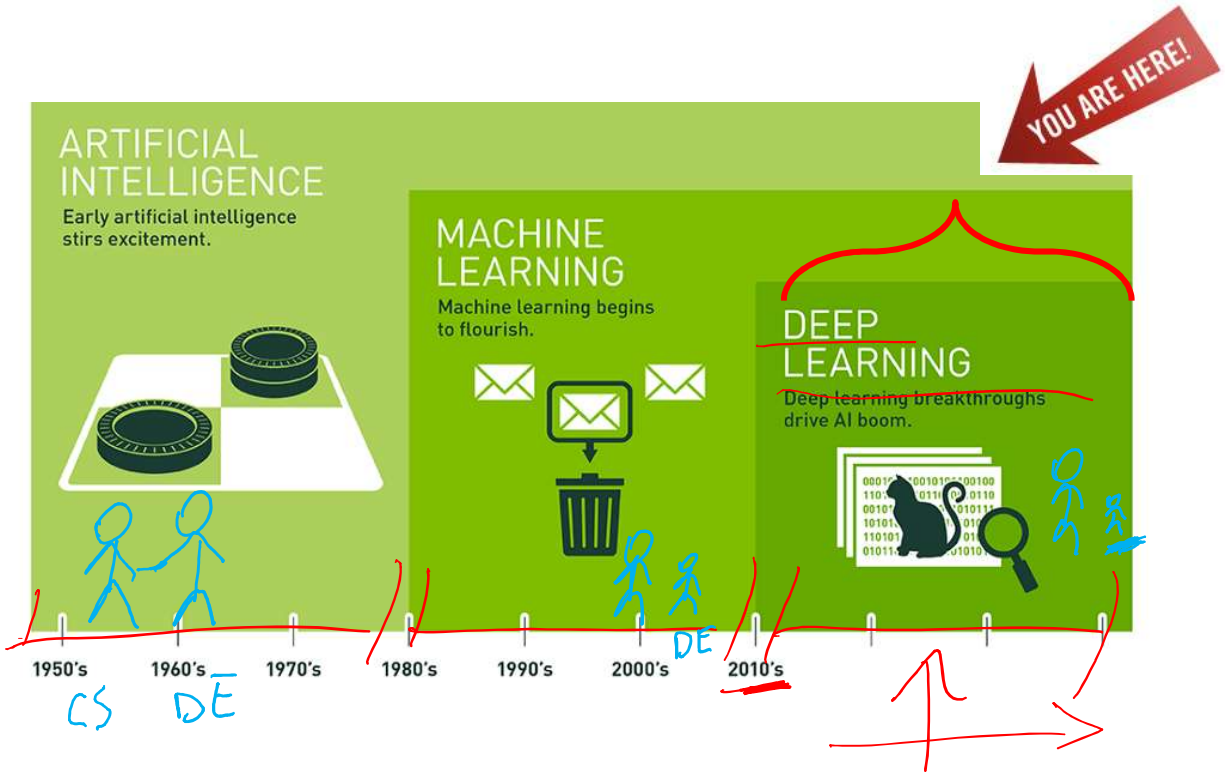
COMP 472: Artificial Intelligence
Deep Learning in 2 minutes

many slides from: Y. Bengio, A. Ng and Y. LeCun

Today

1. Introduction
2. Bag of word model
3. n-gram models
4. Deep Learning for NLP
   1. Word Embeddings
   2. Recurrent Neural Networks

History of AI



https://www.linkedin.com/pulse/goedels-incompleteness-theorem-emergence-ai-eberhard-schoeneburg/

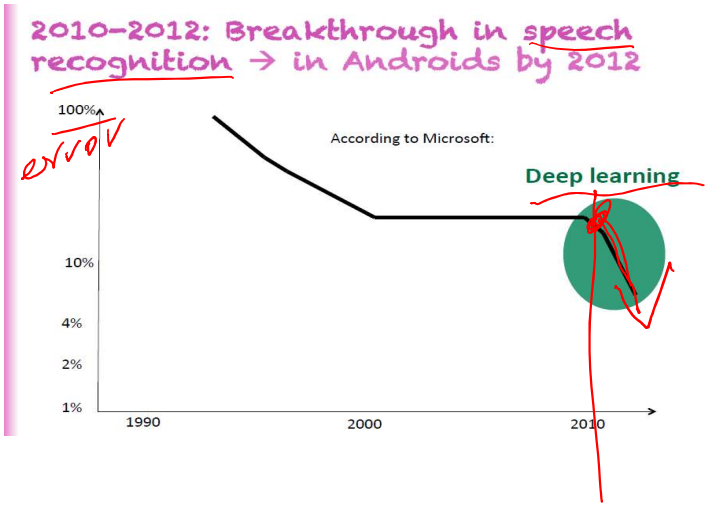
Major Breakthroughs

1. ~2010: Speech Recognition & Machine Translation

Skype Translator

Google now

Google Translate

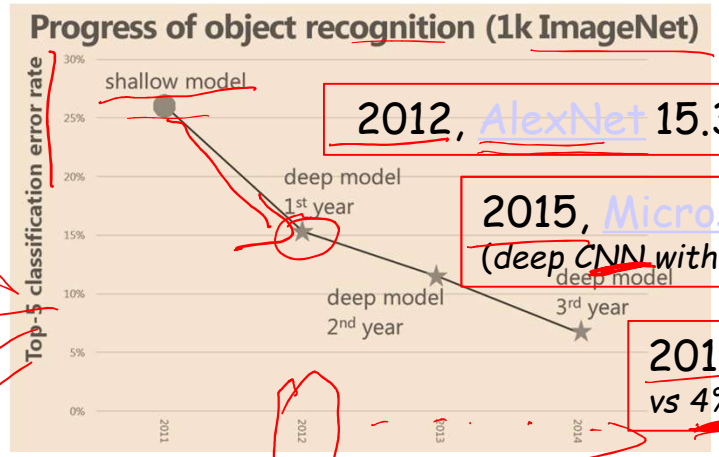


2. ~2012: Image Recognition & Computer Vision
3. ~2014: Natural Language Processing

