Seminar

1. Turing Lecture (ETC 35-40 mins)
2. TMUX 사용법 (ETC 3 mins)

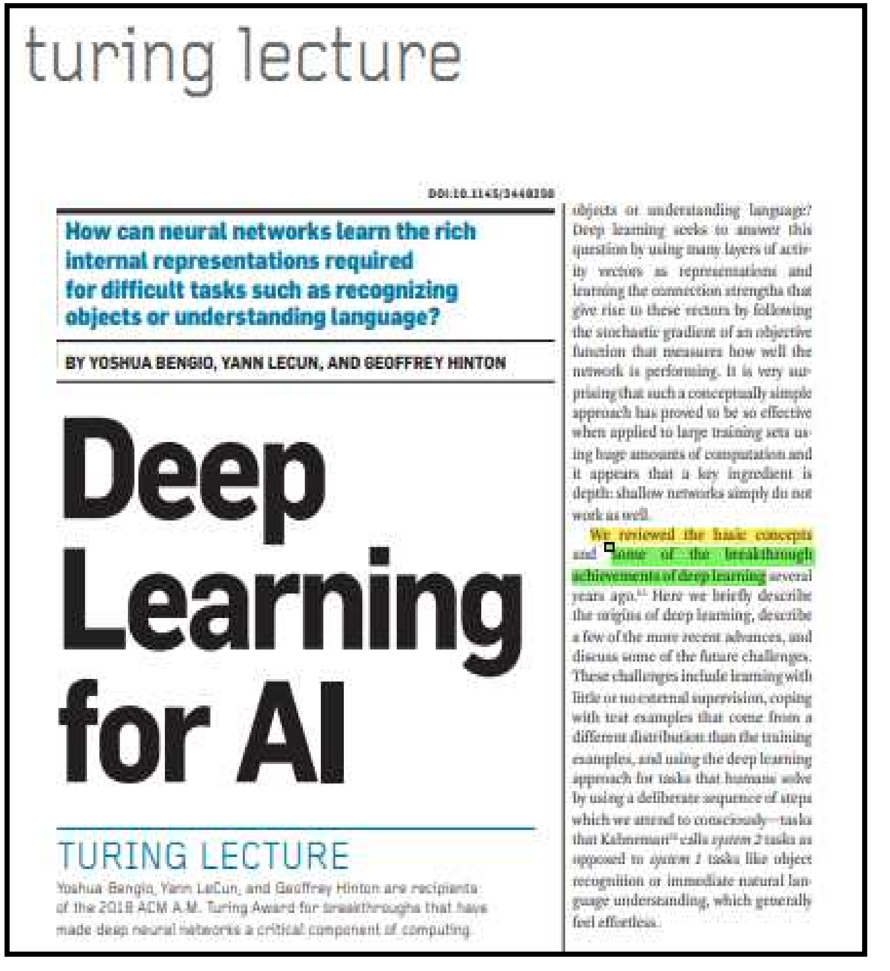
Seminar 1. Turing Lecture

Turing Lecture?

- AI 3대 천왕 (벤지오, 얀라쿤, 힌튼)이 2018년 튜링 상을 공동 수상한 이후, 2년간 공동 작업 후 출판한 논문
  - ➢ Communications of the ACM (2021)
- AI를 위한 딥러닝 요약본 (한계점, 해결책 제시)
  - ➢ Main Stream에서 과거/현재/미래의 관점으로 요약함
- 셋 이 합쳐 1,000,000 Citations (3대장)



YOSHUA BENGIO, YANN LECUN, and GEOFFREY HINTON



turing lecture

DOI:10.1145/3448250

How can neural networks learn the rich internal representations required for difficult tasks such as recognizing objects or understanding language?

