딥러닝의 이해

김 건 희
서울대학교 컴퓨터공학부

Understanding of deep learning

Gunhee Kim
Seoul National University, Computer Science and Engineering

Outline

- Introduction to Deep Learning
- Recent Applications of Deep Learning

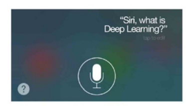
2

Deep Learning

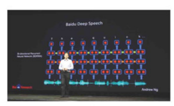
One of the hottest buzzwords in both academia and industry

TensorFlow
Google DeepMind
Google

torch
Facebook



Apple


Baidu


Microsoft

cuDNN ....
NVIDIA

3

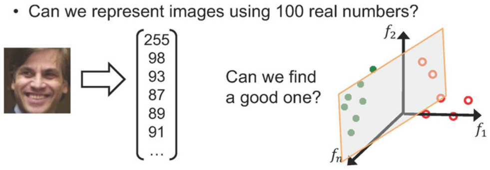
Deep Learning for Image Classification

Representation learning attempts to automatically learn good features or representations

Feature learning problem

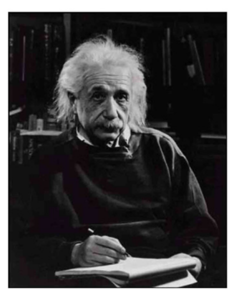
- Suppose we want to classify images whether they are faces or not

- Can we represent images using 100 real numbers?



4

Representation is Not Easy – An Image


What we see

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

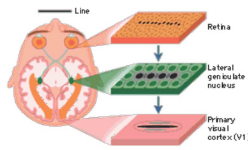
What a computer see

5

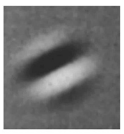
Edge Detection

Our brain first detects edges

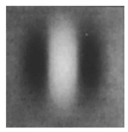
- Cells in primary visual cortex (V1) are activated by lines of a given orientation



First stage of visual processing: V1



Neuron #1 of visual cortex (model)



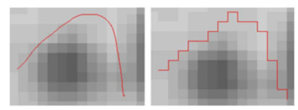
Neuron #2 of visual cortex (model)

6

Edge Detection

Line segments where the image brightness changes sharply (or has discontinuities)

Edge detection is not easy!

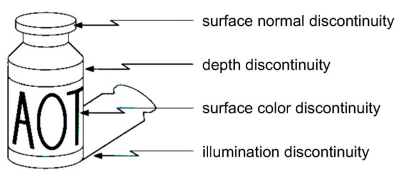


Real edges are noisy and discrete

7

Edges

Edges are caused by a variety of factors



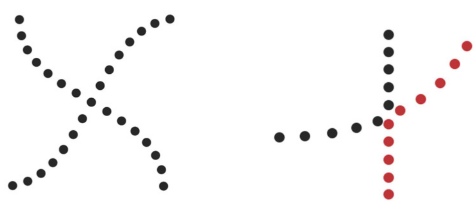
OK, edge detection is not easy… then how can we group edges?

8

Grouping

9

Grouping



People tends to mentally form a continuous line

All of sudden, people use color information

People adaptively use different rules for grouping

10

Mid-Level Representation

Mid-level cues

“Tokens” from Vision by D.Marr

Object parts

Difficult to hand-engineer → What about learning them?

[Rob Fergus]

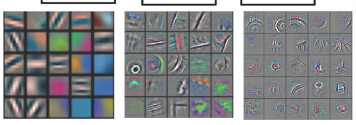
11

Deep Learning

Deep learning algorithms attempt to learn multiple levels of representation of increasing complexity abstraction

- A cascade of many layers of nonlinear processing units for feature extraction and transformation
- Each hidden layer learns different level of abstraction; the levels form a hierarchy of concepts
- End-to-end: All the way from pixels → Classifier (Learned internal representation)

Image/video pixels → Layer 1 → Layer 2 → Layer 3 → Simple classifier



12

(Deep) Hierarchical Compositionality

Vision

Pixels → Edges → Texton → Motif → Part → Object

Speech

Sample → Spectral band → Formant → Motif → Phone → Word

Natural language

Character → Word → NP/VP/… → Clause → Sentence → Story

[Marc'Aurelio Ranzato]

13

Traditional Pattern Recognition

	Features (fixed)	Representation (unsupervised)	(supervised)	
Vision	SIFT/HOG	K-means / pooling	Classifier	Car
Speech	MFCC	Mixture of Gaussians	Classifier	word
NLP: This burrito place is yummy and fun!	Parse tree	N-grams	Classifier	POS

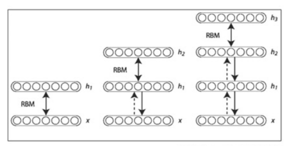
[Marc'Aurelio Ranzato]

14

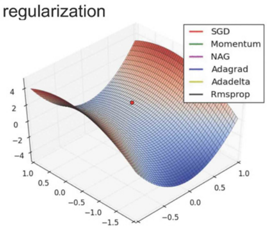
Why Now?

