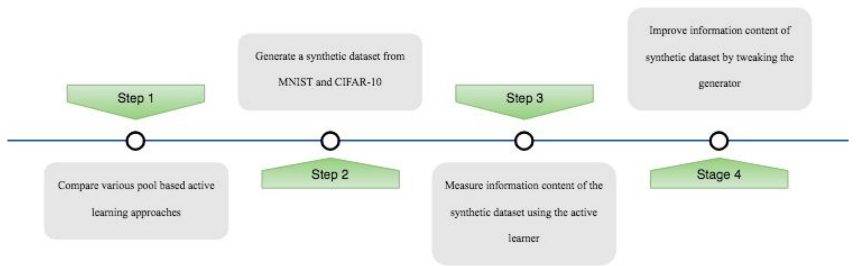
GENERATING INFORMATIVE SAMPLES USING GANs

Anirudha Jitani, Amit Sinha, Deepak Sharma

Objectives



Active Learning

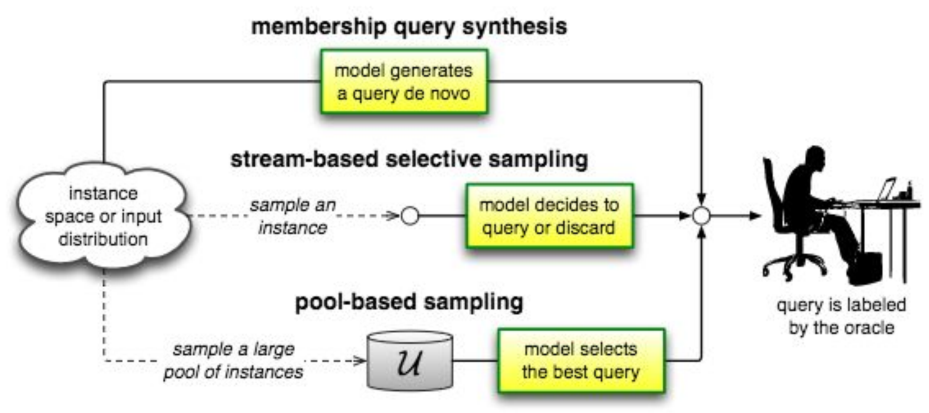


Figure - Active Learning Pipeline [6]

Active learning systems attempt to overcome the labelling bottleneck by asking queries in the form of unlabelled instances to be labeled by an oracle. Active learning is especially useful in applications where the unlabelled data is present in abundance, but obtaining the labels for such datasets is expensive. A good active learner is able to achieve greater performance by choosing the examples that it will learn the most from.

Acquisition Functions

Uncertainty Sampling:

- Least Confidence :
  When the posterior probability for the best class is low, the uncertainty for the selected samples is the highest. The n samples with the highest scores (least confident predictions) are taken for training [4, 5].

$$\phi^{LC}(\mathbf{x}) = 1 - \max_{y \in Y} P(y|\mathbf{x};\theta)$$

- Margin:
  When the difference between posteriors of the best class and the second best class is low, the model is uncertain about the sample between the two classes. The n samples with the highest scores (the smallest margin) are taken for training.

$$\phi^{M}(\mathbf{x}) = -(\max_{y \in C} P(y|\mathbf{x};\theta) - \max_{\substack{y \in Y \\ y \neq y_{max}}} P(y|\mathbf{x};\theta))$$

- Token Entropy:
  The entropy for each sample is indicative of the information content that each sample provides to the model. It is calculated through the standard entropy formula over the posterior prediction. The n samples with the highest score (the most information) are taken for training [4, 5].

$$\phi^{TE}(\mathbf{x}) = -\sum_{y_i \in Y} P(y_i|\mathbf{x}) \log(P(y_i|\mathbf{x}))$$

Uncertainty - Density Sampling:

The concept of density [3, 4, 5] is used to find points that are uncertain, while also similar to many other points. In other words, we look for points that are not outliers, but at the same time the model is uncertain about these data points.

The similarity is evaluated by considering a similarity metric (like cosine) [3,5] between a sample and its neighbors and then taking the average as the score.

