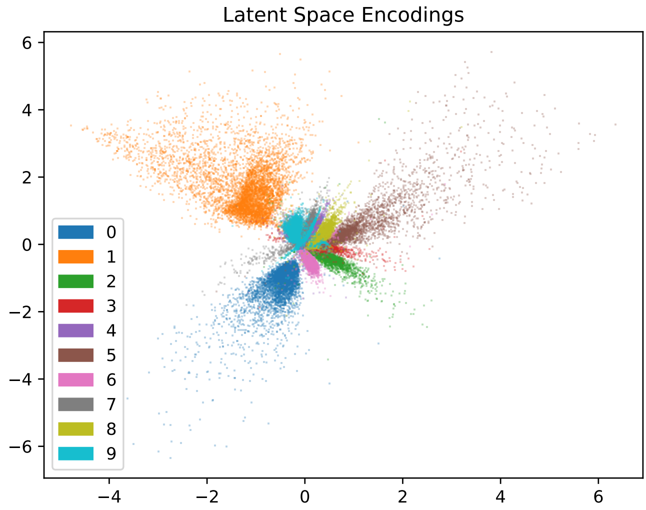
Normal Sampling inside Clusters



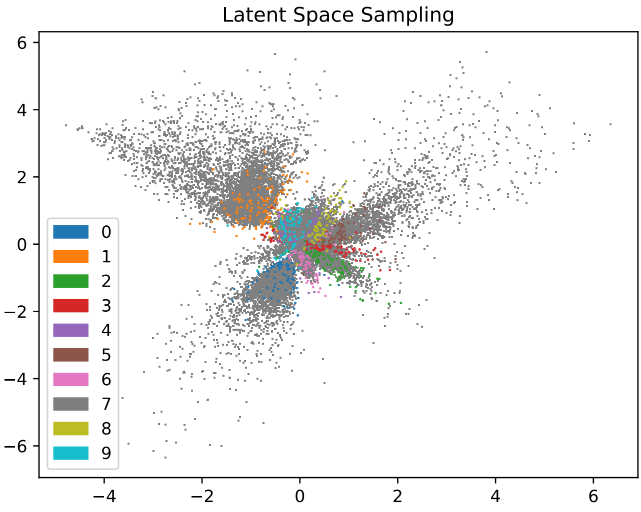
Normal Sampling inside Clusters



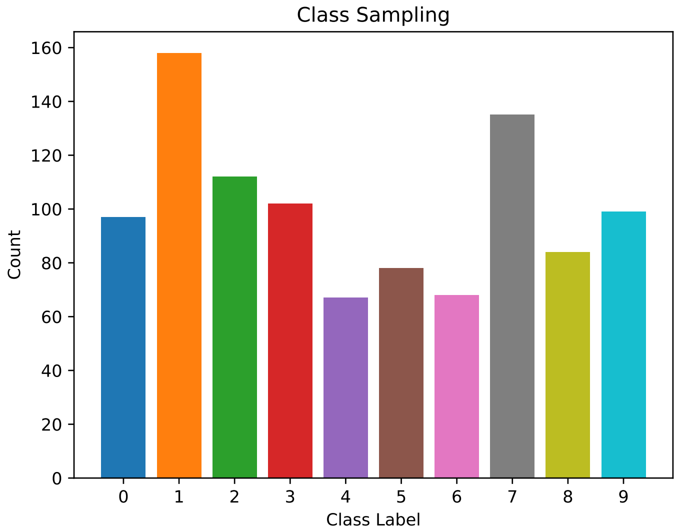
Variational Auto-encoder



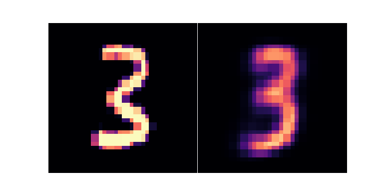
Normal Spatial Sampling

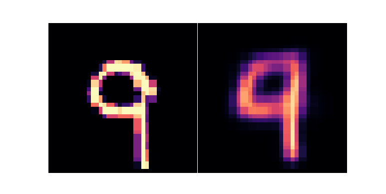


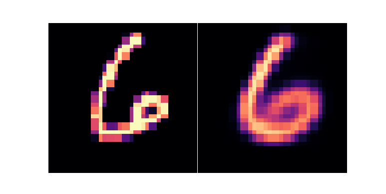
Normal Spatial Sampling