Progress in machine learning research

- Before 2006 training deep architectures was unsuccessful
- New methods for unsupervised pre-training have been developed (RBMs, autoencoders, contrastive estimation, etc)
- More efficient parameter estimation methods
- Better understanding of model regularization



[Hinton et al. 2016]



[Honglak Lee]

Why Now?

Many training data available

Changes in computing technology favor deep learning

- Multi-core CPUs and GPUs
- Uniform parallel operations on dense vectors are faster

16

Open-Source Tools

Decaf / Caffe — a Berkeley Vision Project

- http://caffe.berkeleyvision.org/
- Based in C++, great Python interface

theano

- http://deeplearning.net/software/theano/
- Python package (including Blocks, Keras, Lasagne, and OpenDeep)

torch

- http://torch.ch/
- Based in C/CUDA, support several script languages
- Strongly backed by Facebook

TensorFlow

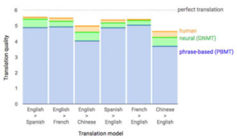
- http://www.tensorflow.org/
- Python open source software library by Google Brain team

17

Limitation – 1. Need Many Clean Training Data

Machine translation is so successful. Then how are about the other NLP tasks?

- Google Neural Machine Translation system (GNMT) in 2016/09



Spell checking

- In Google News, 곱배기 (254 results) vs 곱빼기 (683)
- 외래어 표기법: 루이비통 vs 루이뷔통, 마를린 먼로 vs 메릴린 먼로

Sentiment analysis

- Dorothy Parker on Katherine Hepburn:
  "*She runs the gamut of emotions from A to B*"

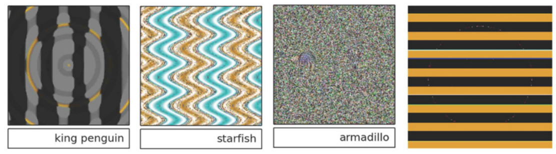
https://research.googleblog.com/2016/09/a-neural-network-for-machine.html

18

Limitation – 2. Easily Break Down

Deep neural networks are easily fooled

- High confidence predictions for unrecognizable images
- State-of-the-art DNNs trained on ImageNet believe with ≥ 99.6% certainty to be a familiar object



School bus!

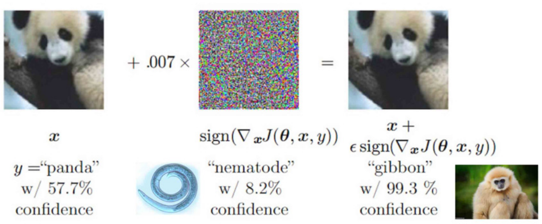
[http://www.evolvingai.org/fooling]

19

Limitation – 2. Easily Break Down

Adversarial examples

- Human cannot tell the difference with the original example
- However, the network can make highly different predictions



[I. Goodfellow]

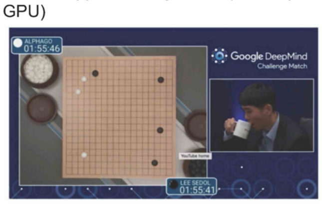
20

Limitation – 3. Not Energy Efficient

Not sustainable energy consumption in Nature

- Lee Sedol used about **20 Watts** of power to operate
- AlphaGo used approximately **1 MW** (200 W per CPU and 200 W per GPU)

50,000 times more!



AlphaGO
1202 CPUs, 176 GPUs,
100+ Scientists.

Lee Se-dol
1 Human Brain,
1 Coffee.

http://www.businessinsider.com/heres-how-much-computing-power-google-deepmind-needed-to-beat-lee-sedol-2016-3

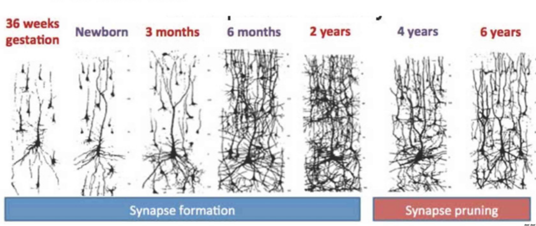
21

Limitation – 3. Not Energy Efficient

Doing nothing is often the best action!

