Artificial Intelligence, Machine Learning, and Deep Learning

Courses We Currently Offer

UMSL
Mathematics & Computer Science

CMP SCI 4300 Introduction to Artificial Intelligence: 3 semester hours
Prerequisites: CMP SCI 3130. This course provides an introduction to artificial intelligence. The list of topics may include search, planning, knowledge-based reasoning, probabilistic inference, machine learning, natural language processing, and practical applications. Credit cannot be granted for both CMP SCI 4300 and CMP SCI 5300.

CMP SCI 4340 Introduction to Machine Learning: 3 semester hours
Prerequisites: CMP SCI 2261 and CMP SCI 3130. This course provides an introduction to machine learning in the context of applications such as data mining, natural language processing, and adaptive computer systems. The course reviews several supervised, unsupervised, and reinforcement machine learning techniques such as naive Bayes networks, clustering, and decision trees. Selected concepts in computational learning theory may also be covered. Credit cannot be granted for both CMP SCI 4340 and CMP SCI 5340.

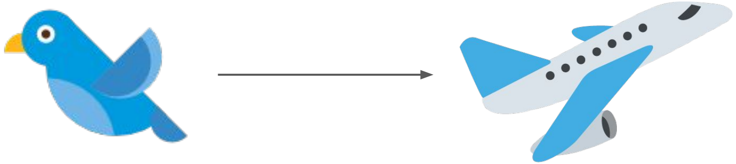
CMP SCI 4390 Introduction to Deep Learning: 3 semester hours
Prerequisites: CMP SCI 3130 or consent of instructor. This course introduces mathematical foundations for deep learning, and follows with practical applications using selected domains such as image classification or protein predictions. It also covers dense neural networks, convolutional neural networks, recurrent neural networks, and other state-of-the-art networks. Credit cannot be granted for both CMP SCI 4390 and CMP SCI 5390.

Topics

- Differences between AI, ML, and DL
- Some recent methods in DL
- Goals of DL & Limitations
- Notes on Learning DL

Definition of Artificial Intelligence (AI) - term coined in 1955 by John McCarthy

- AI is creation of intelligent machines that work and react like humans
    - Learning from the nature is great!



- But, we designed wheels. Not one animal rolls around upon a rotating body part: a biological wheel.

- AI is creation of agents that act rationally
    - Performance can be measured using a well defined “performance measure”

Definition of Machine Learning (ML) - term coined in 1969 by Arthur Samuel

- ML is learning from examples “without being explicitly programmed”

- (some) ML Algorithms / Methods

    - Dimensionality reduction - e.g. Principal Component Analysis
    - Ensemble learning - e.g. Boosting using AdaBoost
    - Linear Classifiers - e.g. Logistic Regression
    - Supervised learning - e.g. Neural Networks and Support Vector Machines
    - Decision tree algorithms - e.g. Random Forest
    - Clustering - e.g. k-means algorithm
    - Unsupervised learning - e.g. Expectation-maximization algorithm
    - Deep Learning methods - e.g. Residual Networks
    - Bayesian networks - e.g. Naive Bayes classifier

XGBoost: A Highly Effective and Widely Used ML Method

- XGBoost is an **open-source library**
- Is a **gradient boosting framework** for C++, Java, Python,R, and Julia
  - a prediction model in the form of an ensemble of weak prediction models (usually decision trees)
- Is popular among the Kaggle community
  - used for a large number of competitions

kaggle.com

- Integrated with scikit-learn for Python, and with the caret package for R

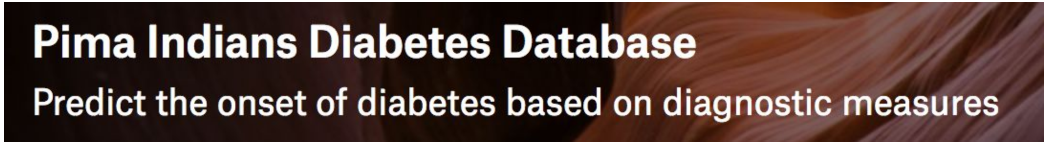
- Can be integrated into data flow frameworks like Apache Spark and Hadoop

hadoop

XGBoost: A Scalable Tree Boosting System

Tianqi Chen, Carlos Guestrin

XGBoost for Pima Indians Diabetes Database



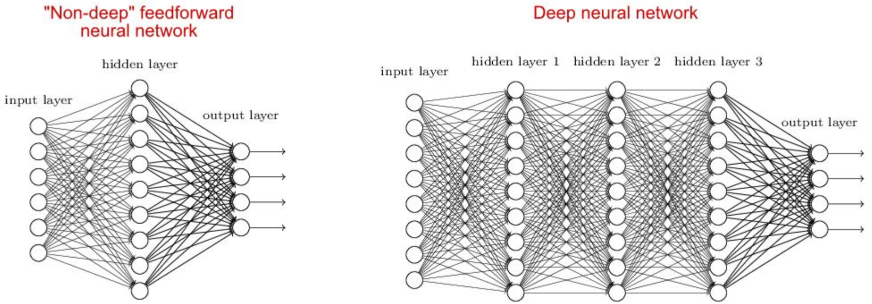
# Pregnancies	# Glucose	# BloodPressure	# SkinThickness	# Insulin	# BMI	# DiabetesPedigreeF	# Age	# Outcome
Number of times pregnant	Plasma glucose concentration a 2 hours in an oral glucose tolerance test	Diastolic blood pressure (mm Hg)	Triceps skin fold thickness (mm)	2-Hour serum insulin (mu U/ml)	Body mass index (weight in kg/(height in m)^2)	Diabetes pedigree function	Age (years)	Class variable (0 or 1)
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1

```
model = XGBClassifier()
model.fit(X_train, y_train)
```