- Density - Max Uncertainty
  To find most informative samples we pick the ones which give a high density and uncertainty score (based on least confidence).

$$\phi^{DMU}(\mathbf{x}) = \frac{\sum_{s_i \in S(\mathbf{x})} \cos(\mathbf{x}, s_i)}{Num(S(\mathbf{x}))} \times \phi^{LC}(\mathbf{x})$$

- Density - Entropy
  To find the most informative samples we pick the ones which give a high density and information score (based on entropy).

$$\phi^{DE}(\mathbf{x}) = \frac{\sum_{s_i \in S(\mathbf{x})} \cos(\mathbf{x}, s_i)}{Num(S(\mathbf{x}))} \times \phi^{TE}(\mathbf{x})$$

CNN Architecture

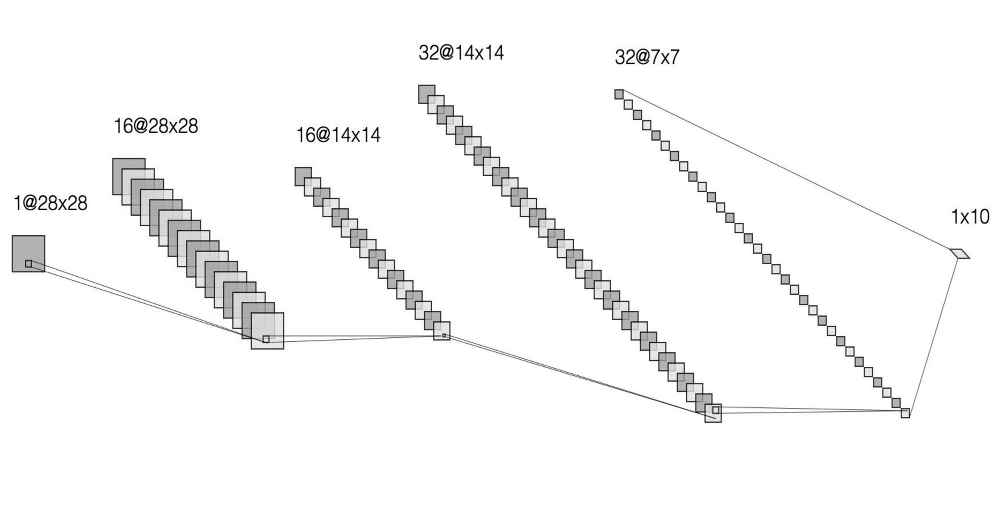
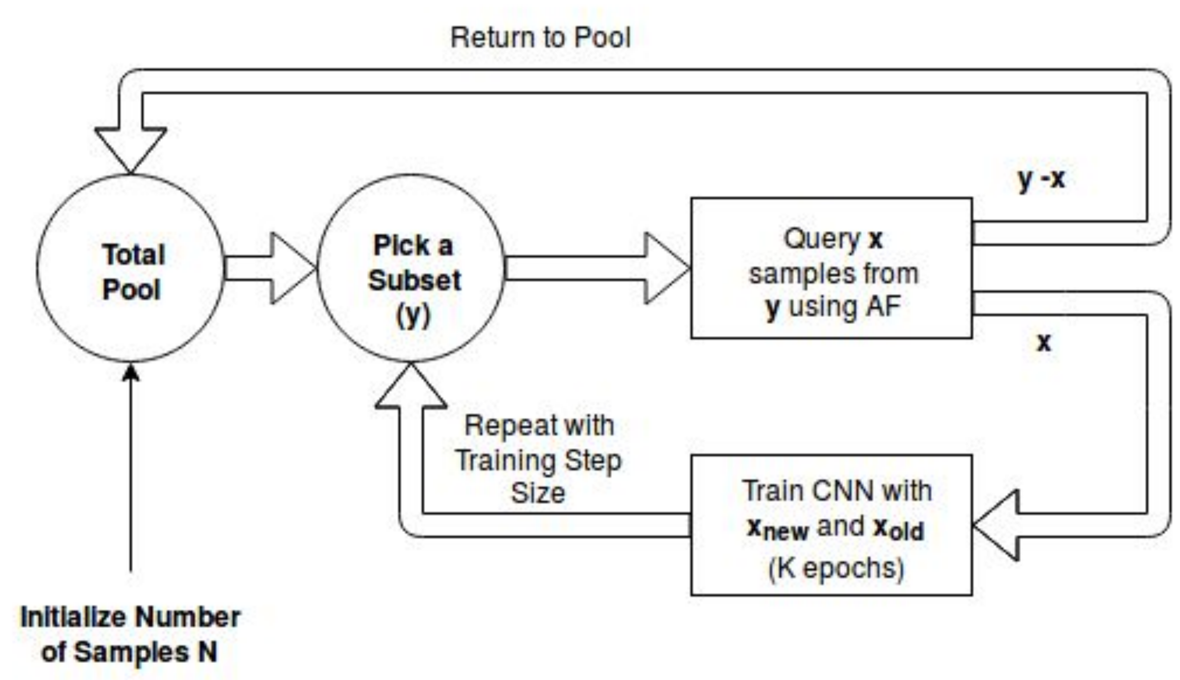


