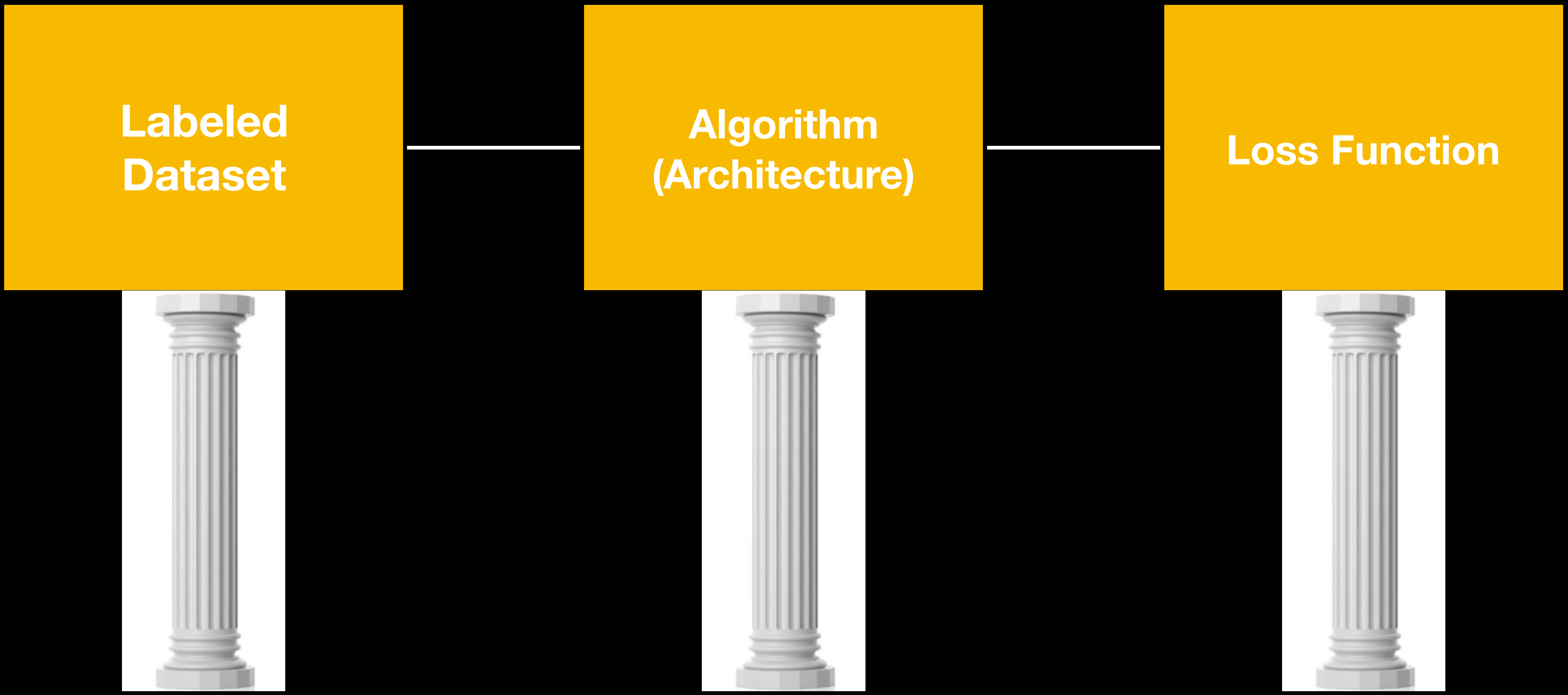
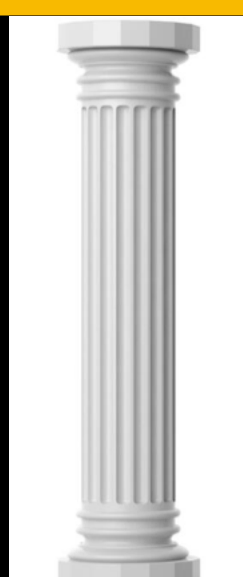
Supervised Learning: A Deeper Dive into Labeled Datasets and Loss Functions