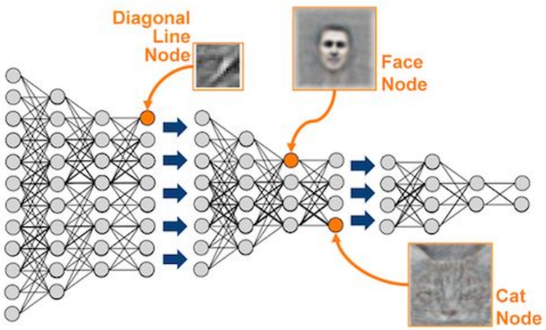
Deep Learning (DL) - term coined in 2000

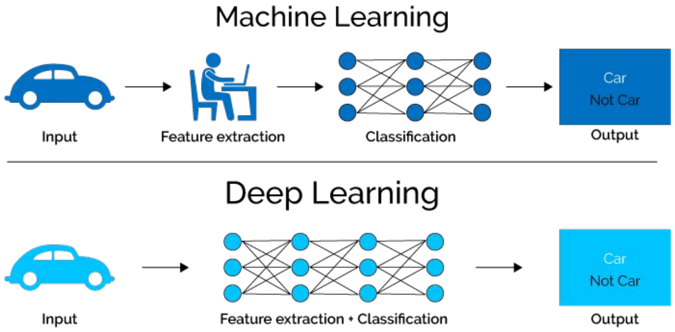
- DL is a subfield of ML

- DL is Large Neural Networks



- DL is Hierarchical Feature Learning





A Hidden Layer

x AND Y

x	y	f-and(x,y)
0	0	0
0	1	0
1	0	0
1	1	1

x OR y

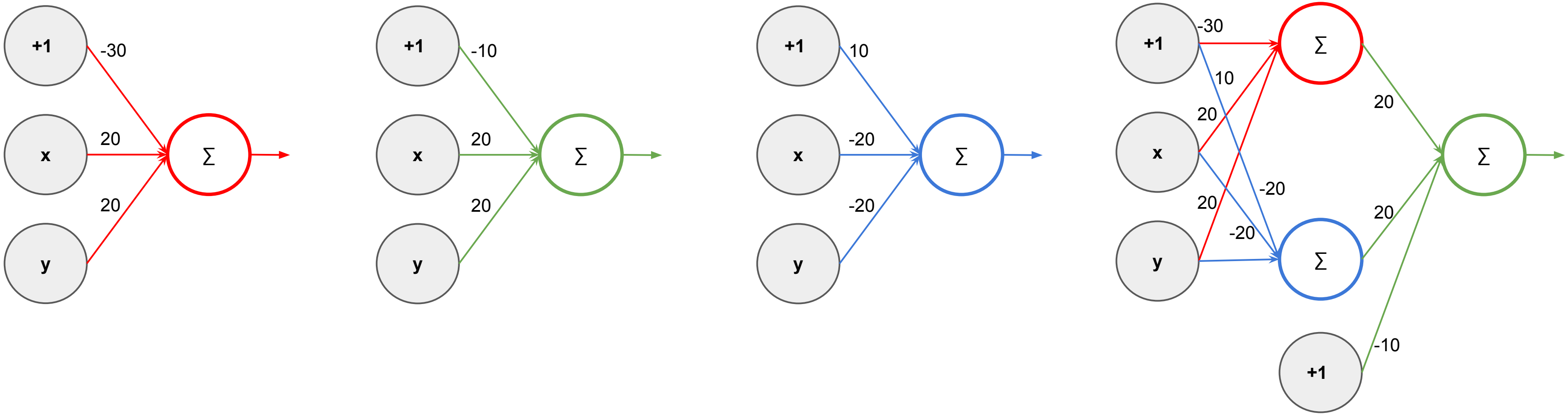
x	y	f-or(x,y)
0	0	0
0	1	1
1	0	1
1	1	1

(!x) AND (!y)