Figure from Yoshua Bengio, 2015

Major Breakthroughs

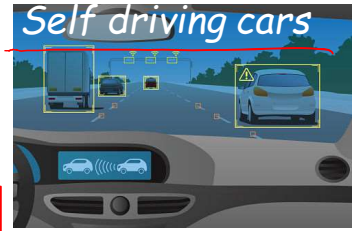
1. ~2010: Speech Recognition & Machine Translation
2. ~2012: Image Recognition & Computer Vision



2012, AlexNet 15.3%

2015, Microsoft Research, 3.58%
(deep CNN with over 100 layers)

2016, Google, 3.08%
vs 4% human error rate



3. ~2014: Natural Language Processing

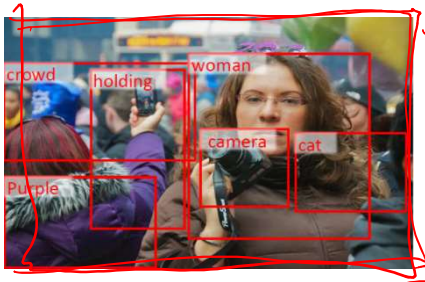
"hello how are you" word2vec

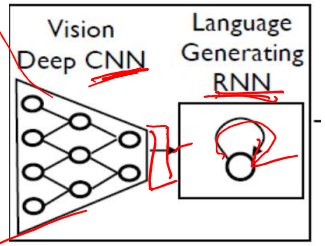
Figure from Yoshua Bengio, 2015

Major Breakthroughs

1. ~2010: Speech Recognition & Machine Translation
2. ~2012: Image Recognition & Computer Vision
3. ~2014: Natural Language Processing

Joe went to the kitchen. Fred went to the kitchen. Joe picked up the milk.
Joe travelled to the office. Joe left the milk. Joe went to the bathroom.
Where is the milk now? A: office
Where is Joe? A: bathroom
Where was Joe before the office? A: kitchen



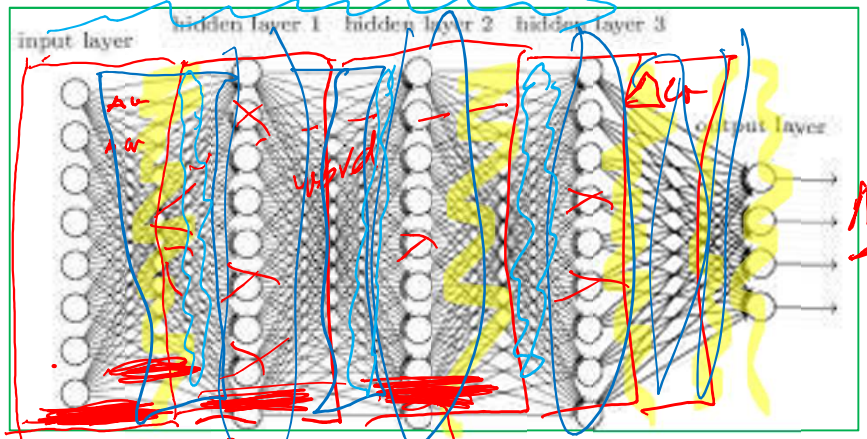


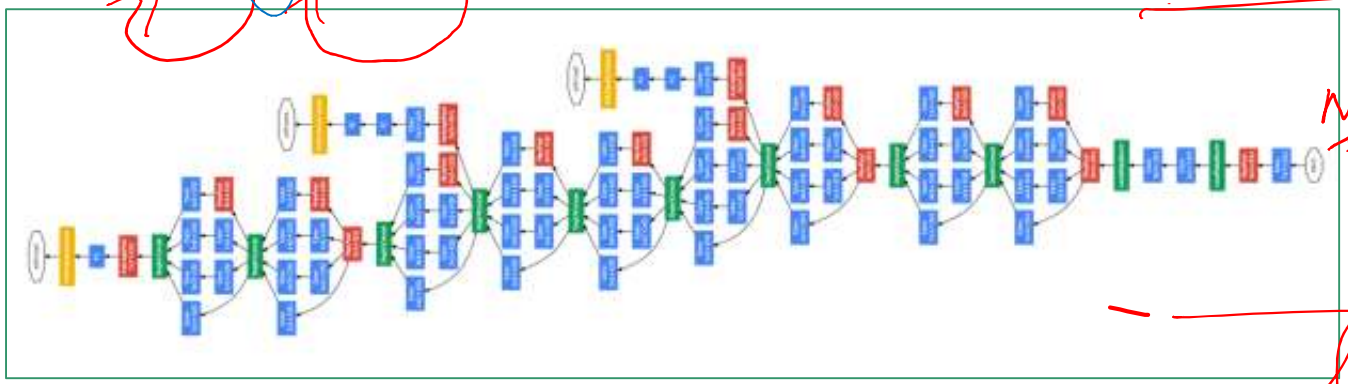
Machine-generated (but turker prefered)	a bicycle is parked next to a river
Human-annotated (but turker not prefered)	a bike sits parked next to a body of water