Earl Wong


Labeled Dataset
Algorithm (Architecture)
Loss Function

Labeled Dataset



Labeled Dataset

- Supervised learning is defined by the presence of a labeled dataset.
- A network can learn valuable internal representations for downstream tasks such as image recognition, object detection, etc., using a labeled dataset.
- However, if your labeled dataset is “missing” information, your network cannot learn what is missing.
- But, a high performing network should be able to extrapolate / generalize to what is similar.

Example: MNIST

Your network cannot learn to correctly classify a “3 “or a “4”, if it has never seen a “3” or a "4”.

However, a high performing network should be able to classify different variants of the number 2 with high success.

Labeled Dataset

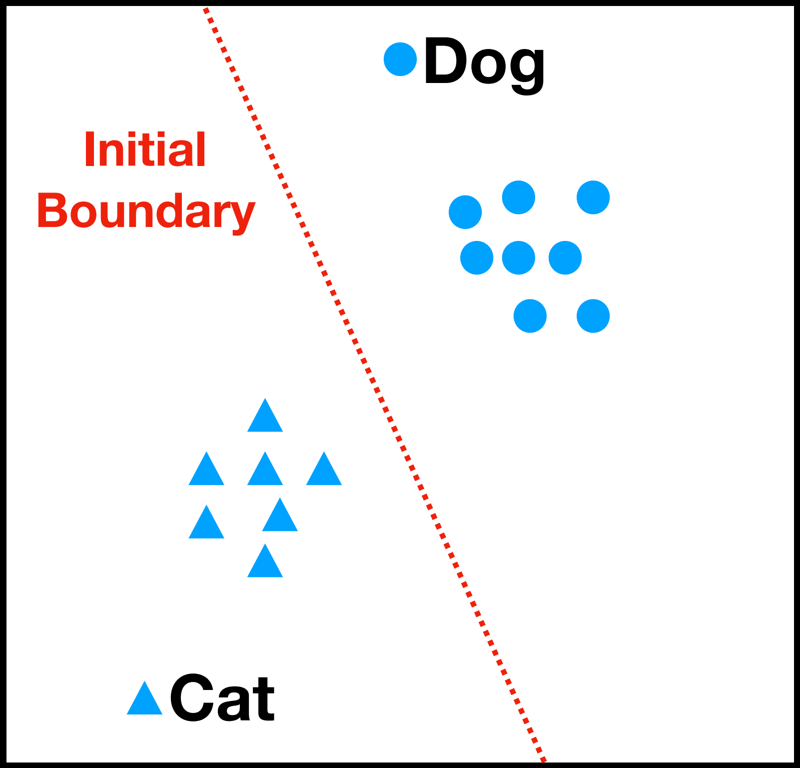
- In the deep learning era, more labeled data usually improves accuracy / performance.
- This is because deep networks are very data hungry.
- However, EVEN MORE IMPORTANT, is collecting the RIGHT labeled data.
- After all, a network can only learn what is represented in the labeled dataset.
- Hence, THINK INTENSELY AND DEEPLY, before making any request for additional labeled data.

An Avoidable Problem

- Example1: Dataset = 10,000 Dog images, 2 Cat Images
- If your **dataset is imbalanced**, your results will probably be misleading and poor.
- Misleading: If you guess dog every time, you will do very well.
- Poor: Unless you significantly re-weigh the errors associated with the 2 cat images, the dog examples will overwhelm the cat examples.