x	y	f-rev-and(x,y)
0	0	1
0	1	0
1	0	0
1	1	0

x XNOR y

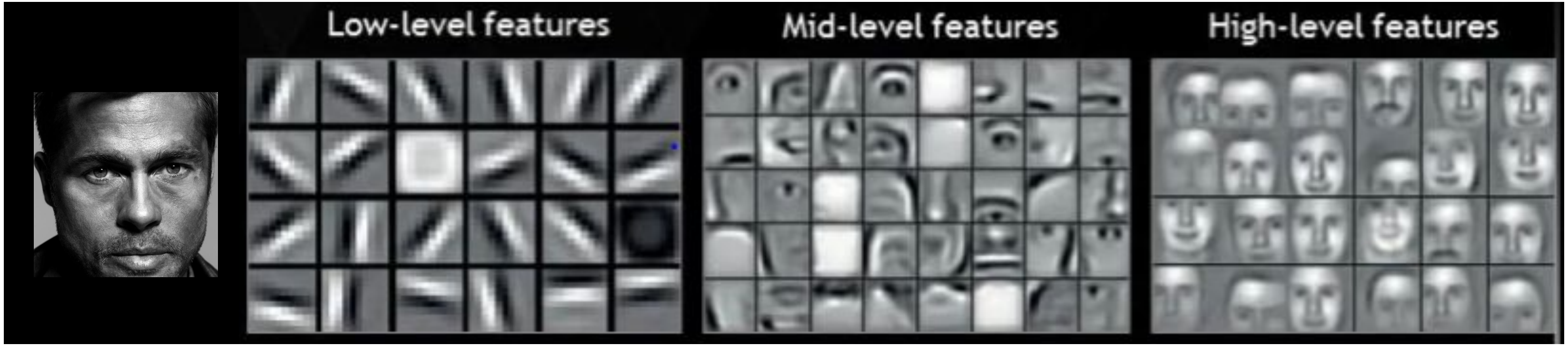
x	y	f-xnor(x,y)
0	0	1
0	1	0
1	0	0
1	1	1



XNOR = (a AND b) OR (!a AND !b)

Many Hidden Layers

- Universal approximation theorem
    - a feed-forward network with a single hidden layer containing a finite number of neurons can approximate (any) continuous functions
    - ability to represent does not mean ability to learn
- Too deep for too little data may not be effective
    - The accuracy on Pima Indian diabetes dataset will not increase significantly if you make the network very deep
- “Deep” is useful when features need to be learned

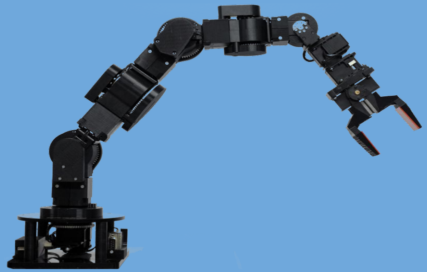


Is Deep Learning simply “more” Layers?

- Yes, it is (in a way).

- But the consequences are what make it a separate field of its own.
  - Zero or minimal feature engineering
  - More computing resources are needed
  - More data is needed
  - Predictions are more accurate, so many further possibilities open
  - Transfer learning is possible
  - We now need to study what the model is learning
  - Smart development practices are needed (wrong directions can be very costly)

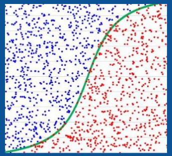
AI vs ML vs DL



ARTIFICIAL INTELLIGENCE

a very broad field

1950s



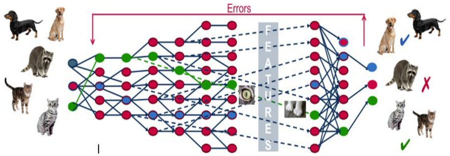
MACHINE LEARNING

Fundamentals of “learning from data”

1980s

Deep Learning

Powerful trending ML methods



2010s

All AI Approaches Have Value

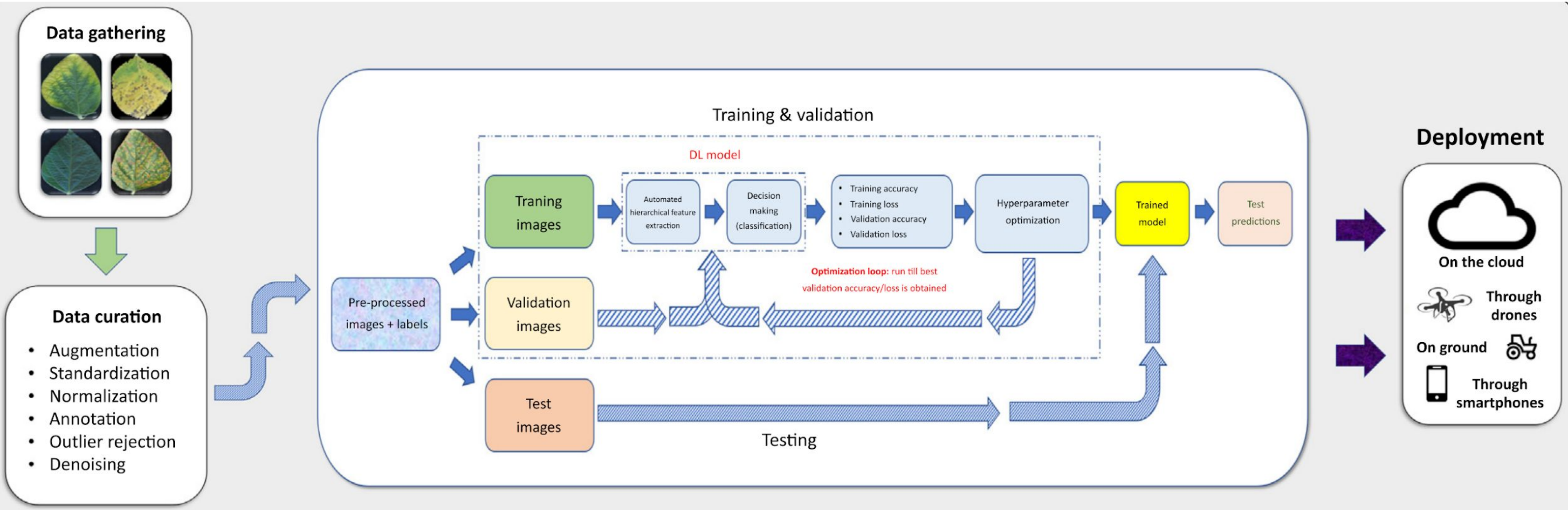
- Perceptrons and the dark age of connectionism
    - Neural network research came to a sudden halt with the publication of Minsky and Papert's 1969 book Perceptrons
    - The effect of the book was devastating
    - Virtually no research at all was done in connectionism for 10 years