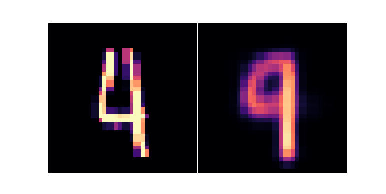


t-SNE Style Auto-encoder

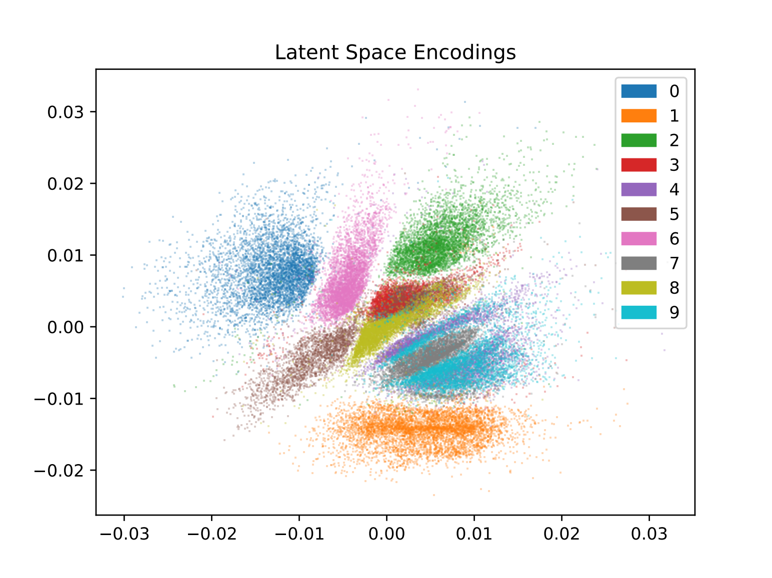




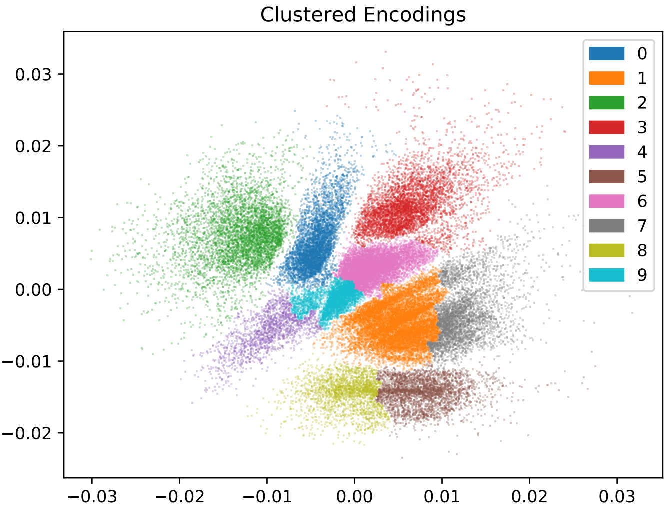




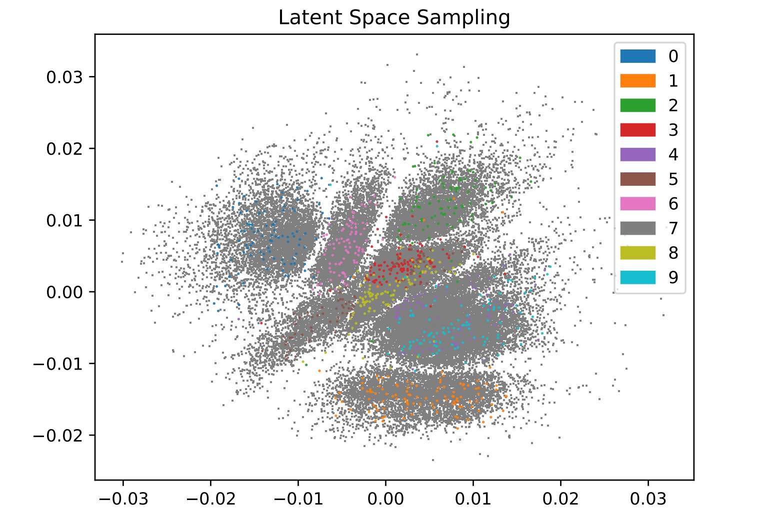
t-SNE Style Auto-encoder



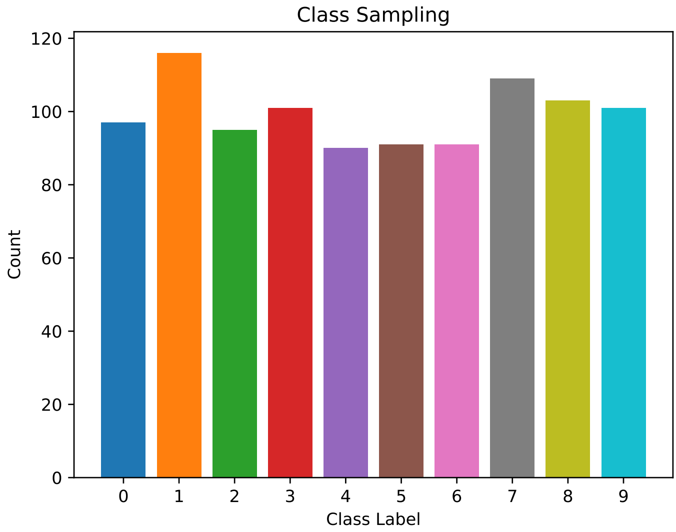
Clustering


Clustered Encodings
0
1
2
3
4
5
6
7
8
9
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
−0.02
−0.01
0.00
0.01
0.02
0.03