Better “Engineering” of Labeled Data

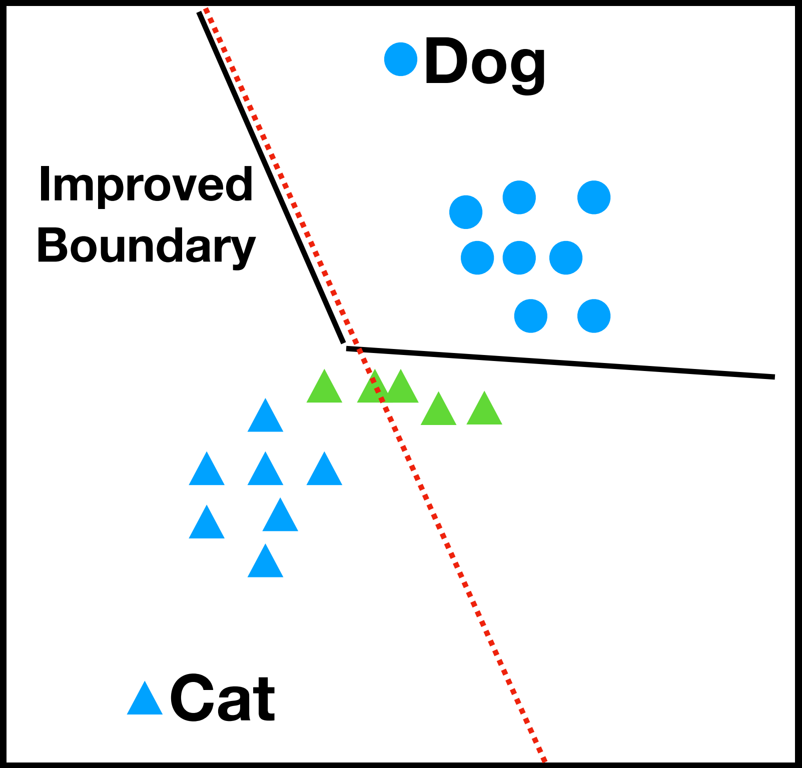
- Example2: Decision boundary for 1000 Easy Dog Images, 1000 Easy Cat Images = **no hard data mining**



Better "Engineering" of Labeled Data

- Example2: New Decision Boundary for 1000 Easy Dog Images, 500 Easy Cat Images + 500 Difficult Cat Images

- Improved learning results from using hard examples.



Augmentation Can Help

Augmentation = Creating More Usable Data

- Transformations: Flip, Crop, etc.
- Synthetic: Simulators & GANS (“Data Augmentation GANS” - 3/18)
- Semi-Supervised: Use labeled data to help label the unlabeled data. (“Self Training with Noisy Student” - 6/20)

Looking More Closely At the Labeled Data

- Naver AI Labs recently relabeled 1.28 million Imagenet training images. (How does one image and one labeled object affect classification outcomes?) (1/21)

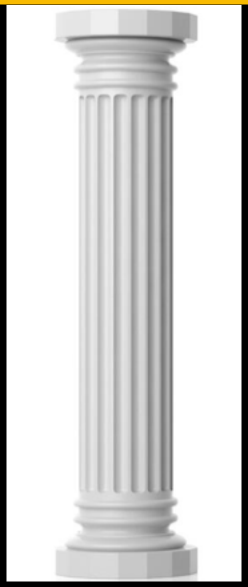
- ObjectNet challenge (How do we robustly address pose, location and angle of objects?) (12/19)

- Nitre challenge for super resolution (Are we even solving the right problem?) (5/20)

Moral

- Know your data very, very well.
- You cannot learn what is not represented / represented correctly.
- Think deeply, before requesting additional labeled data = $$$$$$.

Loss Function



Loss Function

How good is your teacher? = How are you trying to learn?