- AlphaGo relies on two different components
    - a tree search procedure - (basic AI - symbolic AI)
    - convolutional networks (DL - geometric AI)

- This suggests, in future
    - we will come back to the topics discussed in AI
    - we will combine deep learning approaches with the primitive AI approaches like the ones related to logic reasoning.

Methods in Deep Learning

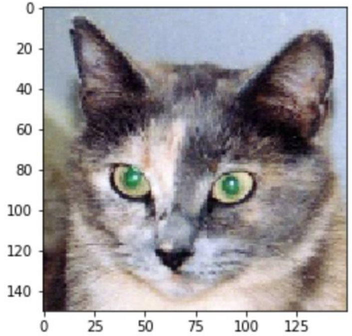
DL Tool Chain: From Gathering Data to Decision Making



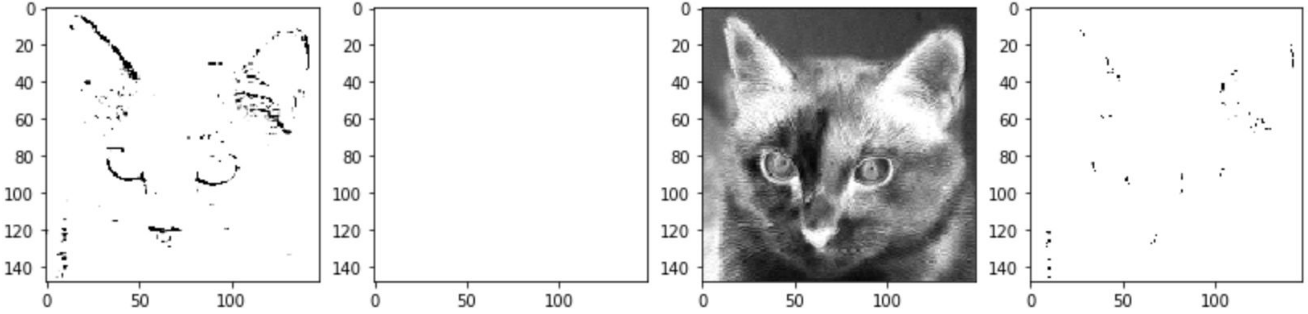
Deep Learning for Plant Stress Phenotyping: Trends and Future Perspectives

Asheesh Kumar Singh,[1] Baskar Ganapathysubramanian,[2] Soumik Sarkar,[2,*] and Arti Singh[1,*]