BY YOSHUA BENGIO, YANN LECUN, AND GEOFFREY HINTON

Deep Learning for AI

TURING LECTURE

Yoshua Bengio, Yann LeCun, and Geoffrey Hinton are recipients of the 2018 ACM A.M. Turing Award for breakthroughs that have made deep neural networks a critical component of computing.

objects or understanding language? Deep learning seeks to answer this question by using many layers of activity vectors as representations and learning the connection strengths that give rise to these vectors by following the stochastic gradient of an objective function that measures how well the network is performing. It is very surprising that such a conceptually simple approach has proved to be so effective when applied to large training sets using huge amounts of computation and it appears that a key ingredient is depth: shallow networks simply do not work as well.

We reviewed the basic concepts and some of the breakthrough achievements of deep learning several years ago.[1] Here we briefly describe the origins of deep learning, describe a few of the more recent advances, and discuss some of the future challenges. These challenges include learning with little or no external supervision, coping with test examples that come from a different distribution than the training examples, and using the deep learning approach for tasks that humans solve by using a deliberate sequence of steps which we attend to consciously—tasks that Kahneman[2] calls *system 2* tasks as opposed to *system 1* tasks like object recognition or immediate natural language understanding, which generally feel effortless.

Seminar 1. Turing Lecture

Why Turing Lecture?

- 지난 세미나에서 SSL을 자세히 다루었기 때문에, 또 SSL을 발표하기 보다는 정보 공유 관점에서 준비
- 이미 친숙한 개념 (AI, Attention, Transformer. . .)들에 대해서 대가들이 어떻게 바라보는지 확인
  - ➢ 향후 연구에 다양한 형태로 도움이 될 것이라고 생각

- 총 4개의 섹션

Section 1. From Hand-Coded Symbolic Expressions to Learned Distributed Representations (Past : ~ 2012)

Section 2. The Rise of Deep Learning (Past-Present : 2012 ~ 2017)

Section 3. Recent Advances (Present : 2017 ~ 2021)

Section 4. The Future of Deep Learning (Future : 2022 ~)

Seminar 1. Turing Lecture

- 총 4개의 섹션

Section 1. From Hand-Coded Symbolic Expressions to Learned Distributed Representations (Past : ~ 2012)

- AI를 어떻게 생각하고있는지?

Section 2. The Rise of Deep Learning (Past-Present : 2012 ~ 2017)

- 왜 깊은 네트워크가 성능이 좋은지?

- 왜 unsupervised pre-traininig이 성능에 도움이 되는가?

Section 3. Recent Advances (Present : 2017 ~ 2021)

- Attention / Transformer

Section 4. The Future of Deep Learning (Future : 2022 ~)

Section 1. From Hand-Coded Symbolic Expressions to Learned Distributed Representations

(Past: ~ 2012)

제목은 거창하나, 단순히 AI가 무엇이라고 생각하는지 알려주는 정도

Turing Lecture

Section 1. From Hand-Coded Symbolic Expressions to Learned Distributed Representations (Past: ~ 2012)

AI

Learning Paradigm

Logic-Inspired Paradigm

Sequential reasoning as the essence of intelligence

→ Rule-based AI

Brain-Inspired Paradigm

Learning representations from data as the essence of intelligence

→ Data-based AI

Turing Lecture

Section 1. From Hand-Coded Symbolic Expressions to Learned Distributed Representations (Past: ~ 2012)

AI

Learning Paradigm

Logic-Inspired Paradigm

Sequential reasoning as the essence of intelligence

컴퓨터 공학적 발전

→ Rule-based AI

Brain-Inspired Paradigm

Learning representations from data as the essence of intelligence

컴퓨터 과학적 발전

→ Data-based AI

Turing Lecture

Section 1. From Hand-Coded Symbolic Expressions to Learned Distributed Representations (Past: ~ 2012)

AI

Learning Paradigm

→ **Logic-Inspired Paradigm**

Sequential reasoning as the essence of intelligence

→ **Rule-based AI**

하이브리드 AI

→ **Brain-Inspired Paradigm**

Learning representations from data as the essence of intelligence

→ **Data-based AI**

Turing Lecture

Section 1. From Hand-Coded Symbolic Expressions to Learned Distributed Representations (Past: ~ 2012)