- Teacher = Linear regression: MSE or MAE Loss
- Teacher = Multivariate Logistic regression: Cross Entropy Loss
- Teacher = Multitask loss
- Teacher = Triplet Loss

Loss Function

A Multi-Task Teacher for two outputs: Classification and Regression

- Multi-Task Loss - “Rich Feature Hierarchies for Object Detection” (6/14)
- L = Cross Entropy Loss + /alpha * (MSE or MAE Loss)
- Choose an appropriate /alpha weight to balance the losses due to regression and cross entropy

Loss Function

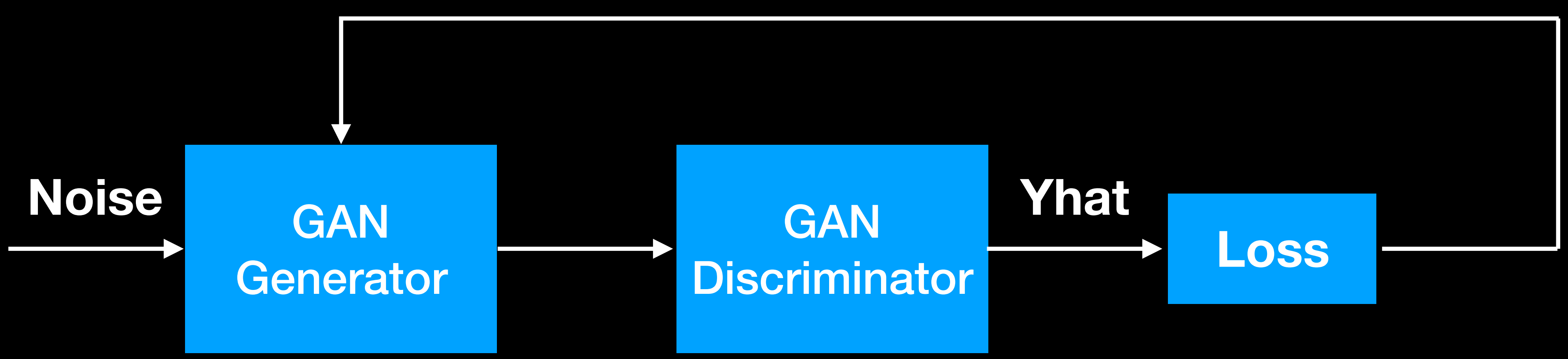
<u>A Specialized Teacher for a Siamese Architecture</u>

- Triplet Loss (Siamese network) - “FaceNet” (3/15)
- L(anchorImage, positiveImage, negativeImage) = MAX(distance(anchorImage, positiveImage) - distance(anchorImage, negativeImage) + /alpha, 0)
- Choose hard triplets pairs (anchorImage, positiveImage, negativeImage) to train on.

Loss Function

<u>How good is your teacher = How are you trying to learn?</u>

The Generative Adversarial Network (GAN) discriminator outputs a loss that is used to train the parameters of a neural net associated with a GAN generator.



Loss Function

- **Enhancements to the Teacher**
- Bayesian statistics: Prior knowledge is utilized
- WeightRegularization: L1 (sparsity of weights is encouraged) and L2 (large weights are discouraged)

Moral

- The loss function tells your algorithm how to improve its results.
- A poorly constructed loss function creates a weak signal for improvement.
- A good loss function creates a strong signal that helps your algorithm learn the required representations.
- Think DEEPLY AND CAREFULLY about this feedback signal.

References

- The MNIST DATABASE of Handwritten Images, LeCun, et. al.
- ImageNet: A Large Scale Hierarchical Image Database, Deng, et. al., CVPR 2009
- ObjectNet: A Large Scale Bias-Controlled Dataset for Pushing the Limits of Object Recognition Models, Barbu, et. al., NeurIPS 2019
- NITRE 2020 Challenge on Real World Image Super Resolution, Lugamayr, et. al., arXiv May 2020
- Data Augmentation using GANS, Tanaka, et. al., arXiv April 2019
- Self Training With Noisy Student Improves ImageNet Classification, Xie, et. al., arXiv November 2019
- Rich Feature Hierarchies For Accurate Object Detection and Semantic Segmentation, Girshick, et. al., CVPR 2014
- FaceNet: A Unified Embedding for Face Recognition and Clustering, Schroff, et. al., arXiv March 2015
- Focal Loss for Dense Object Detection, He, et. al., arXiv August 2017