Deep Learning Models are not Black Boxes

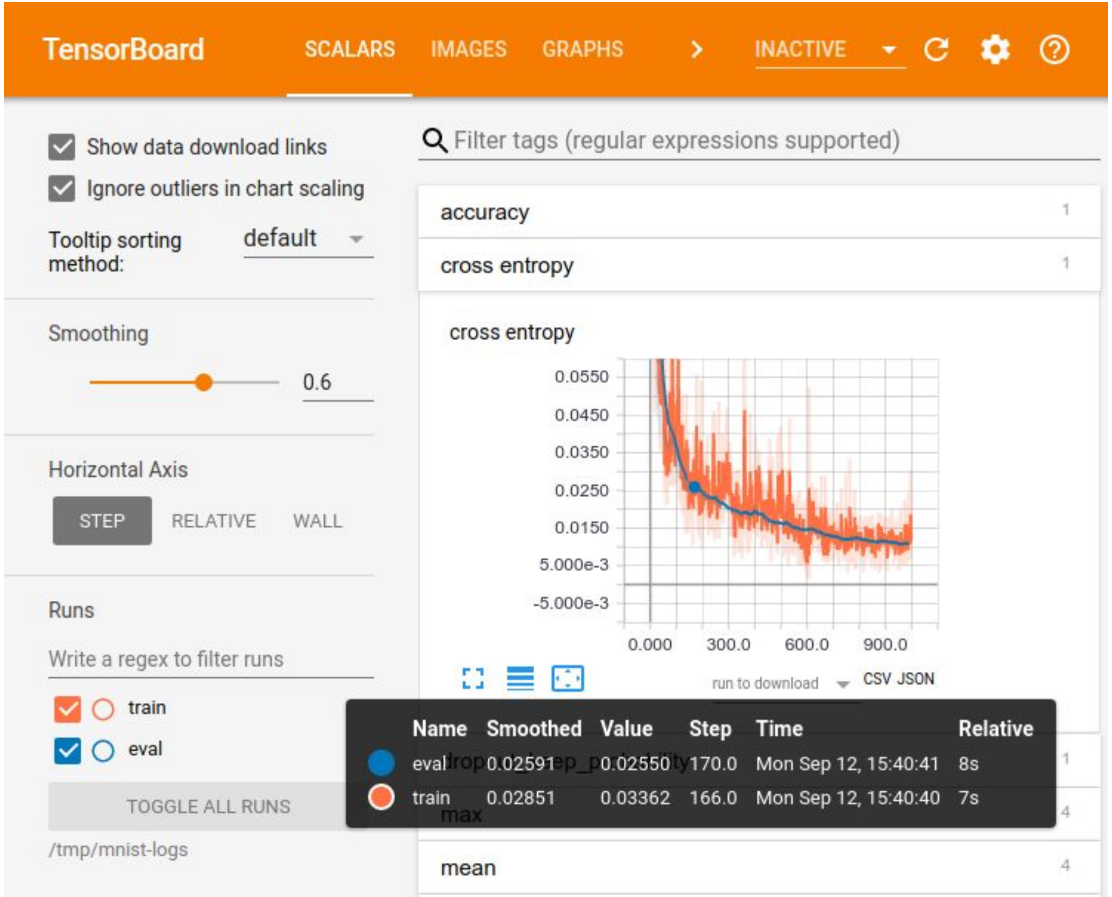


Input Image



Activations of the first few filters in the first layer

Deep Learning Models are not Black Boxes



Deep Learning Models are not Black Boxes

African elephants detected correctly by VGG16

Apply activation heatmap back to the original image

Convolutional Neural Networks and Transfer Learning

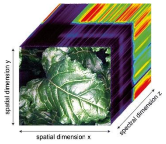
Current Practice:

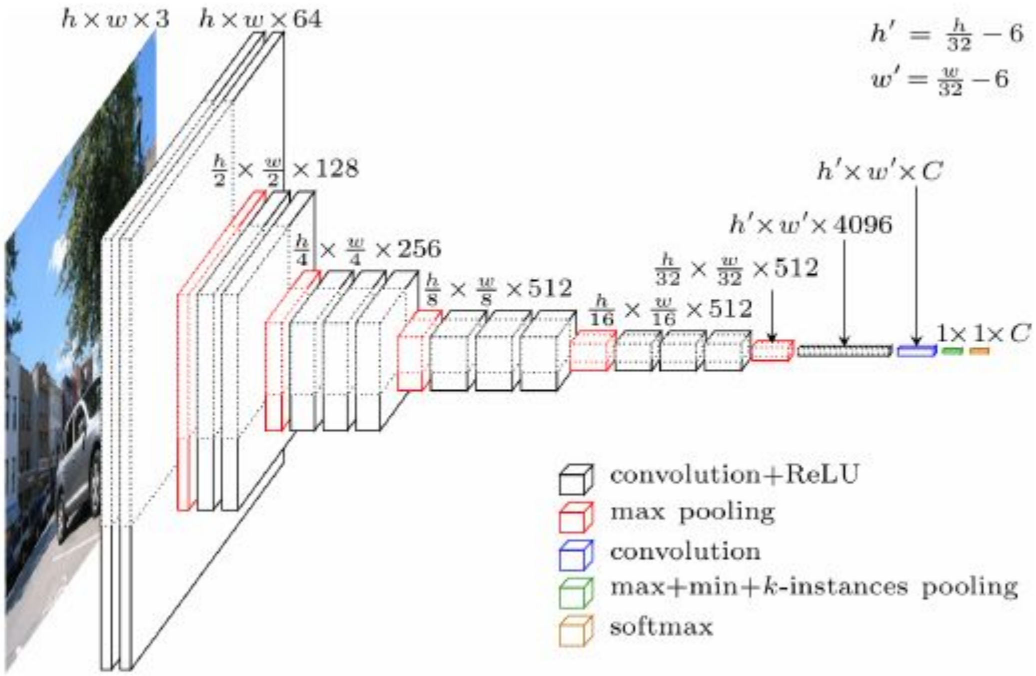
- Use pretrained models such as VGG16, Inception-v3 (by Google), etc.
- Most of them are independent of image size (the convolutional layers)

Example 1:
You want to build your own face recognizer to unlock your door

Example 2:
You want to build a facebook app that tells you if someone posted a happy or sad picture of you (tagging you).

