The Truth Behind Machine Learning and AI

Includes a couple of extra slides in response to the questions at the meeting

Avi Kak

kak@purdue.edu

November 5, 2019

Presented at the "Advanced Development for Security Applications Workshop (ADSA)", Boston, MA

Robot Vision Lab,

Purdue University

First Things First: What Are the Take-Aways for DHS

- **For Non-Mission-Critical Applications And When Decisions Must be Based on Large Datasets:** It would be foolish to not use the tools based on deep-learning.

- **For Mission-Critical Applications:** It is still too early to jump into the deep-learning bandwagon. We do not yet fully understand all of the “failure modes” of such tools.

- **The goal of this presentation is to justify these statements.**

Unquestionably, Modern ML and AI (aka Deep Learning) Have Brought Us Incredible Tools

- **ResNet:** One of the best deep networks for classifying images
- **Variants of R-CNN and SSD:** For detecting and localizing objects in images, these are the best
- **Cycle-GAN and Conditional-GAN:** With truly amazing abilities to carry out domain adaptation and domain repair.
- **Recurrent Networks:** Ideal for what is known as sequence learning
- **And so on ….**

Given All the Media Attention These Tools Have Received, We Need Answers to the Following Questions

- **The Media Attention:** Is there anything to be said about all this hype and the hyperbole?
- **The Robustness Issue:** Are these tools ready for mission critical applications?
- **The Fragility Issue:** Can the tools be fooled into giving wrong answers?
- **Understandability Issue:** When the answers produced defy credibility, can we tell why?

The Media Attention: The Hype and The Hyperbole

- **If you are under 35:** It must seem that ML and AI will rule every facet of our lives going forward.

- **If you are over 45:** Your reaction is likely to be: "I have seen this before. This phase shall pass too."

mid 1980s (when AI was super hot)	mid 2010's (when AI became hot again)
• Planning • LISP, symbolic reasoning • Robotics • Expert Systems	• Deep nets for classification • Deep nets for detection • Nets for machine translation • Nets for reinforcement learning

Mid 1980’s vs. Now

- **Just as was the case in 1980’s**, the current excitement is based primarily on the potential of the tools that have been developed.
  - There is no question that the deep-learning tools that have been developed are “**objects of great beauty**”
  - But that was also the case of the tools that were developed by the AI community in the 1980s
- **Just as was the case in 1980’s,** the current excitement in the deep learning tools does not factor in the fact that we do not yet fully understand all of their limitations.

Let's Now Talk About the Robustness Issues: Training Deep Networks

- **The most impressive demos of deep learning** are based on collecting humongous training datasets by scraping the internet and getting volunteers to label the objects in the images of the dataset.

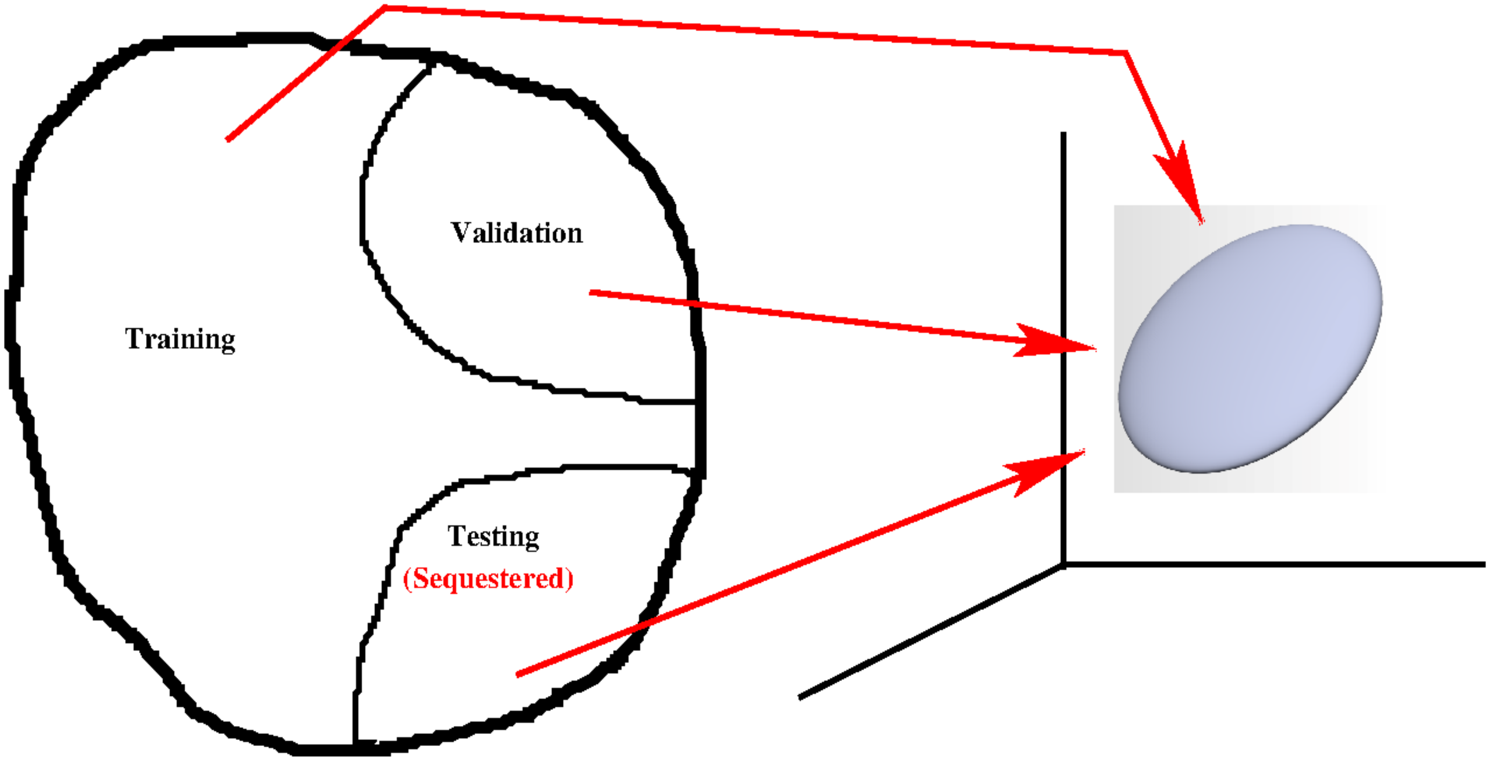
- **A ground-truthed dataset thus created is randomized and divided into three parts:**