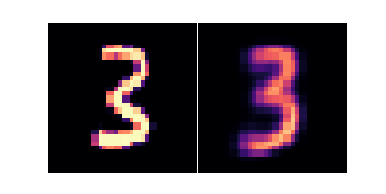
Normal Sampling inside Clusters

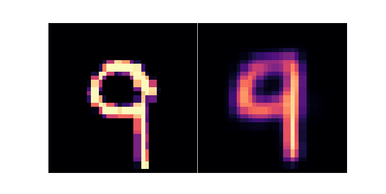


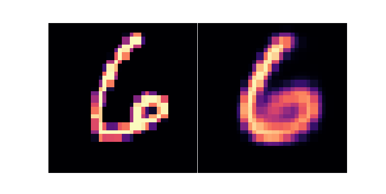
Clustered Normal Spatial Sampling

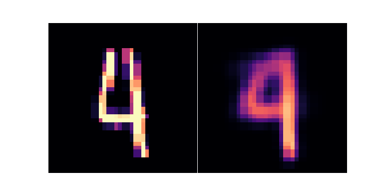


Variational t-SNE Style Auto-encoder

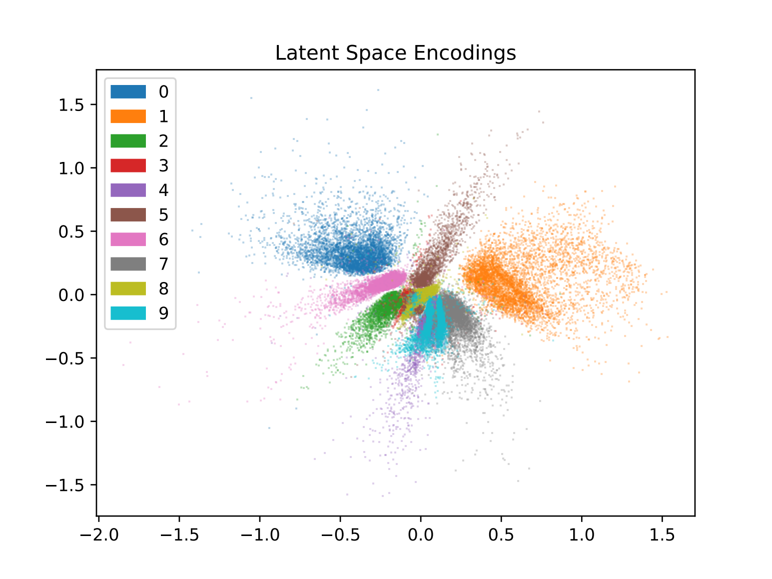




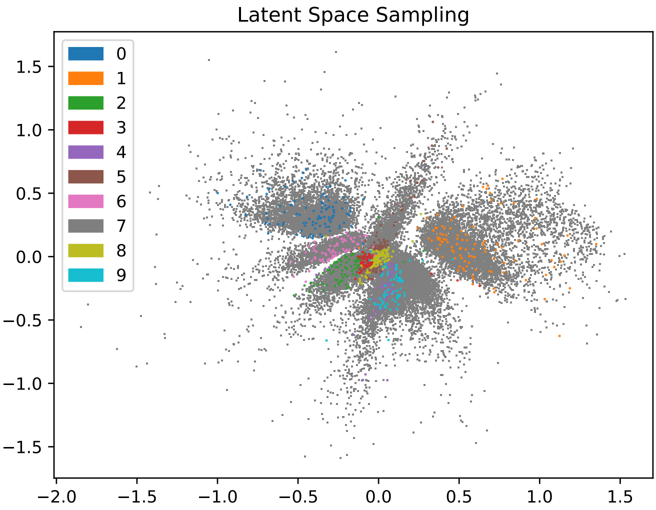




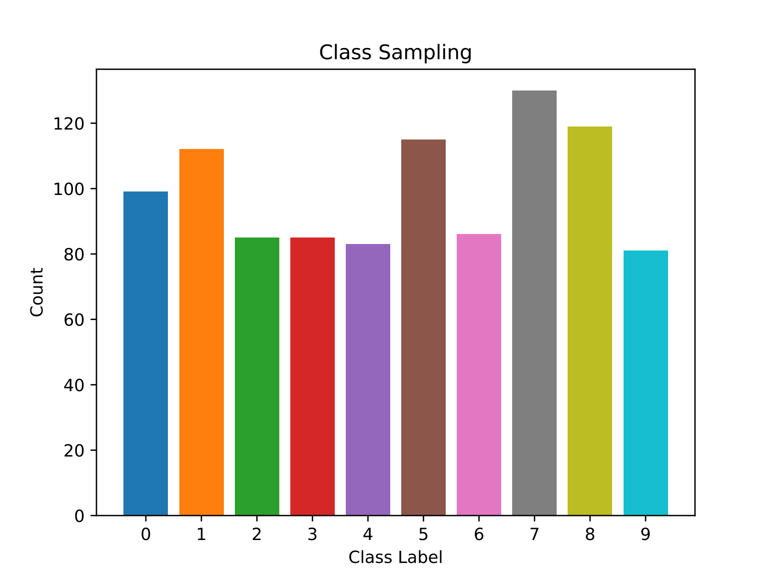
Variational t-SNE Style Auto-encoder



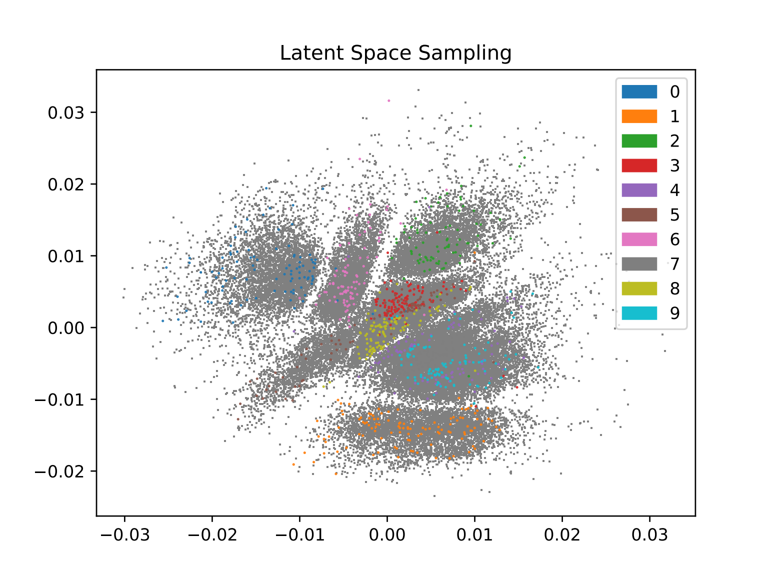
Normal Sampling inside Clusters



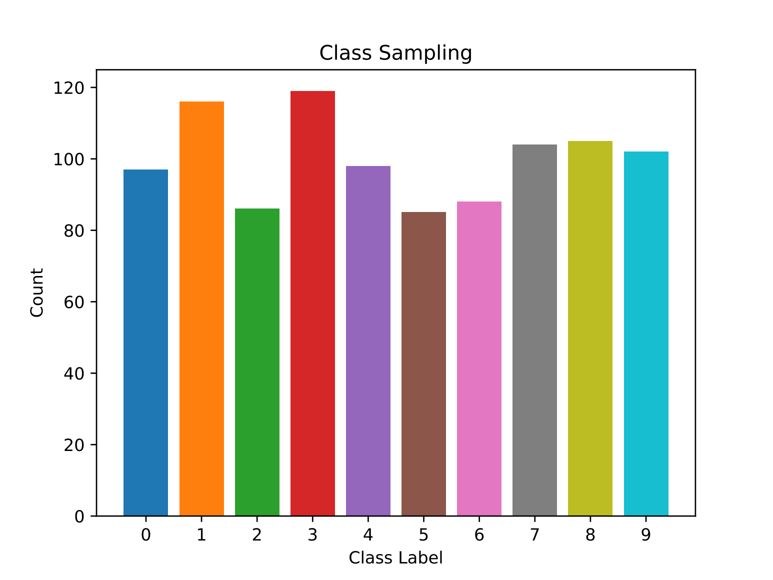
Clustered Normal Spatial Sampling



Reconstruction Error Sampling



Reconstruction Error Sampling



Discussion

- Normal sampling inside clusters yields **balanced samples**.

Discussion

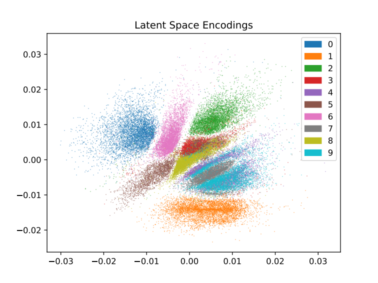
- Normal sampling inside clusters yields **balanced samples**.
- Are auto-encoders the **best choice** for latent-space representation?