AI

Learning Paradigm

Logic-Inspired Paradigm

Sequential reasoning as the essence of intelligence

→ Rule-based AI

Brain-Inspired Paradigm

Learning representations from data as the essence of intelligence

→ Data-based AI

Turing Lecture

Section 1. From Hand-Coded Symbolic Expressions to Learned Distributed Representations (Past: ~ 2012)

Brain-Inspired Paradigm

Learning representations from data as the essence of intelligence

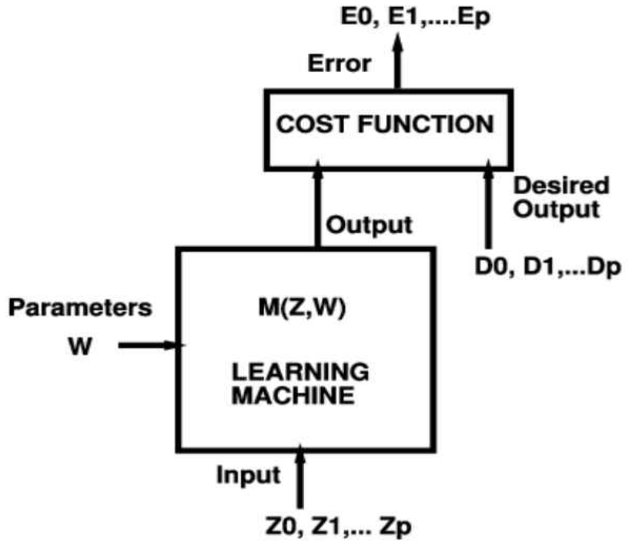
→ Data-based AI

- 'Gradient based Learning Applied To Document Recognition' 에서 시작
- Machine Learning 알고리즘 학습을 위해
  - Data로 부터 패턴을 학습
  - Gradient에 기반

1998

BENGIO PUBLISHES A GROUNDBREAKING PAPER

Bengio publishes a groundbreaking paper, "Gradient based Learning Applied To Document Recognition," showing that, if provided the right architecture, certain algorithms can recognize images, such as handwritten characters, more accurately than conventional technology, which typically relied on hand-designed heuristics.



Yann LeCun's Figure

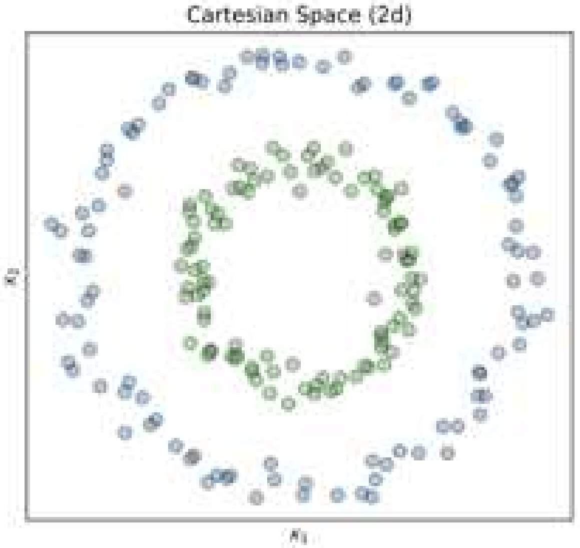
Turing Lecture

Section 1. From Hand-Coded Symbolic Expressions to Learned Distributed Representations (Past: ~ 2012)