Image Captioning (deep vision + deep NLP)

CNN + RNN

Hierarchical Learning and Deep Architectures





The Google "Inception" deep neural network architecture for image recognition (27 layers)

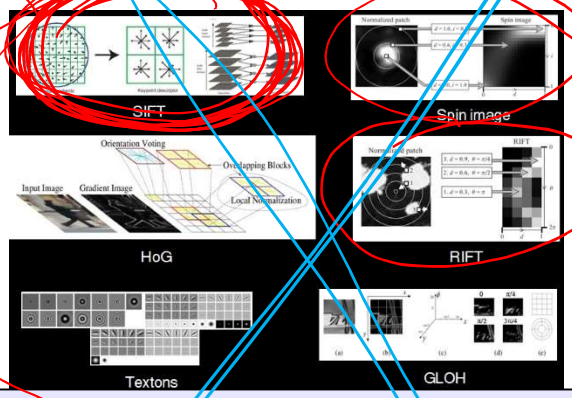
https://qph.ec.quoracdn.net/main-qimg-970d2b5f57b6b5cd13dc11f5371166b2-c
https://devblogs.nvidia.com/parallelforall/mocha-jl-deep-learning-julia/

Initial Drawbacks

1. Standard backpropagation with *sigmoid* does not scale well with multiple layers
   - ☹ Vanishing gradient -> weight of early layers change too slowly -> no learning
   - ☺ Mitigation: Use other activation functions (ReLU, Leaky-ReLU,…)
   - ☹ Exploding gradient -> very large weight updates -> network is unstable
   - ☺ Mitigation: "gradient clipping" (i.e. set maximum bounds on the gradients)
2. Overfitting
   - ☹ Large network -> lots of parameters -> capacity to "learn by heart"
   - ☺ Mitigation: regularization & dropout
3. Need lots of labeled data
   - ☹ Most data is not labeled
   - ☺ Mitigation: "pre-train" the network with features found automatically using unsupervised data -> Automatic feature learning…

Classic ML

Input

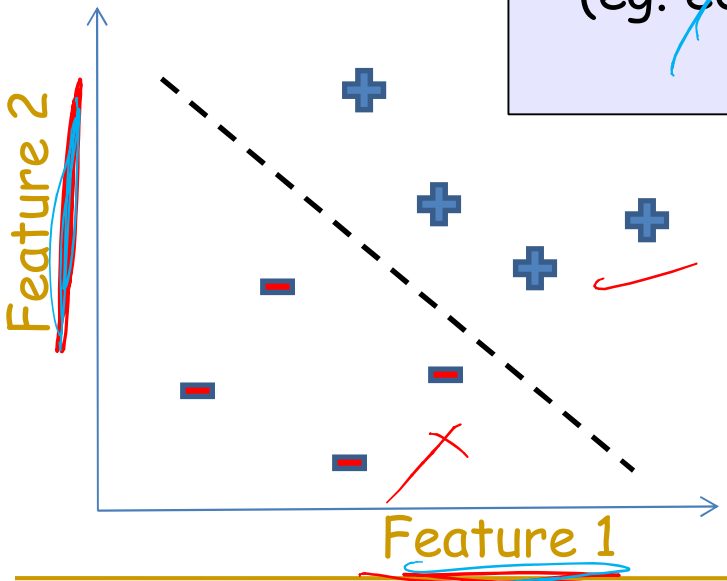


Manual Extraction of Features (eg. edge detection, colors, texture,…)

Learning algorithm

Classic ML, requires labeled data and hand-crafted features

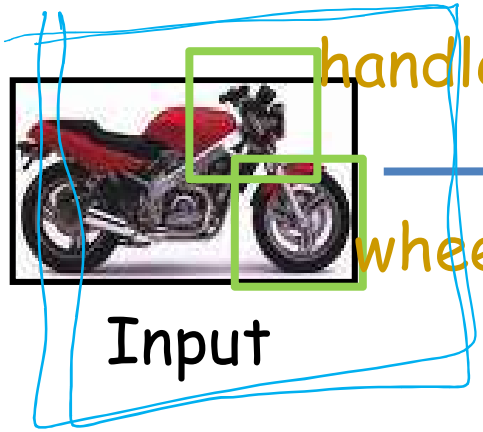
1. Needs expert knowledge
2. Time-consuming and expensive
3. Does not generalize to other domains