  - **one part for training**

  - **one part for validation**

  - **and one part for testing (this part is sequestered)**

Training Deep Networks (contd.)



Ground–Truthed Data

Randomized

Covariance Matrix of the Data

In-Distro vs. Out-of-Distro Testing of Pre-Trained Networks

- Let's consider **ResNet** - this is one of the world's most famous deep networks for solving image classification problems. In addition, I'll also consider **Inception** and **AlexNet**.

- And let's consider ImageNet --- this is the world's most famous image dataset for benchmarking convolutional networks.

- In-Distro means the images that can be expected to be similar to those in ImageNet. And Out-of-Distro means the opposite.

How Did I Choose the Images for the In-Distro vs. Out-of-Distro Test

- My wife and I are avid cyclists. So the first thing that popped up in my mind were bicycle images. [ImageNet includes the bicycle category and its various subcategories.]
- From the web, I downloaded 6 images that show bicycles as you would see them in the streets. These were my In-Distro images.
- By using search strings like “wall stored bicycles”, “bicycles in repair shops”, etc., I also downloaded what I considered to be 6 Out-of-Distro images.

About the Results Shown in the Next Two Slides

- The next slide shows the classification results on what I believe are **In-Distro** images.
- It is surprising to see the errors for the In-Distro images, but the errors for **Out-of-Distro** images on the second slide are much more frequent.
- ResNet used for these results is ResNet-18 and, when the ResNet was trained, ImageNet had 1 million images with 1000 categories.

Results for What I Believe are In-Distro Images

ResNet	bbf2	mo bike	bbf2	bbf2	unicycle	bbf2
Inception	tricycle	mo bike	bbf2	bbf2	unicycle	bbf2
AlexNet	tricycle	mo bike	tricycle	bbf2	unicycle	lionfish

Blue label: Wrong answer

mo bike : mountain bike
bbf2 : bicycle built for two

Results for What I Believe are Out-of-Distro Images






ResNet	whistle	unicycle	turnstile	tricycle	unicycle	tricycle
Inception	mag cmp	unicycle	bbf2	moped	mo bike	mo bike
AlexNet	bow	bbf2	bow	tricycle	jinrikisha	stetho

Blue label: Wrong answer

stetho	:	stethoscope
mo bike	:	mountain bike
mag cmp	:	magnetic compass
bbf2	:	bicycle built for two

But What About the Fact That It is Possible to Adapt Deep Networks to New Data?