Brain-Inspired Paradigm

Learning representations from data as the essence of intelligence

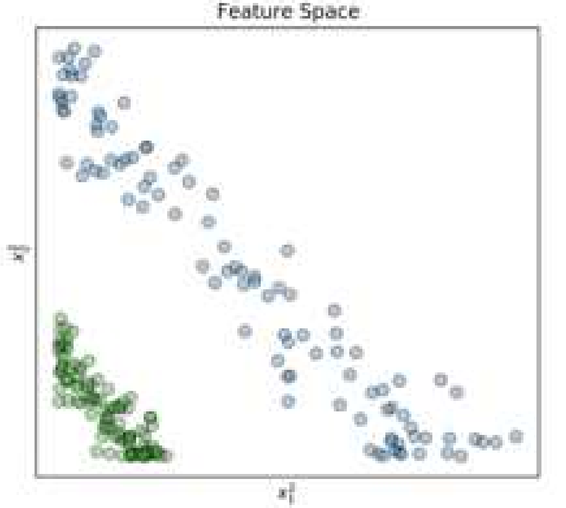
→ **Data-based AI**



Mapping

학습 된 ML 알고리즘은
데이터의 표현을 Task-Specific하게 바꿈

→ Representation Learning



Turing Lecture

Section 1. From Hand-Coded Symbolic Expressions to Learned Distributed Representations (Past: ~ 2012)

Representation learning이 가능한 이유:

'Neural activity of vectors'

Neural activity: capture relationships between concepts

Goal: Automatic and Ultimate generalization

Brain-Inspired Paradigm

Learning representations from data as the essence of intelligence

→ Data-based AI

Mapping

데이터의 표현을 Task-Specific하게 바꿈

→ Representation Learning

Section 2. The Rise of Deep Learning

(Past-Present: 2012 ~ 2017)

Turing Lecture

Section 2. The Rise of Deep Learning

(Past-Present: 2012 ~ 2017)

1) Depth

2) Unsupervised pre-training

3) The mysterious success of rectified linear units

4) Breakthroughs in speech and object recognition

Turing Lecture

Section 2. The Rise of Deep Learning
(Past-Present: 2012 ~ 2017)

1) Depth

- Network 'depth'
  - <u>네트워크의 구조가 '깊어'질수록</u>, 성능이 좋더라 !
  - 네트워크가 깊어지면서 Learnable Parameter가 많아져서 (대중)

    No ! 같은 Parameter를 갖는 다른 네트워크보다 성능이 높더라

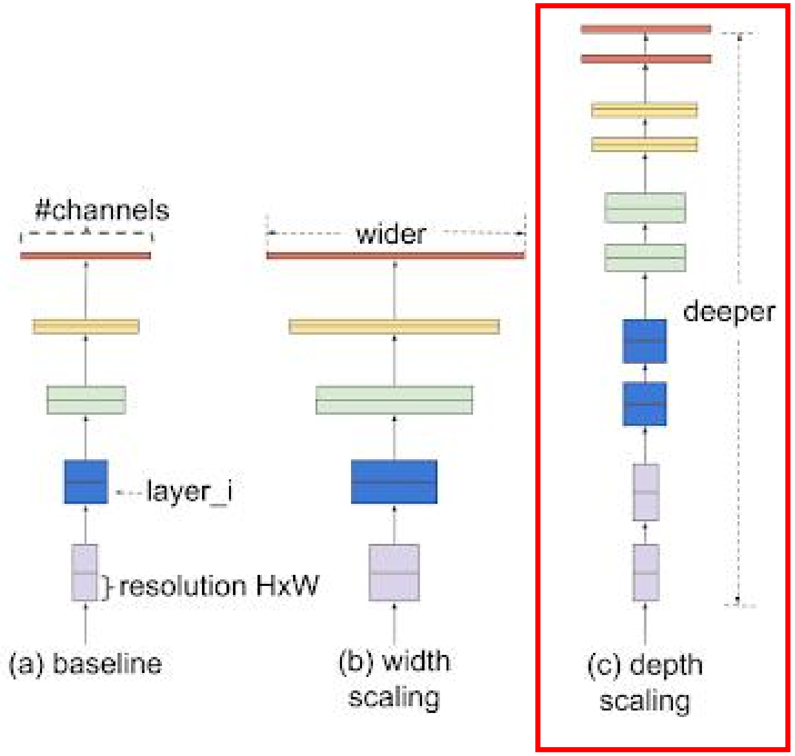
- Why?
  - 매 layer에서 추출된 features가 다양한 방식으로 combined 되게 때문
  - Exploitation on particular form of compositionality
  - 예시)