Motorbikes
"Non"-Motorbikes

Slide from Y. LeCun

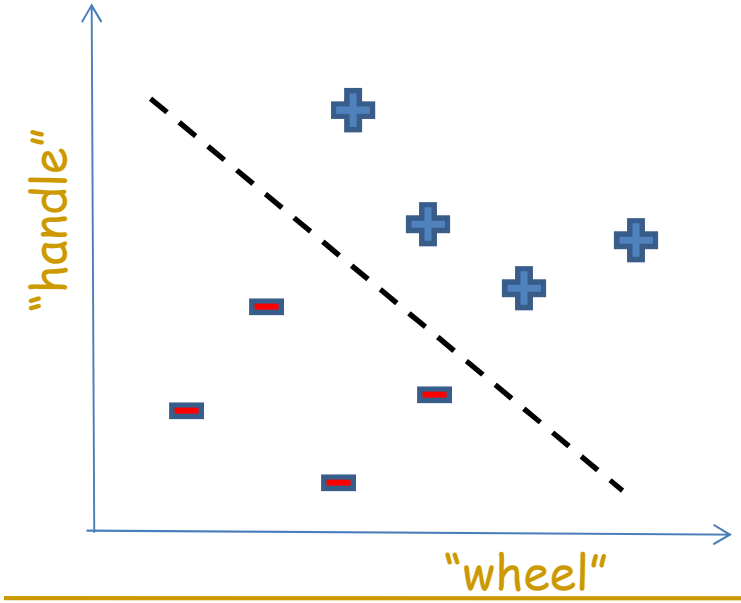
Automatic Feature Learning



Automatic Feature Representation

eg. handle, wheel, …

Learning algorithm



Motorbikes

"Non"-Motorbikes

With Automatic Feature Learning:

1. We feed the network the raw data (not feature-curated)

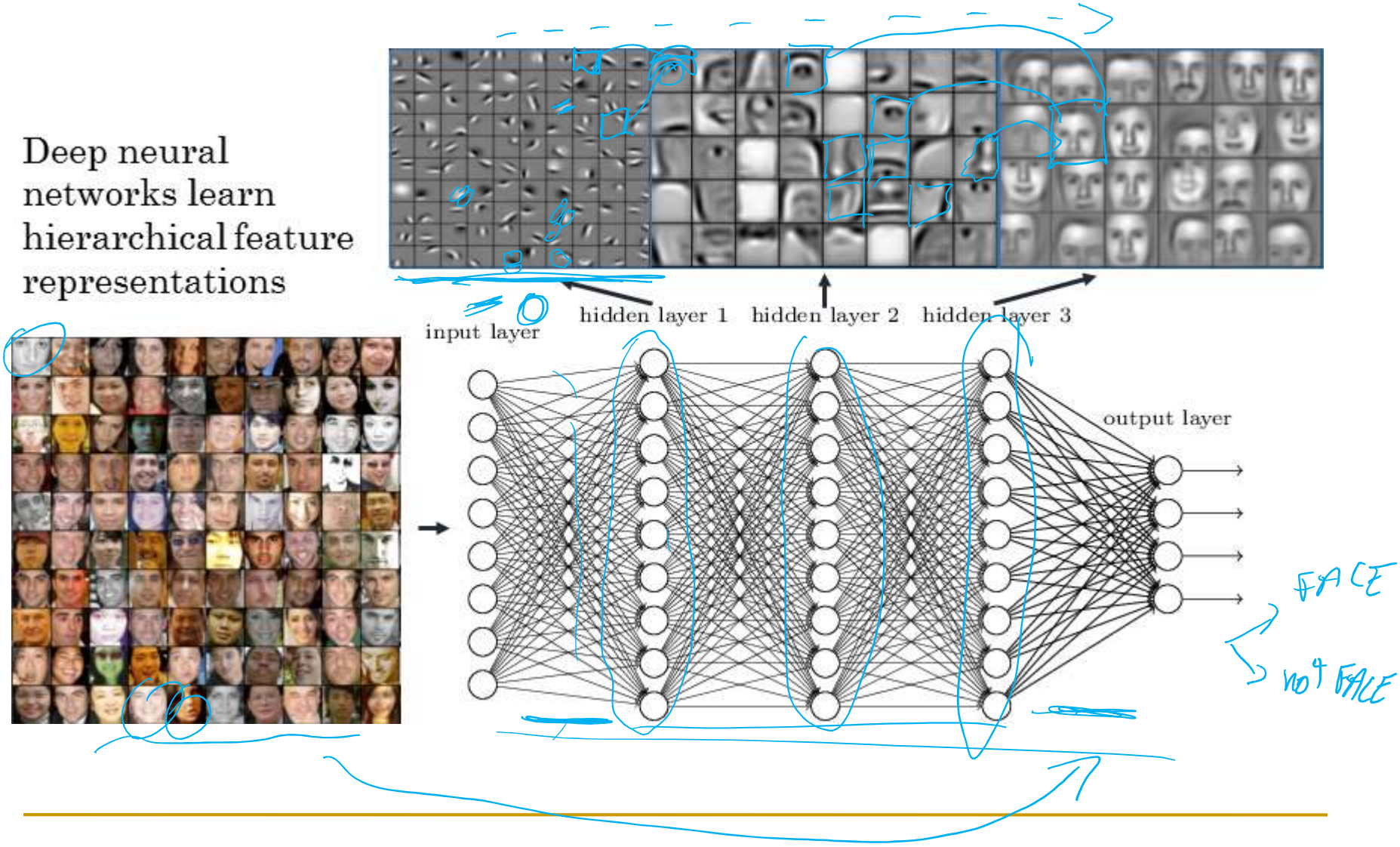
2. The features are learned by the network