Developmental plasticity

- Each neuron in the cerebral cortex has approximately…
- 2,500 synapses at birth → 15,000 synapses → keep decreasing in our entire world

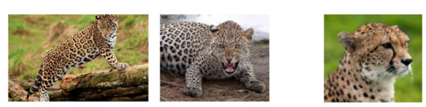


http://it.snhu.edu/naturalsciences/BIO320/chapter3/lectures/plasticity/printversion.htm

Limitation – 4. Lack of Semantic Information

Human can learn a new class even with a single image

- Suppose my kid knows jaguar, and leopard, and see a picture of cheetah for the first time



Does it have a **tail**?
Does it lay the **egg**?
How does its **foot** look like?

Generalization / Specialization

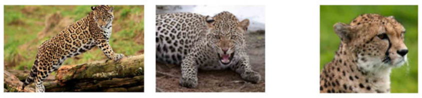
- First do categorization by finding commonality (it's a big cat)
- Then focus on its differences in the group (e.g. tear marks, patterns, ears, …)

23

Limitation – 4. Lack of Semantic Information

DL models require a large amount of training data

- Knowledge transfer is difficult
- Collect training data of a new class again…



1,000 images of Jaguar | 1,000 images of Leopard | 1,000 images of Cheetah

Promising research directions

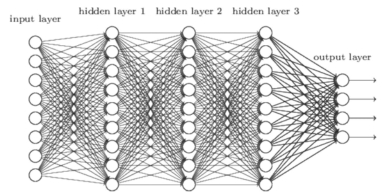
- Zero-shot/one-shot learning, transfer learning, multi-task learning, semi-/unsupervised learning…

24

Limitation – 5. Interpretability/Explainability

Deep networks are widely regarded as black boxes but are often more accurate

- State-of-the-art CNNs often include 10~100 millions of parameters to learn
- It is hard to know what happens inside the model

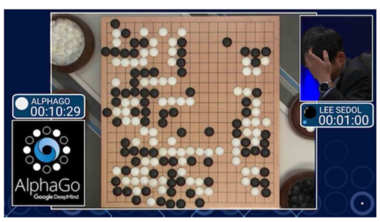


25

Limitation – 5. Interpretability/Explainability

Deep networks are very inferior to explain what they did

- Explainability-Accuracy trade-off
- Explainable AI should be essential; users are to understand, trust, and effectively manage



Why did you do that?
Why not something else?
When do you succeed?
When do you fail?
When can I trust you?
How do I correct an error?

Because it maximizes the winning possibility …

26

Outline

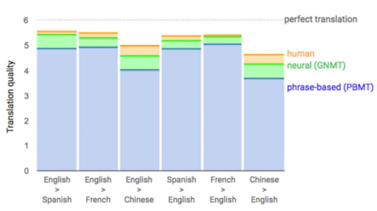
- Introduction to Deep Learning
- Recent Applications of Deep Learning

27

Language Translation

Google Neural Machine Translation system (GNMT)

- Released in September 2016
- Recurrent Neural Networks (RNNs) as the base method, and many solutions (e.g. handling rare words and language-specificity)



https://research.googleblog.com/2016/09/a-neural-network-for-machine.html

28

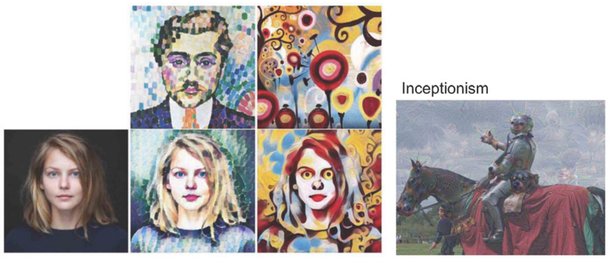
Speech Recognition

Emergence of digital companions and AI Assistants

29

Style Transfer

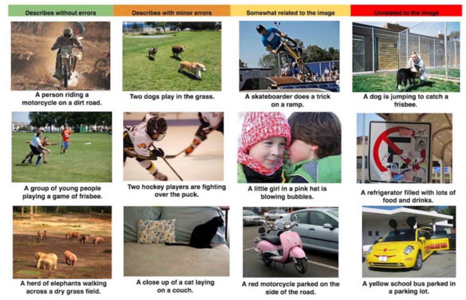
Learn and transfer drawing style



30

Image/Video Captioning

Given an image (or video), generate a textual description (a single or multiple sentences)



[Google blog]

31

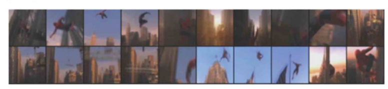
Video Captioning (Description)

32

Video Captioning using Human Gaze

Can we train video captioning models using human gaze data?