Example 3:
Hyperspectral imaging at Donald Danforth Plant Science Center

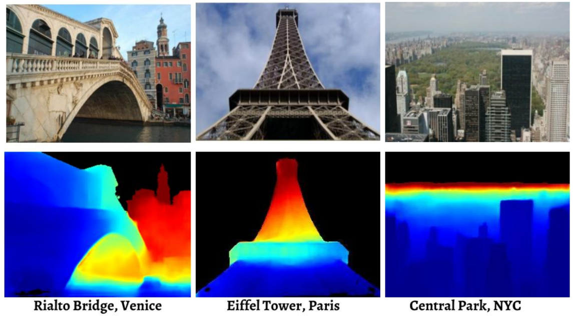


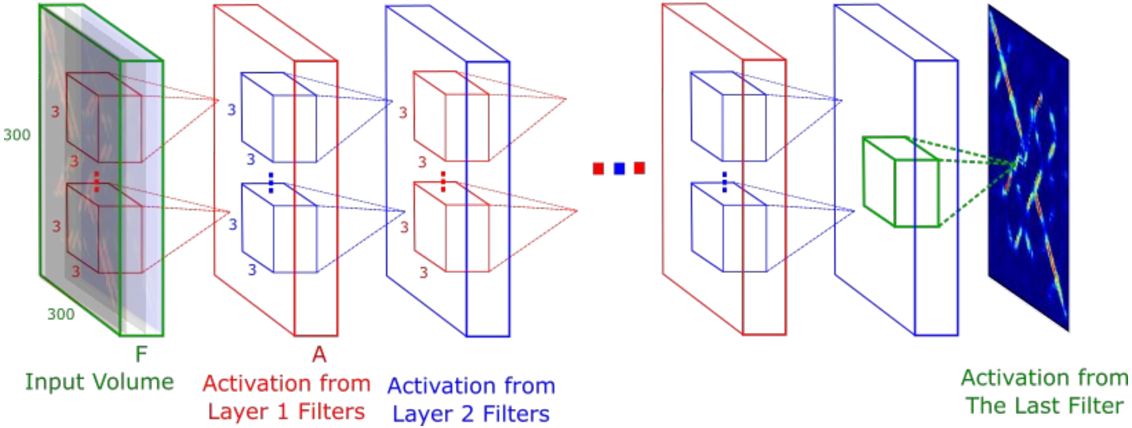


The VGG-16 Architecture

- A deep convolutional network for object recognition developed and trained by Oxford's renowned Visual Geometry Group (VGG)
- VGGNet performed very well in the Image Net Large Scale Visual Recognition Challenge (ILSVRC) in 2014

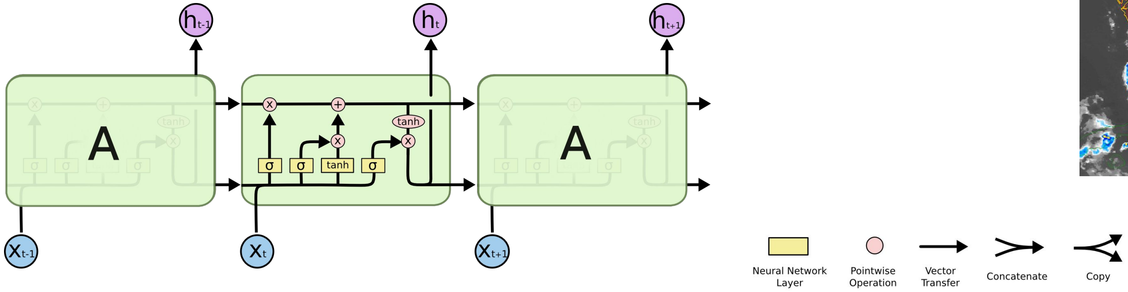
CNNs for Image Depth Prediction and Protein Contact Prediction

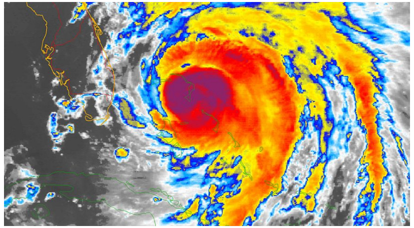




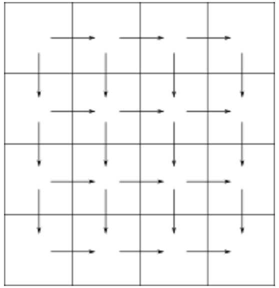
Long Short Term Memory networks (LSTMs)

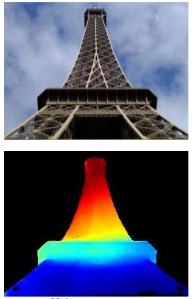
1D LSTM

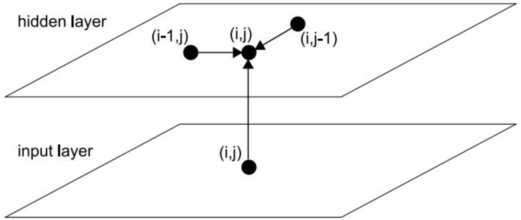




2D LSTM



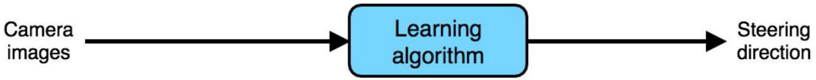

Eiffel Tower, Paris

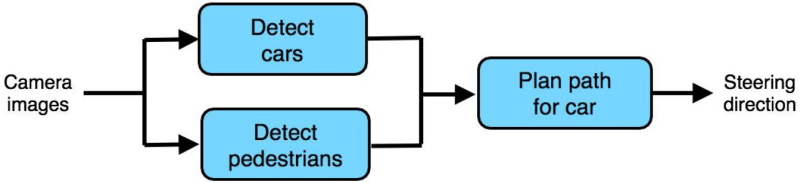


End-to-End Systems

- Less or no Feature engineering
  - Minimize hand-crafted features
- Examples
  - Object recognition
  - Neural Machine Translation: take raw words as the input and all components trained together
  - Neural Caption Generation: produce image descriptions from raw images

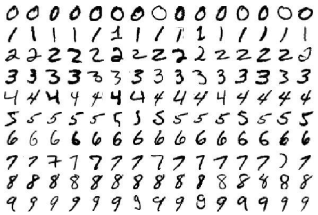
- When NOT to use end-to-end systems?
  - When we do not have a lot of hand engineered features end-to-end may not be the best solution
    - For example, a lot of hand engineered features text or audio data enable end-to-end
  - And, when we do not have enough data to train the model
    - We need a lot of [image, steering] data for end-to-end automatic driving
    - But we may still have enough data to train individual components