Figure - Convolutional Neural Network for Classification of MNIST Dataset

Experimental Setup



Training Results

MNIST **CIFAR10**

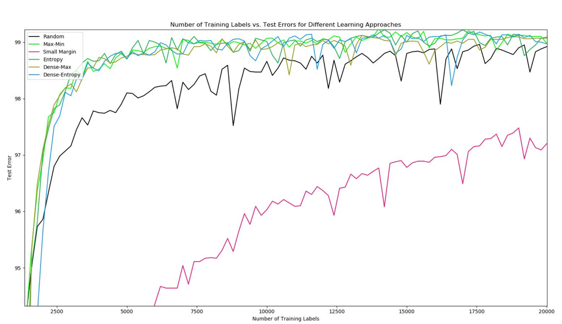


Figure - 2 Epochs per Acquisition Iteration

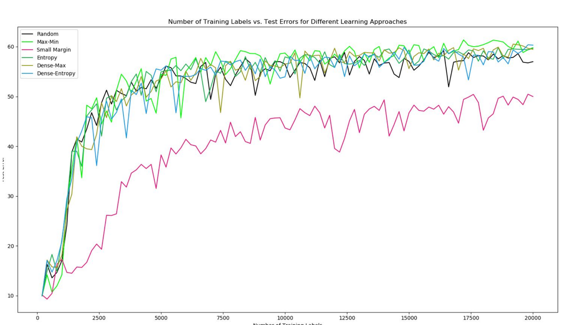


Figure - 2 Epochs per Acquisition Iteration

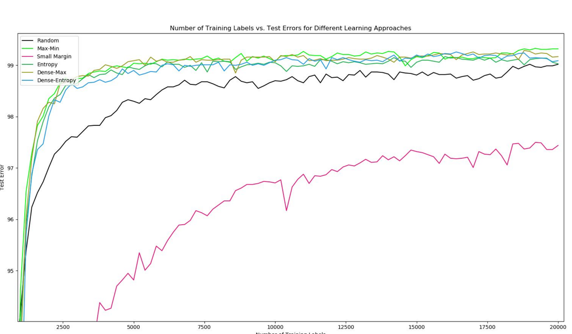


Figure - 5 Epochs per Acquisition Iteration

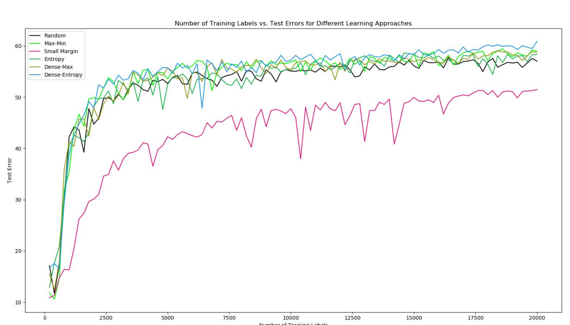


Figure - 5 Epochs per Acquisition Iteration

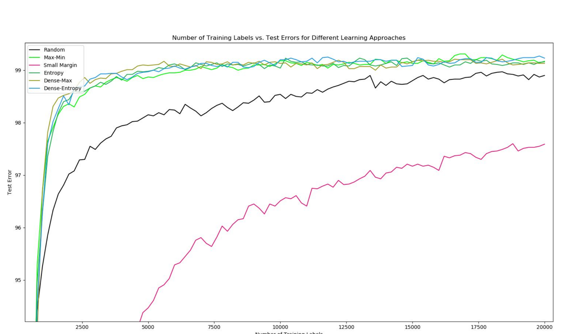


Figure - 10 Epochs per Acquisition Iteration

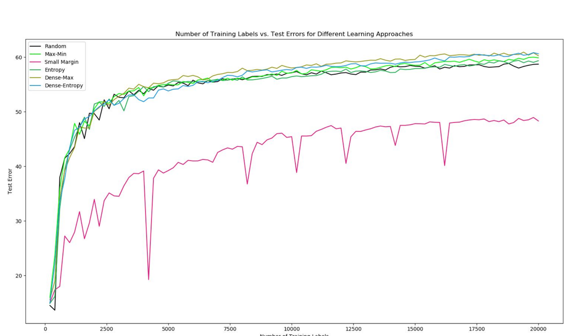


Figure - 10 Epochs per Acquisition Iteration

DC-GAN Architecture

Modifications to GAN architecture

1. Use strided convolution instead of spatial pooling functions, allowing the network to learn its own spatial downsampling
2. Eliminating fully connected layers on top of convolutional features.
3. Batch Normalization to stabilize learning. However, to avoid model instability, batchnorm was not applied to generator output layer and discriminator input layer.