Discussion

- Normal sampling inside clusters yields **balanced samples**.
- Are auto-encoders the **best choice** for latent-space representation?
- Distribution of **difficult examples** in latent space.

Discussion

- Normal sampling inside clusters yields **balanced samples**.
- Are auto-encoders the **best choice** for latent-space representation?
- Distribution of **difficult examples** in latent space.



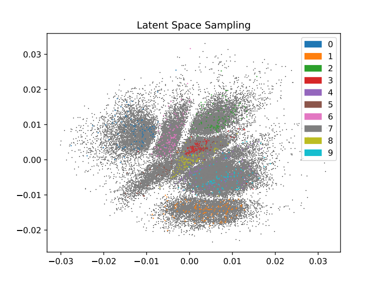


Figure: Visualizing points which are difficult to learn.

Future Work

- Transfer learning **without overfitting**.

Future Work

- Transfer learning **without overfitting**.
- Evaluate performance of classifiers **trained on query selections**.

Future Work

- Transfer learning **without overfitting**.
- Evaluate performance of classifiers **trained on query selections**.
- Explore higher dimensions.

Future Work

- Transfer learning **without overfitting**.
- Evaluate performance of classifiers **trained on query selections**.
- Explore higher dimensions.
- Labeling **artificial examples** produced by an auto-encoder.

References I

Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, and others, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," p. 8.

L. M. Nash, D. Kleckner, A. Read, V. Vitelli, A. M. Turner, and W. T. M. Irvine, "Topological mechanics of gyroscopic metamaterials," *Proceedings of the National Academy of Sciences*, vol. 112, no. 47, pp. 14 495–14 500, 2015. [Online]. Available: http://www.pnas.org/content/112/47/14495

A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *arXiv:1312.6114 [cs, stat]*, Dec. 2013, arXiv: 1312.6114. [Online]. Available: http://arxiv.org/abs/1312.6114

L. Maaten, "Learning a parametric embedding by preserving local structure," in *Artificial Intelligence and Statistics*, 2009, pp. 384–391.