3. Features learned can be re-used in similar tasks.

Slide from Y. LeCun

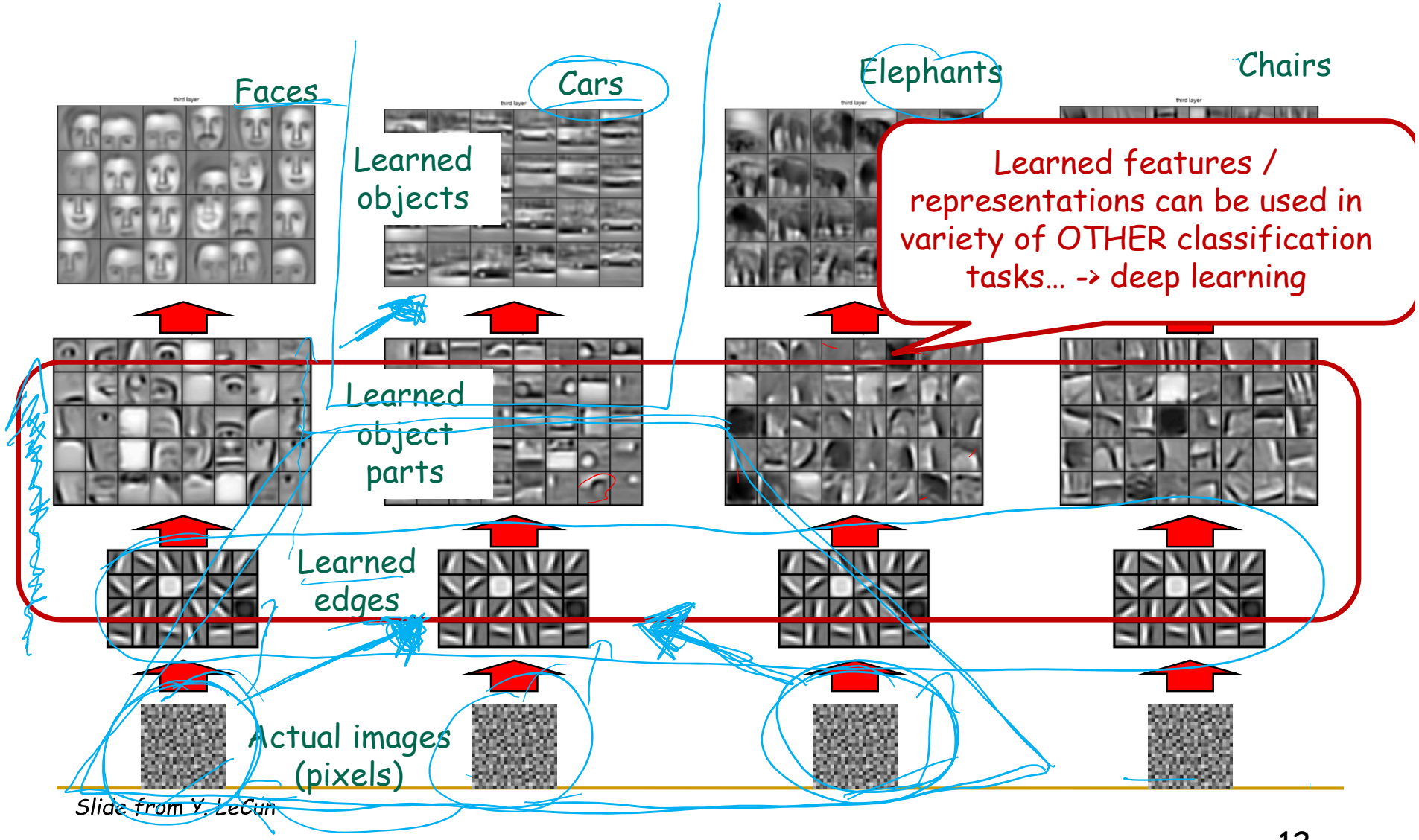
Automatic Feature Learning

Deep neural networks learn hierarchical feature representations

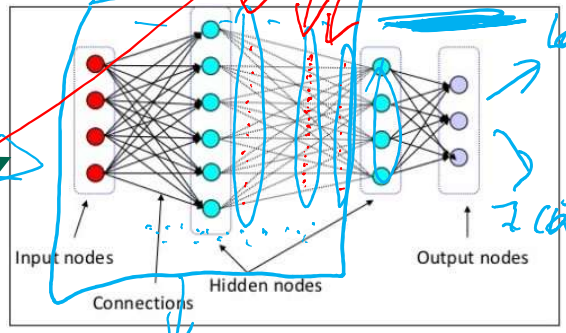


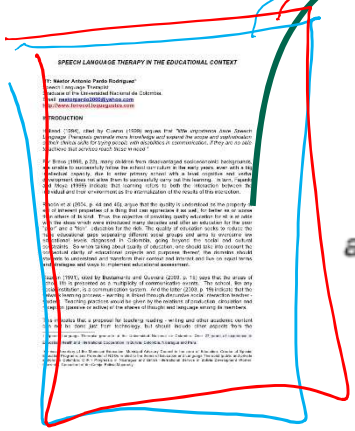
https://www.strong.io/blog-images/movie-posters/Slide6.png