    (b): 5회의 Compositionality

    (C): 10회의 Compositionality

    동일한 파라미터를 갖는 모델임에도, 성능 차이가 나는 원인

Better

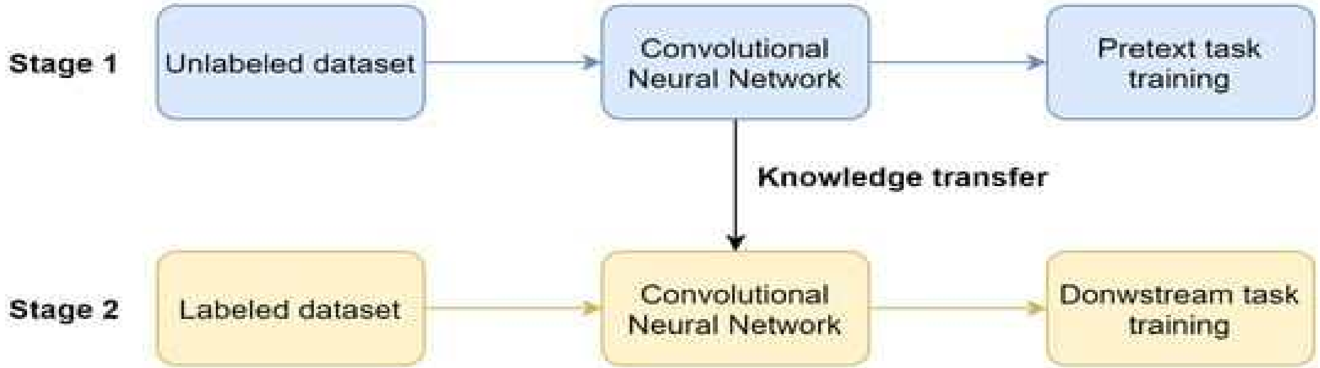


Related works: GoogleNet, VGG, Inception, ResNet

Turing Lecture

Section 2. The Rise of Deep Learning
(Past-Present: 2012 ~ 2017)

2) Unsupervised pre-training



- **Definition:** Learning with other source of information → fine-tune the layers with labeled-data
  - ➢ Ex) Using ImageNet pre-trained networks
- Why is it called "Unsupervised" ?
  - ➢ The first paper: 'Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles'
  - ➢ Can be called 1) transfer learning, 2) self-supervised, 3) unsupervised representation learning

2) Unsupervised pre-training

Why good performance?

- Pre-trained Layers를 사용하면, final classification과는 상관없는 여러 종류의 features를 추출
  - 여러 종류의 features가 AI모델의 일반화성능을 우수하게 함
  - Pre-training + Fine-tuning 한 Network가 generalization 성능도 우수하더라
  - Follow-up 장점
    - 1) 데이터를 구하고, 2) 값비싼 레이블링을 진행하는 것보다 훨씬 경제적이고 현실적

- 결론적으로
  - Pre-training은 General Deep Learning (Ultimate goal of AI) 에 필수적인 전처리 작업
  - 최근 Meta-learning을 포함한 모든 학습 paradigm에도 도움이 됨
  - Unsupervised pre-training을 발견해서 다행이다 !