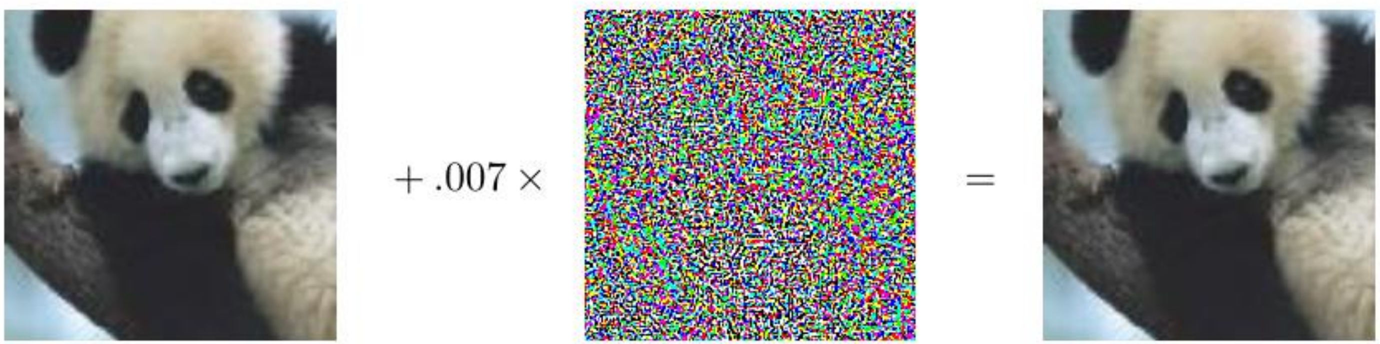
- **Yes,** deep learning does provide us with Transfer Learning techniques, GANs, etc., for domain adaptation.

- **But I am NOT talking about domain adaptation.**

- **I am talking about a clever adversary recognizing the fundamental limitations of your deep-learning based approach and creating a one-off example of a deadly threat.**

What About the Fragility Issues?

- **Can a deep network be fooled into giving a wrong answer? The answer is: YES**



$\boldsymbol{x}$ + .007 × $\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$ = $\boldsymbol{x} + \epsilon\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

"panda" 57.7% confidence | "nematode" 8.2% confidence | "gibbon" 99.3 % confidence

Gibbon

[From: Goodfellow, Shlens, & Szegedy ICLR 2015]

What About the Understandability Issue?

- When a deep-learning based tool makes an error, can the human users understand the reason for that error. **The answer is: NO. Deep networks operate like “black boxes”.**

- **On the other hand, tools based on traditional machine learning can explain their decisions.**

- My own open-source Decision Tree module for classification can show you which specific training samples influenced a classification decision.

Introspection Ability of the Python Module DecisionTree-3.4.3

When you invoke the Introspection API of my module after you have trained a decision tree, it shows which training samples (sample_1, sample_2, etc) contribute directly or indirectly to each node.

```
sample_1:
   nodes affected directly: [2, 5, 19, 23]
   nodes affected through probabilistic generalization:
        2=> [3, 4, 25]
            25=> [26]
        5=> [6]
            6=> [7, 13]
                7=> [8, 11]
                    8=> [9, 10]
                    11=> [12]
                13=> [14, 18]
                    14=> [15, 16]
                        16=> [17]
        19=> [20]
            20=> [21, 22]
        23=> [24]

sample_4:
   nodes affected directly: [2, 5, 6, 7, 11]
   nodes affected through probabilistic generalization:
        2=> [3, 4, 25]
            25=> [26]
        5=> [19]
            19=> [20, 23]
                20=> [21, 22]
                23=> [24]
        6=> [13]
            13=> [14, 18]
                14=> [15, 16]
                    16=> [17]
        7=> [8]
            8=> [9, 10]
        11=> [12]

...
...
...
```

Deep networks can not provide such functionality.

What is the Way Forward for Mission Critical Applications?

- **Let's consider the problem of threat detection for airport baggage inspection systems.**

- Since the ultimate truth is always in the material composition (as measured by, say, Z_{eff} and ρ) of the contents of a bag, **an approach that gives greater importance to the underlying physics is likely to be more robust than a purely data-driven approach based on deep learning.**

Mission Critical Applications (contd.)

- If a threat detector could be initialized with physics based considerations and then further fine-tuned with a deep-learning framework, **that might yield the best of both worlds.**

- **But what about the training data needs of whatever part is based on deep learning?**

- Fortunately, the baggage simulators being developed in the research labs have now become so powerful that **generating the training data is not a challenge any longer.**

DEBISim --- A Baggage Simulator from Purdue RVL

- I believe that this tool will play an important role in figuring out how to best combine the power of the (Z_{eff}, ρ) based approach and the DL based approaches to threat detection.

- Regarding the precision of the simulations: The 3D DECT reconstructions of the Battelle phantom as produced by DEBISim are virtually identical to those produced by the IDSS 1000 scanner.

- DEBISim also includes a powerful GUI for packing a virtual bag with objects composed of different materials (including threat materials like RDX, H2O2, etc.)

THANK YOU