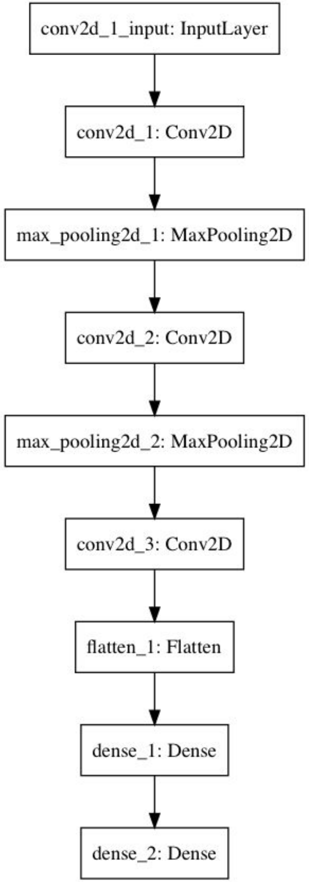


GPUs for Deep Learning

- The MNIST dataset of classifying images
  - contains 60,000 training images and 10,000 testing images



```python
with tf.device('/device:GPU:0'):
    model = models.Sequential()
    model.add(layers.Conv2D(32, (3, 3), activation='relu', input_shape=(28, 28, 1)))
    model.add(layers.MaxPooling2D((2, 2)))
    model.add(layers.Conv2D(64, (3, 3), activation='relu'))
    model.add(layers.MaxPooling2D((2, 2)))
    model.add(layers.Conv2D(64, (3, 3), activation='relu'))
    model.add(layers.Flatten())
    model.add(layers.Dense(64, activation='relu'))
    model.add(layers.Dense(10, activation='softmax'))
    model.compile(optimizer='rmsprop', loss='categorical_crossentropy', metrics=['accuracy'])
    model.fit(train_images, train_labels, epochs=8, batch_size=64)
```



Hardware Platforms

- cuDNN library
  - NVIDIA CUDA Deep Neural Network library (cuDNN)
  - For high-performance GPU acceleration
  - Older GPUs don't support cuDNN

- NVIDIA GPUs
  - Titan X (cuda cores = 3072, mem = 12 GB)
  - Quadro P2000 (cuda cores = 1024, mem = 5 GB)
  - Quadro P6000 (cuda cores = 5840, mem = 24 GB)
  - Tesla V100 (125 teraFLOPS, mem = 16/32 GB)

- MU RCSS Research Cluster
  - 15 GPU nodes (most of them with 12 GB memory)

- Setting up CUDA, cuDNN, and GPU Driver (version issues)
  - Python version and Virtual environments
  - Configuring multiple GPUs in one machine