Turing Lecture

Section 2. The Rise of Deep Learning
(Past-Present: 2012 ~ 2017)

3) The mysterious success of rectified linear units

- 여러 Activation Layer 중 유독 'ReLU'가 왜 잘 working하는지 아직 분명히 밝혀지지 않음
- 저자들은 다른 layer에 비해 'make it easy to train deep networks'

4) Breakthroughs in speech and object recognition

- Speech와 Object를 '인식'하는 Task에서 breakthrough가 되어 주었기 때문에 빠른 속도로 발전하지 않았을까 생각

Section 3. Recent Advances

(Present : 2017 ~ 2021)

Turing Lecture

Section 3. Recent Advances

(Present : 2017 ~ 2021)

1) Soft Attention (Self-Attention)
2) Transformer Architecture
3) Representation Learning
   - Unsupervised and Self-Supervised Learning
   - Contrastive Self-Supervised Learning

Turing Lecture

Section 3. Recent Advances
(Present : 2017 ~ 2021)

1) Soft Attention (Self-Attention)

2) Transformer Architecture → 모든 딥러닝 아키텍처/메커니즘을 개발
Ex) CNN/GAN/GNN/RNN . . .

3) Representation Learning

- Unsupervised and Self-Supervised Learning
- Contrastive Self-Supervised Learning

Turing Lecture

Section 3. Recent Advances
(Present : 2017 ~ 2021)

1) Soft Attention (Self-Attention)

- (Selective) Attention은 어떻게 AI가 '특정한 객체/정보'에 집중할 수 있게 할 것인가에 대한 방법론
- 직관적으로, Attention은 불필요한 정보를 덜어내고, 정말 중요한 정보에만 집중하도록 도움
  - From Jonathan Hui (Google Deep Learning Researcher)

Input

Label: 한 남자가 여러 플라스틱 박스를 들고 교차로에서 나에게 건너오고 있다

Turing Lecture

Section 3. Recent Advances

(Present : 2017 ~ 2021)

1) Soft Attention (Self-Attention)

Selective Attention

Soft Attention

Reflect information globally

Hard Attention

Reflect information locally

Turing Lecture

Section 3. Recent Advances
(Present : 2017 ~ 2021)

1) Soft Attention (Self-Attention)

Selective Attention

Soft Attention

Reflect information globally

좋은 방법론
Recent Advancements!

Hard Attention

Reflect information locally

Turing Lecture

Section 3. Recent Advances
(Present : 2017 ~ 2021)

1) Soft Attention (Self-Attention)

삼대장이 정의하는 Attention:

- Attention은 신경망의 구조를 바꾸는 방법론
  - NNs이 input에 대해서 동적처리를 할 수 있게 바꿈
    - ✓ 기존에는 학습이 끝나면 파라미터가 고정되었고, 고정된 파라미터는 '동적처리'가 불가했음
  - Attention의 구조는 인풋을 동적으로 처리 하게하면서도 '미분이 가능'하기때문에 의미가 큼

삼대장이 말하는 Attention의 활용법:

- Input의 순서에 independent한 output을 산출 가능
  - Input에 노이즈가 많거나, 무작위성을 띄는 데이터에 활용 시 좋음
- 서로 다른 Input들에 대해서 relationships을 modelling 가능
  - 데이터의 관계를 파악하기 어렵고, 직관적인 수치해석이 불가능한 데이터에 활용 시 좋음

Turing Lecture

Section 3. Recent Advances
(Present : 2017 ~ 2021)

1) Soft Attention (Self-Attention)

Attention이 성능을 향상시키는 이유:

- Soft Attetion은 input이 modules (layers/units)을 지나가면서 '동적으로 정보를 처리'
- 이는 Out-of-Distribution 관점에서 일반화 성능을 높인다
  - 불필요한 정보가 필요 없을 때는 활용 X
  - 불필요한 정보가 필요할 때는 적절히 활용 O

- Attention으로만 구현 된 Transformer-based 모델이 일반화 성능이 좋은 이유
  - 그러나, Data-Efficient 하지 않음 → Data-Effective 함, 많은 학습 데이터 필요
  - 최근 잘나가는 Behchmark 데이터셋은 모두 Large-Scale이기 때문에 우수한 일반화 성능이 필수적