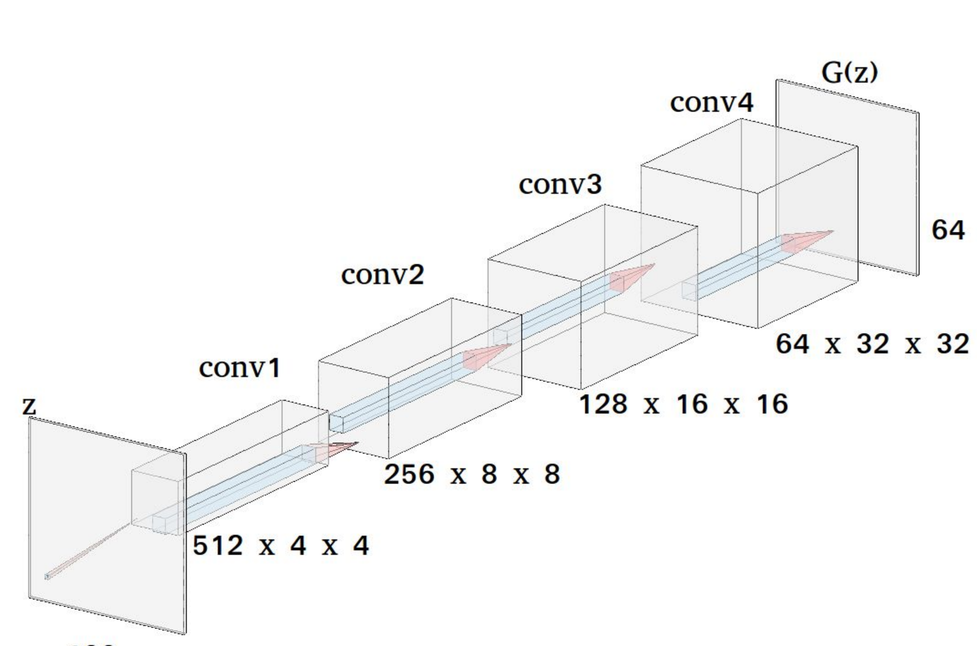


Figure - DCGAN Generator [1] used for MNIST data generation. A 100 dimensional uniform distribution Z is projected to a small spatial extent convolutional representation with many feature maps. A series of four fractionally-strided convolutions then convert this high level representation into a 64 * 64 pixel image.

Generated Data

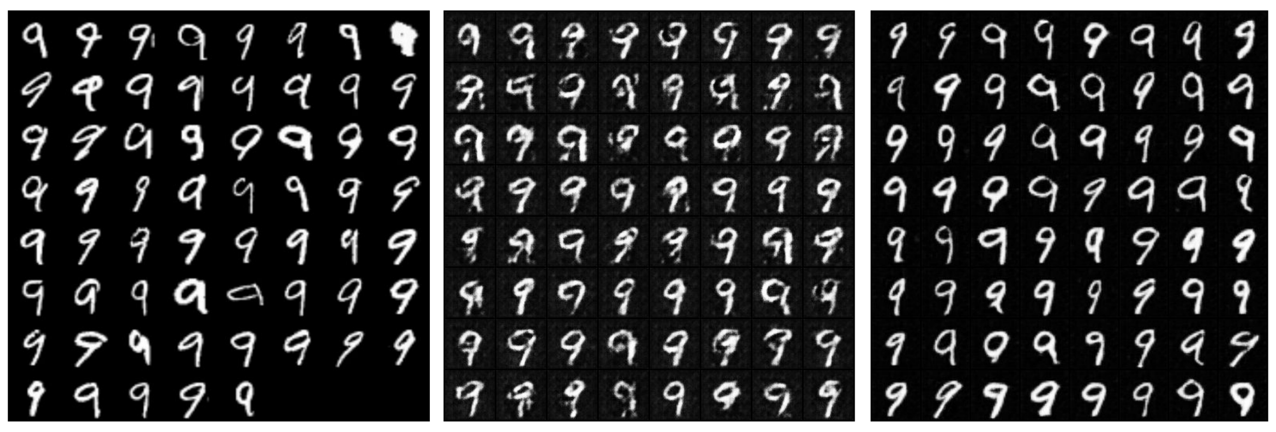


Figure 1 - The figure on the (left) shows the actual MNIST data, the image generated by the GAN after training for 5 epochs (center) and 25 epochs (right)