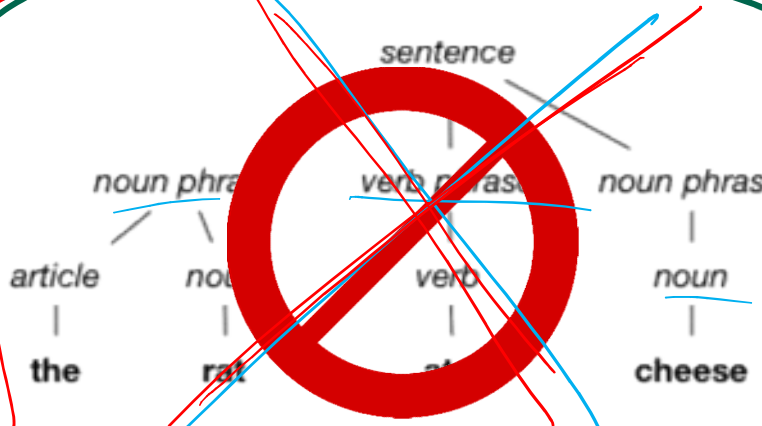
Re-use of Features

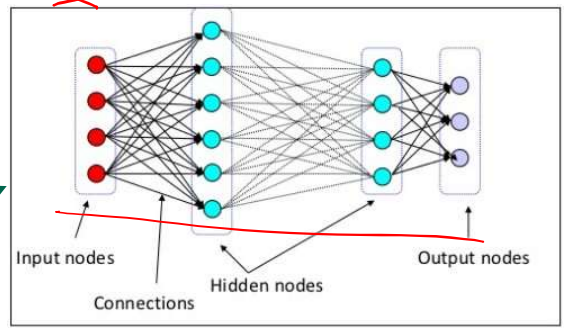


Advantages of Unsupervised Feature Learning