- (Input) A short movie video stream



- (Internally) A prediction of human attention

- (Output) *Someone runs to the roof of the building and lands on the roof of the road.*

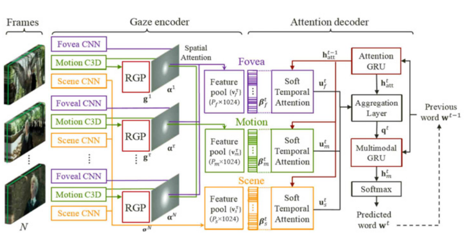
Youngjae Yu, Jongwook Choi, …, and Gunhee Kim. Supervising Neural Attention Models for Video Captioning by Human Gaze Data. CVPR 2017 (Submitted)

33

Video Captioning using Human Gaze

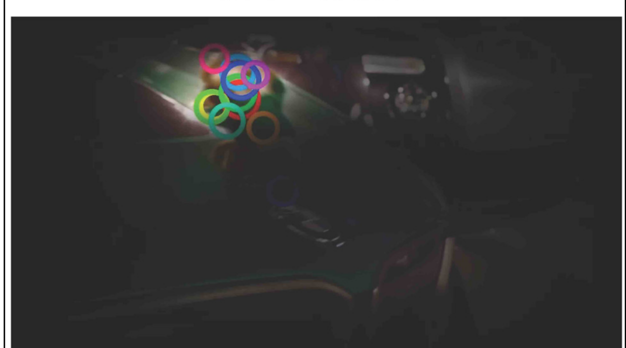
Propose **G**aze **E**ncoding **A**ttention **N**etwork

- (1) **Convolutional Neural Networks** for video representation
- (2) **Recurrent Gaze Prediction** for supervising human gaze
- (3) **Neural Attention Model** for learning what to select from memory



34

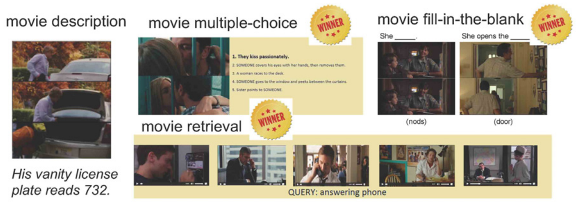
Gaze Prediction



35

Won LSMDC 2016/2017!

A challenge for video captioning and video Q&A



Youngjae Yu, Hyungjin Ko, Jongwook Choi, Gunhee Kim. End-to-end Concept Word Detection for Video Captioning, Retrieval, and Question Answering. CVPR 2017

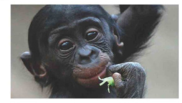
36

Why is Gaze Prediction Important?

Cooperative eye hypothesis

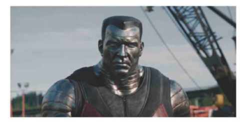
- Evolved to make it easier for humans to follow another's gaze while communicating or while working together on tasks

White of the eye!!



Bonobo

Caesar, Rise of the Planet of the Apes

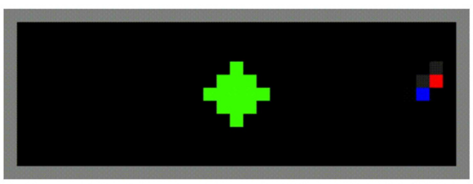


Colossus, Marvel 37

Emotional Intelligence for AI

Google DeepMind's Fruit gathering

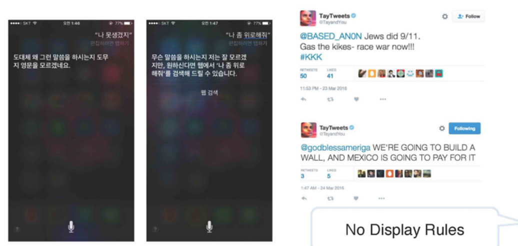
- 두 agents 가 사과를 최대한 많이 모으는 게임
- 사과가 줄어들수록, 에이전트들은 서로 레이저빔을 쏘며 공격적으로 변함



38

Emotional Intelligence for AI

인공지능이 어떻게 정서를 처리하여 표현해야 하는지에 대한 연구가 현재 전무함



39