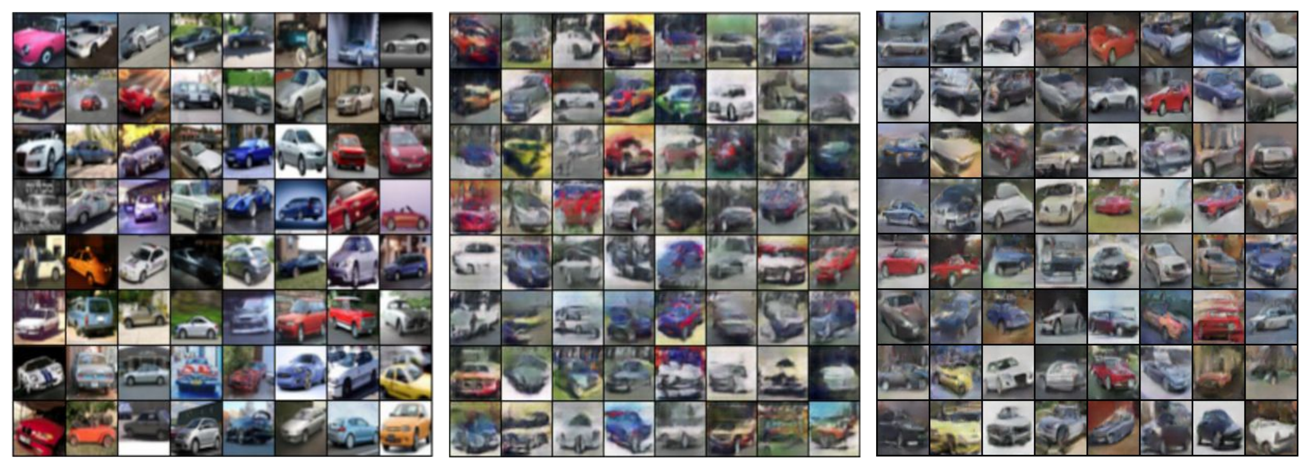


Figure 2 - The figure on the (left) shows real CIFAR data for category 'automobile', the image generated by the GAN after training for 35 epochs (center) and 49 epochs (right)

Conclusions

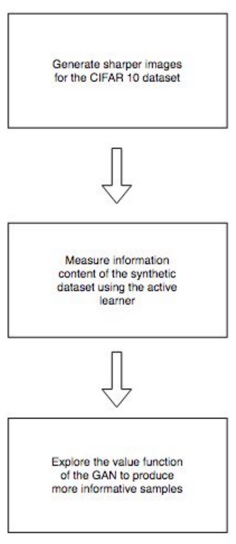
The curves for the acquisition functions based on CNNs has better performance (in terms of AULC) than the random sampling approach.

The gap between the random curves and the acquisition curves decreases as the number of epochs increases.

The learning curves smoothen out more when the number of epochs is higher.

The gap between the random curves and the acquisition function curves decreases as the dataset gets more complicated (This can be seen between CIFAR10 and MNIST).

Next Steps



References

1. Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434 (2015).
2. Shen, Xuehua, and ChengXiang Zhai. "Active feedback in ad hoc information retrieval." Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2005.
3. Zhu, Jingbo, et al. "Active Learning With Sampling by Uncertainty and Density for Data Annotations." IEEE Trans. Audio, Speech & Language Processing 18.6 (2010): 1323-1331.
4. Settles, Burr, and Mark Craven. "An analysis of active learning strategies for sequence labeling tasks." Proceedings of the conference on empirical methods in natural language processing. Association for Computational Linguistics, 2008.
5. Huijser, Miriam, and Jan C. van Gemert. "Active Decision Boundary Annotation with Deep Generative Models." Computer Vision (ICCV), 2017 IEEE International Conference on. IEEE, 2017.
6. Settles, Burr. "Active learning." *Synthesis Lectures on Artificial Intelligence and Machine Learning* 6.1 (2012): 1-114.