Software Platforms

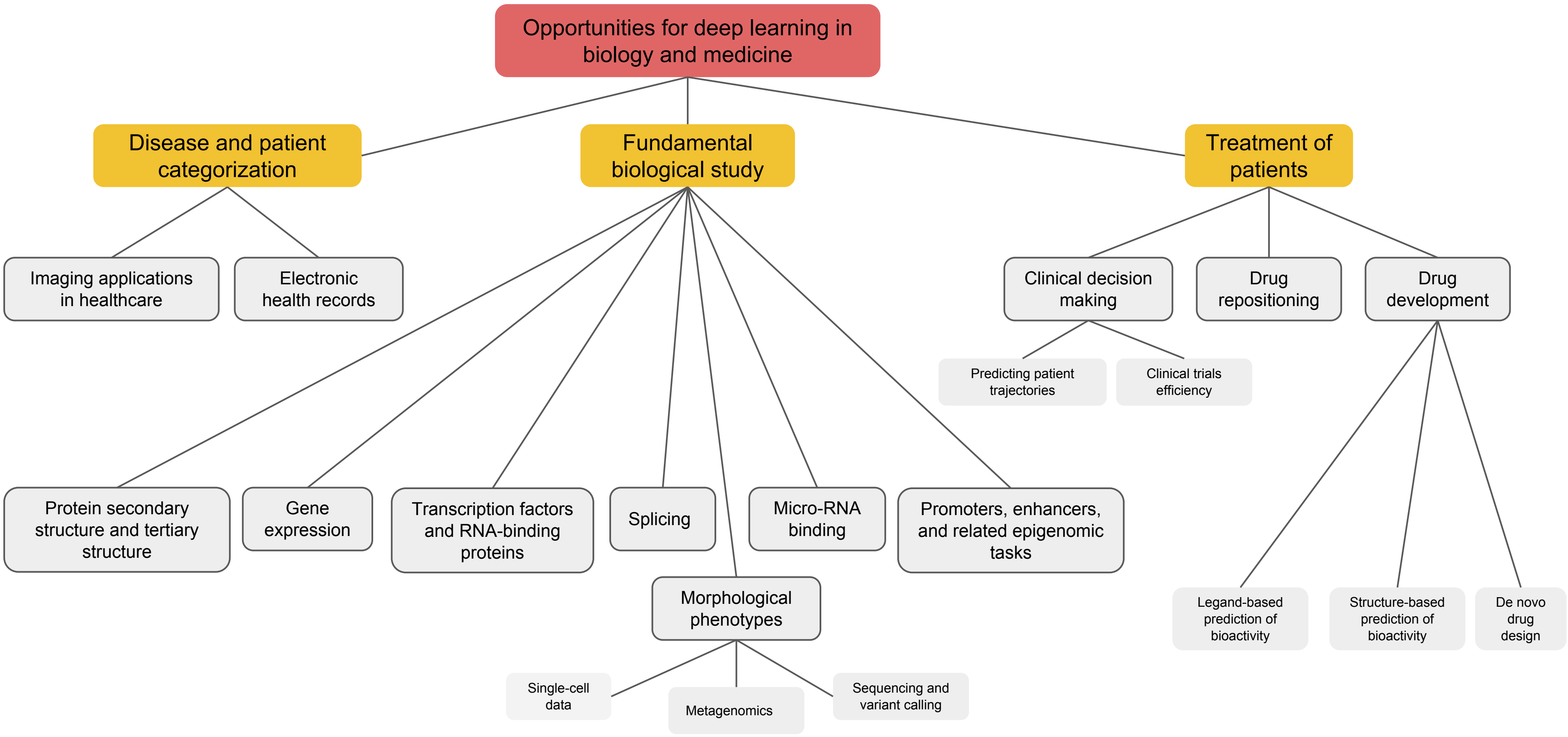
Some Current Goals of AI/DL

Replace Mental Labor

Physical Labor

Mental Labor

Solving Problems in Biology and Medicine



The state of Computer Vision and AI: we are really, really far away

http://karpathy.github.io/2012/10/22/state-of-computer-vision/

Some of the things "we" understand easily

There are 3 mirrors in the scene so some of those people are "fake" replicas from different viewpoints

Recognize Obama from the few pixels that make up his face

You recognize that there's a person standing on a scale, even though the scale occupies only very few white pixels that blend with the background

Obama has his foot positioned just slightly on top of the scale

Working physics - Obama is leaning in on the scale, which applies a force on it. Scale measures force that is applied on it, that's how it works => it will over-estimate the weight of the person standing on it.

The person measuring his weight is not aware of Obama doing this

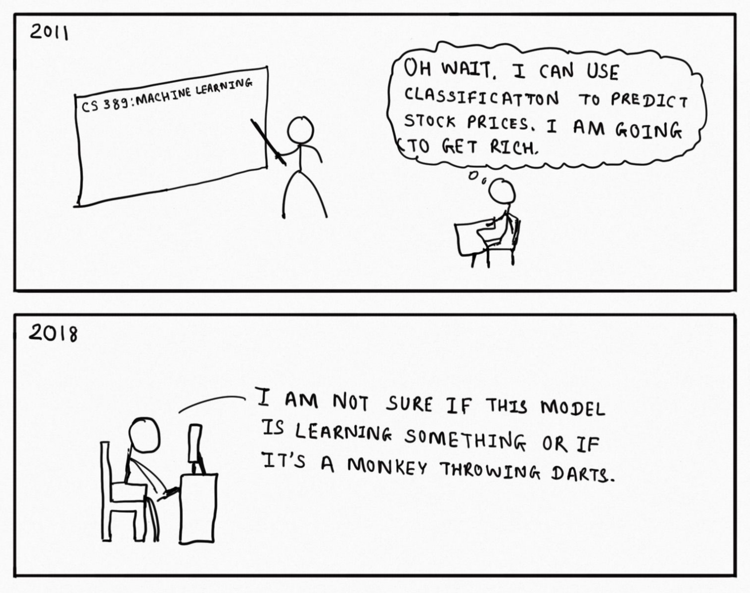
There are people in the back who find the person's imminent confusion funny

Limitations of DL

- Deep learning model is just a chain of simple continuous geometric transformations mapping one vector space into another
- All it can do is map one data manifold X into another manifold Y
    - assuming the existence of learnable continuous transform from X to Y
- A deep learning model can be interpreted as a kind of program; **but inversely most programs can't be expressed as deep learning models**
    - algorithm ≠ deep learning model
- For most tasks, either there exists no corresponding deep-neural network that solves the task or, even if one exists, it may not be learnable
    - The corresponding geometric transform may be far too complex, or there may not be appropriate data available to learn it
- Extreme generalization vs Local generalization
    - Extreme generalization: an ability to adapt to novel, never-before-experienced situations using little data or even no new data at all (abstraction and reasoning)
    - Local generalization: mapping from inputs to outputs

Learning the Bleeding-Edge “Deep Learning” because deep learning is already cutting-edge

Approach 1:
Learn basic mathematics in machine learning
then, learn basics of machine learning
then, learn basic concepts of deep learning
then, practice deep learning



Approach 2:
Pick a project
then, learn everything you need to learn to solve it

Approach 3:

Theory and application is tied together very closely - you could not just study the theory and go to your cave and solve a problem

Run others’ examples → Practice your own coding → Brush concepts → Read Literatures → Run others’ examples

Workshop on Deep Learning

- Rehearsal Workshop on 11-7-2018, Wednesday - 4:45 PM to 6:45 PM
- Actual Workshop on 11-14-2018, Wednesday - 4 PM to 6 PM

Thank You !!