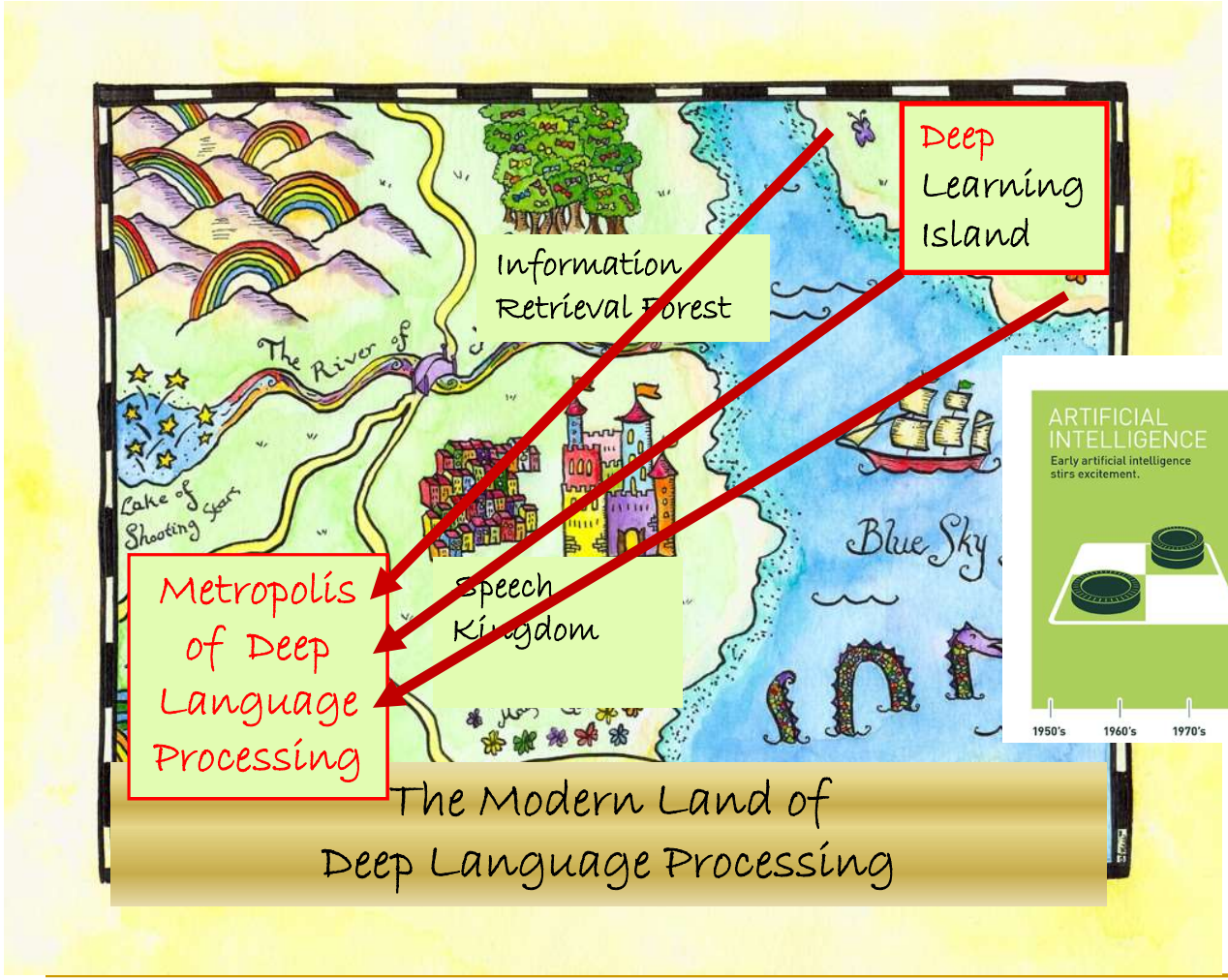


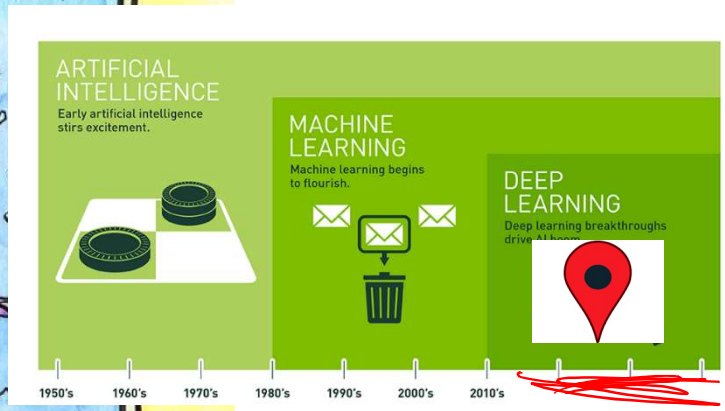




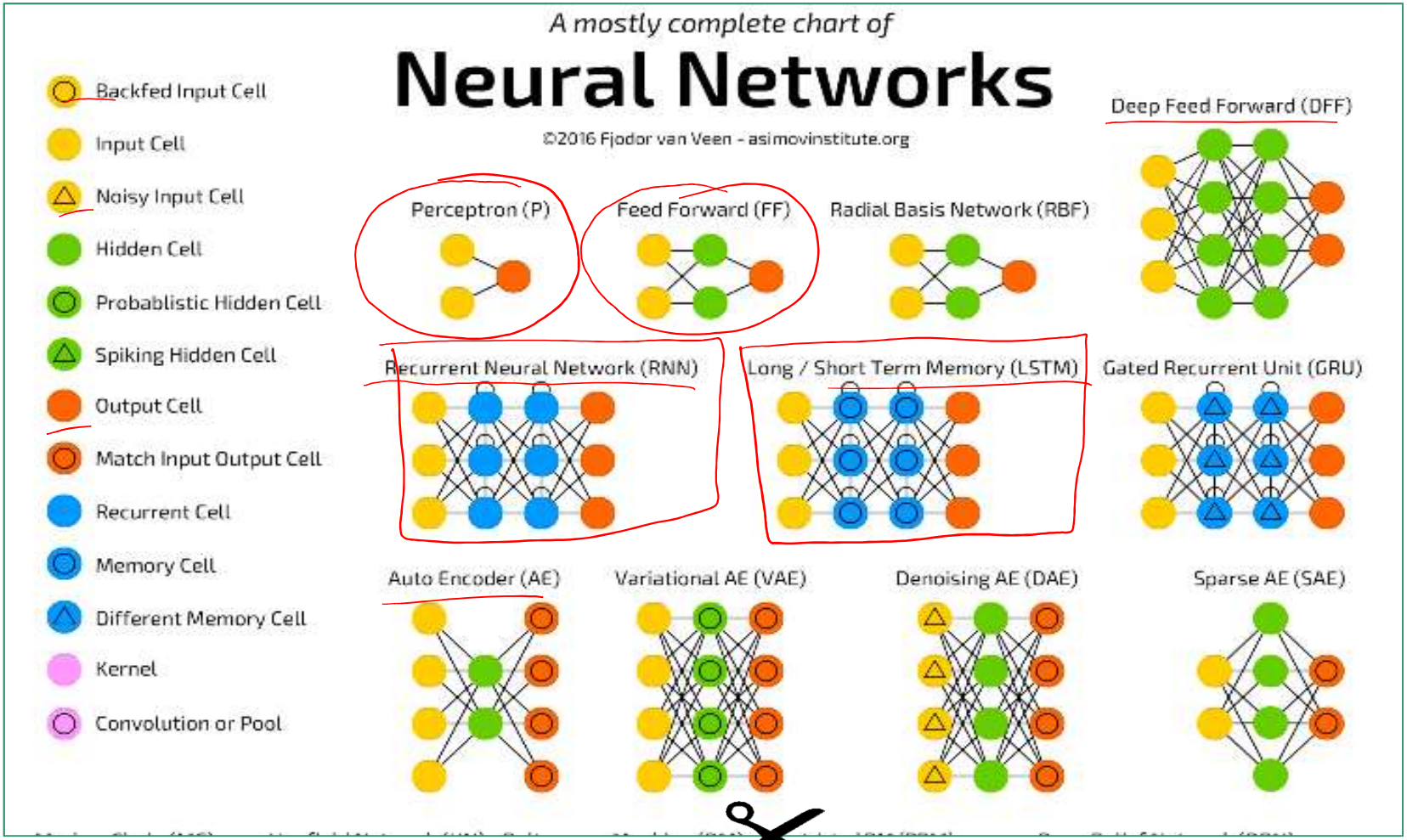


2nd Invasion of NLP, by Deep Learning (circa 2010-today)



