Applied Machine Learning: Challenges and Considerations

KEELEY EDWARDS
BISON PROGRAMMING TEAM – MARCH 11, 2022

Overview

- Machine learning overview
- CIFAR Demo
- My research: challenges and considerations

Machine learning overview

- Key concepts:
  - Network architecture
  - Training set
  - Loss function

Network architectures

- Image-to-image: Convolutional Neural Networks
  - Image enhancement
  - Image classification
- Data-to-x:
  - Data-to-parameter is accomplished with fully connected networks (in our research)
  - Data-to-image: hybrid networks

Training set

- **Supervised** machine learning requires a large amount of labelled data
- Training data must be reasonably representative of whatever you want to use the neural network for
- Should be diverse so that the network can generalize

- Classic example: if you train an image classifier to identify birds on only blue jays, what happens if it sees an ostrich?

Loss function

- At the end of the day, training a neural network is just solving an optimization problem
- An alternative name for this is the "objective function"
- The loss function is a **choice**, and is defined by the network designer
  - During training, the network updates the weights in an attempt to minimize the value of the loss function

$$L(\underline{\mathbf{p}}) = \frac{(\underline{\mathbf{p}} - \underline{\mathbf{p}}_{true})^T(\underline{\mathbf{p}} - \underline{\mathbf{p}}_{true})}{\underline{\mathbf{p}}_{true}{}^T\underline{\mathbf{p}}_{true}}$$

Availability of training data

- Do we have labelled experimental data?
- Can we model the system synthetically?
  - How accurate is this model?
  - Is it resistant to change?

Cost of building a training set

- More general network requires more diversity in the training set:
  - This could mean hours of labeling images by hand
  - This could mean tens of thousands of synthetic data generation
- Synthetic training sets:
  - What do we have to model, and what can we ignore?

Quick demo:
CIFAR-10 using TensorFlow / Python

https://www.cs.toronto.edu/~kriz/cifar.html

CIFAR-10 confusion matrices

Fully Connected Network

	Airplane	Auto	Bird	Cat	Deer	Dog	Frog	Horse	Ship	Truck
Airplane	592	33	66	27	39	11	26	13	146	47
Auto	39	616	23	28	15	10	15	20	96	138
Bird	61	18	445	113	104	74	80	62	30	13
Cat	29	11	92	423	54	174	106	55	27	29
Deer	46	8	143	112	440	56	78	74	29	14
Dog	24	5	92	274	66	370	57	72	26	14
Frog	18	16	85	127	88	55	563	16	15	17
Horse	38	15	71	84	63	59	28	585	17	40
Ship	86	63	27	26	11	12	8	10	708	49
Truck	59	128	25	54	18	21	28	32	82	553

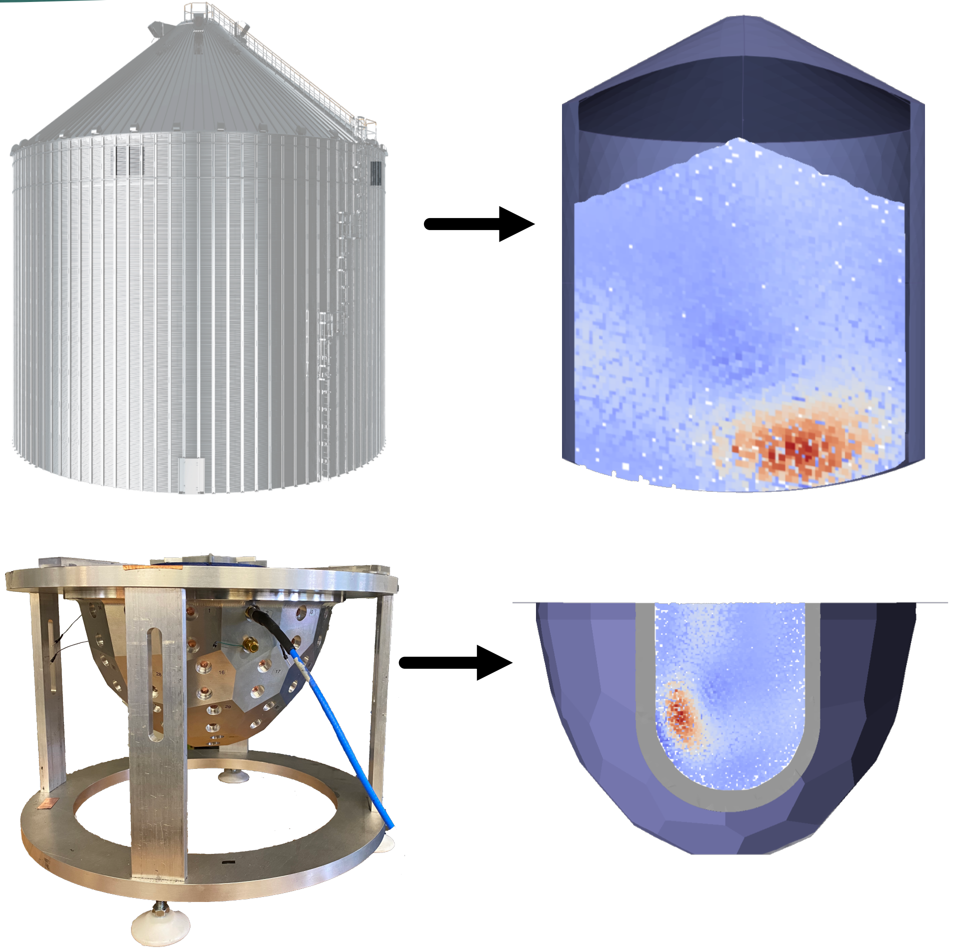
Convolutional Network

	Airplane	Auto	Bird	Cat	Deer	Dog	Frog	Horse	Ship	Truck
Airplane	687	11	79	27	29	2	8	21	84	52
Auto	29	702	12	13	5	5	19	6	47	162
Bird	76	5	527	71	114	64	56	56	18	13
Cat	24	13	81	453	83	180	62	72	14	18
Deer	19	1	109	52	605	44	48	97	17	8
Dog	12	3	70	204	68	497	32	100	3	11
Frog	4	11	63	91	87	43	668	9	8	16
Horse	19	0	43	35	74	63	8	723	6	29
Ship	69	37	21	31	13	4	7	5	772	41
Truck	34	114	13	25	12	5	12	38	37	710

My research:
challenges and considerations

Electromagnetic imaging

- Use electromagnetic imaging to recover information about the inside of an object or container non-invasively
- Existing methods for reconstructing an image are computationally expensive
- **Apply machine learning to enable rapid processing of data from these systems to produce images**



Synthetic vs. experimental data

- Synthetic:
  - Ideal, clean data
  - Empirically calculated

  - Modelling error
  - Simplifications / assumptions made

- Experimental:
  - Prone to noise
  - System drift

  - Measurement error
  - Requires calibration

Calibration

- In an ideal, cooperative system, calibration involves recording measurements of a known target in within the system:

$$H^{unknown} = \frac{H^{cal}}{S^{cal}} S^{unknown}$$

- So what happens when we can't calibrate?