

CS839: Special Topics in Deep Learning

Uni of Wisconsin Madison

Sharon Yixuan Li

Lecture 1

- Neural architecture design
- Trustworthy deep learning
- Interpretable deep learning
- Deep learning generalization and theory
- Learning with less supervision
- Lifelong learning
- Deep generative modeling

Evolution of NN architecture

LeNet

AlexNet

InceptionNet

DenseNet

~~DesNet~~ DenseNet \rightarrow AutoML \rightarrow NASNet [Zoph 2017]
goal is to replace human expert

Trustworthy DL

Out of distribution reliability

Closed-world \rightarrow train and test dist match

Open world \rightarrow train and test dist differ.

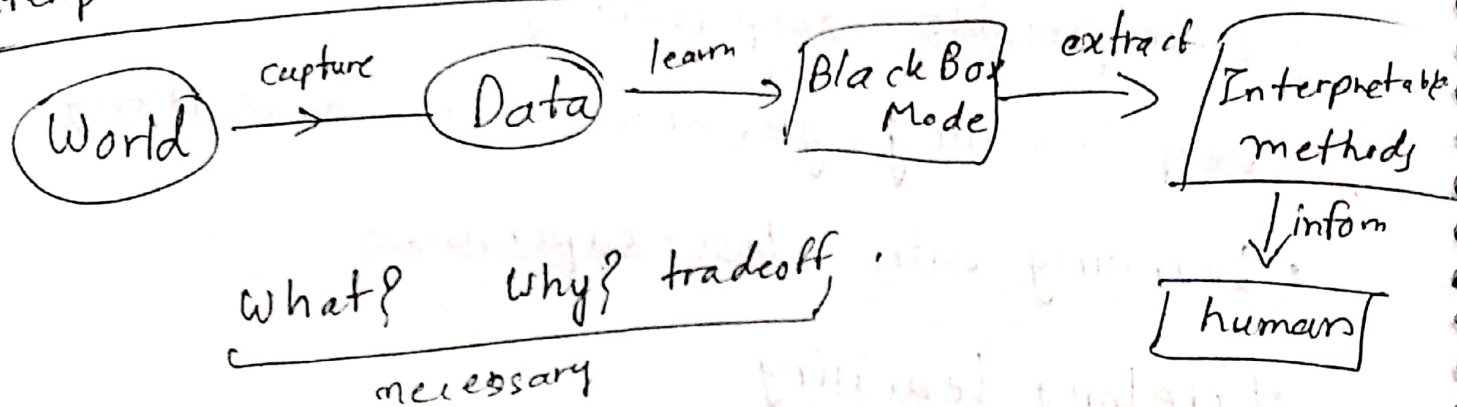
Trustworthy Deep learning

Out-of-distribution reliability

Adversarial Robustness

Fairness/Group Robustness

Interpretable Deep Learning



Generalization and Theory

Learning with less supervision

fully supervised \rightarrow weakly supervised \rightarrow Self supervised
(true notion of artificial intelligence)

Lifelong learning

Machines that ~~say~~ improve with experience
and become 'smarter' over time

Deep Generative Modeling

45 years of face generation style transfers

Lecture 2: Convolutional Neural Network

- Brief review of convolutional computing
 - 2D convolutions
 - Padding, stride
 - Multiple input and output channels

Basic convnets

- LeNet
- AlexNet
- ResNet
- DenseNet

Why Convolution?

- Translation invariance
- Locality

Kernel

edge detector
sharpen
gaussian blur

$X = n_h \times n_w$ input

$W = k_h \times k_w$ kernel

$b = \text{scalar bias}$

$Y = (n_h - k_h + 1) \times (n_w - k_w + 1)$ output

Padding and Striding

Multiple input channels:

Have a kernel for each channel and then sum results over channels

Why do we need multiple output channel?

→ To recognize particular different pattern
we don't define this particularly neural
nets learn this by itself.

MNIST dataset:

- centered and scaled
- 50,000 training data
- 10,000 test data
- 28x28 images
- 10 classes

ALEXNET [^{won} ImageNet 2012 competition]
[2009]

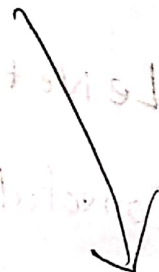
- Deeper and bigger LeNet
- Dropout
- Data Augmentation

ResNet [He et al. 2015]

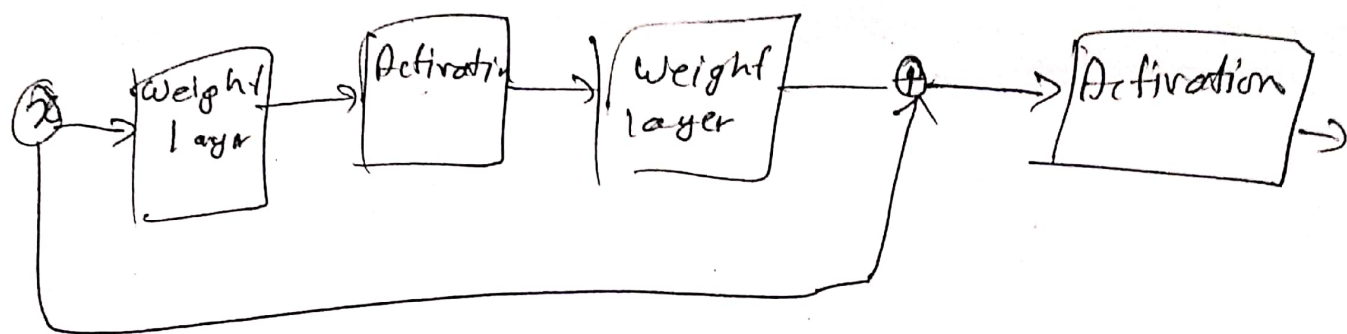
• Deeper model are harder to optimize

Copy the learned layer from the shallower
model and setting additional layers to
identity & mapping

Sigmoid



ReLU



DenseNet (Huang et al. 2016)

ResNet adds x and $f(x)$

DenseNet concatenates x and $f(x)$

Other architectures

ResNeXt [Xie 2016]

Wide ResNet [Zagoruyko 2016]

Deep Networks stochastic Depth [Huang et. al. 2016]

Fractal Net [Larsson et. al. 2017]

SqueezeNet [Iandola et. al. 2017]

ShuffleNet [Zhang et. al. 2018]

Auto Neural Architecture Search

Auto ML

NasNet [Zoph et. al. 2017]