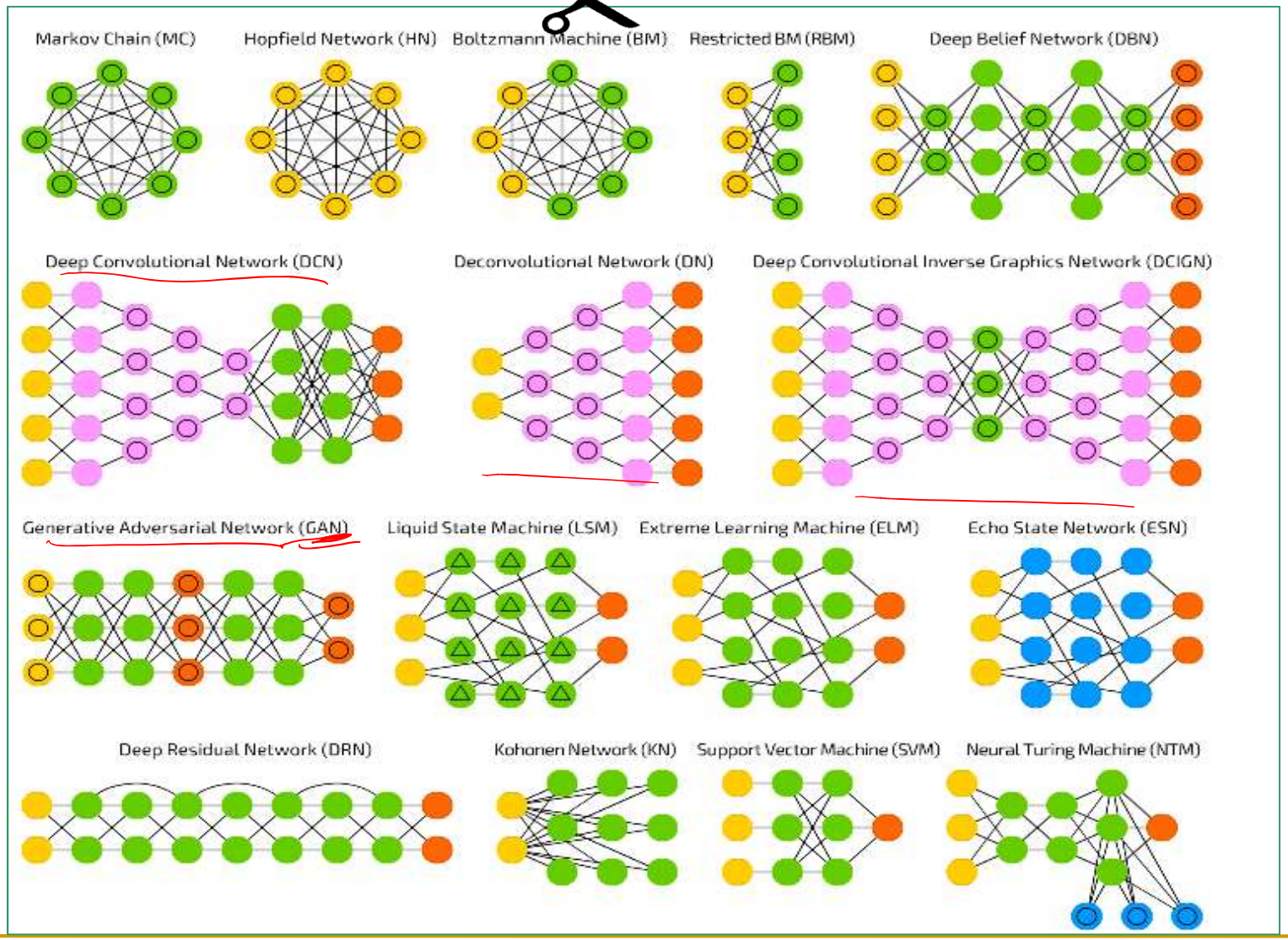


Many Types of Neural Networks



https://becominghuman.ai/cheat-sheets-for-ai-neural-networks-machine-learning-deep-learning-big-data-678c51b4b463

Many Types of Deep Networks (con't)



https://becominghuman.ai/cheat-sheets-for-ai-neural-networks-machine-learning-deep-learning-big-data-678c51b4b463

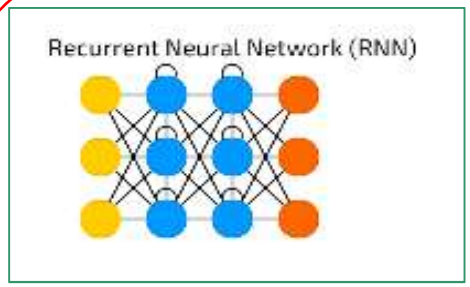
Deep Learning for NLP

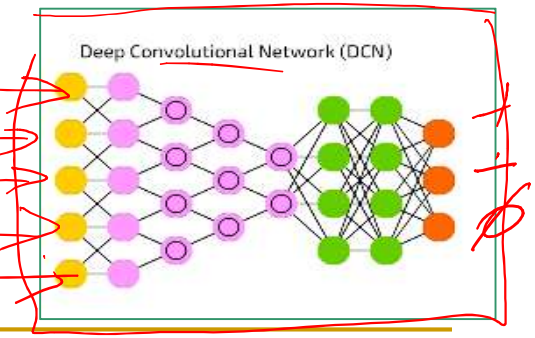
Deep learning models for NLP use

- Vector representation of words
  - i.e., word embeddings
- Neural network structures
  - Recurrent Neural Networks (RNNs)
  - Convolutional Networks (CNNs)
  - Recursive Neural Networks
  - …

next video

last video





Today

1. Introduction ✓
2. Bag of word model ✓
3. n-gram models ✓
4. Deep Learning for NLP ✓
    1. Word Embeddings
    2. Recurrent Neural Networks

Up Next

1. Introduction
2. Bag of word model
3. n-gram models
4. Deep Learning for NLP
    1. Word Embeddings
    2. Recurrent Neural Networks