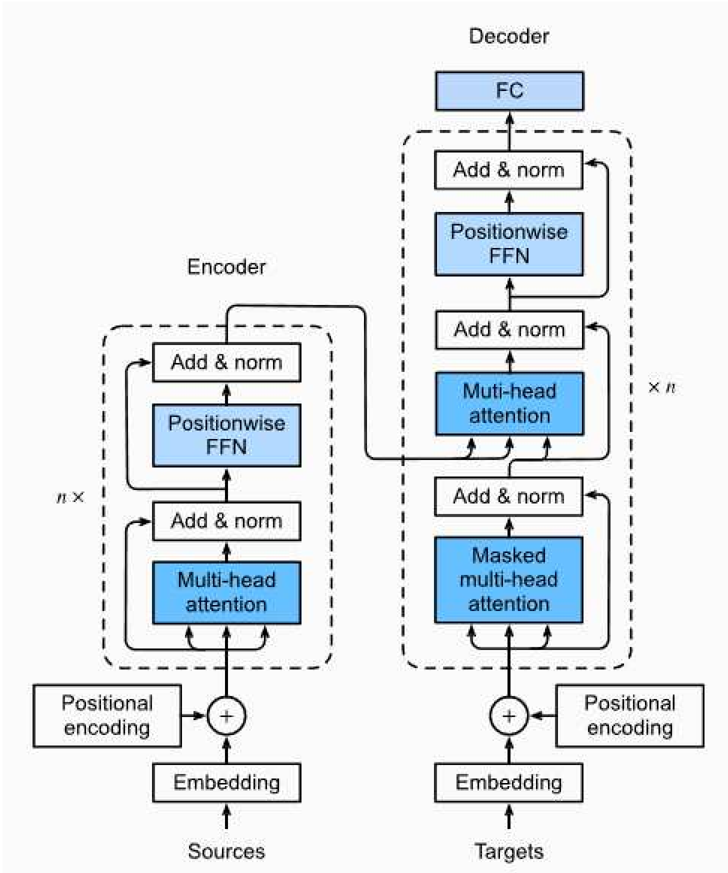
Turing Lecture

Section 3. Recent Advances
(Present : 2017 ~ 2021)

2) Transformer Architecture

Transformer?

- Stacking many layers of '(soft) self-attention' modules
- 성능이 좋은 이유는 앞서 설명 (동적처리/일반화 성능 우수)



특징

- Dominant models in many applications
- Transformer는 pre-trained를 기본전제로 함

성능이 좋은 이유 (추가설명)

- making it possible to operate on sets of vectors rather than single vectors as in traditional neural networks.

Turing Lecture

Section 3. Recent Advances
(Present : 2017 ~ 2021)

2) Transformer Architecture

Transformer를 다른 구조보다 최고의 구조로 선택한 이유

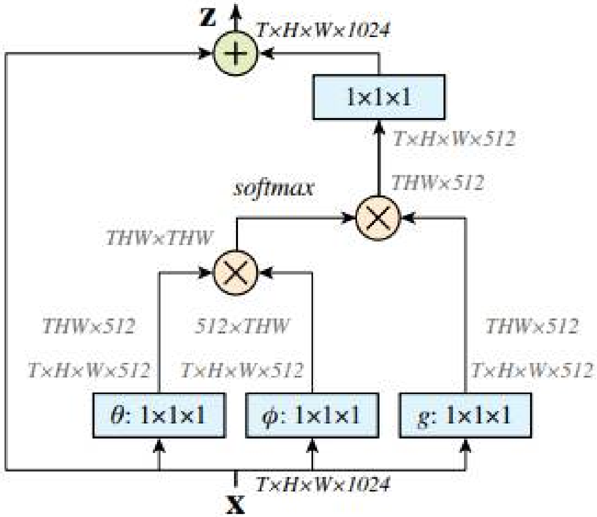
- Perhaps more surprisingly, transformers have been used successfully to solve integral and differential equations symbolically
  - Symbol로 표현 되어있는 미적분 문제를 성공적으로 풀었기 때문임
  - 이는 향후 Human-Level AI를 연구하는데 큰 가능성을 제시

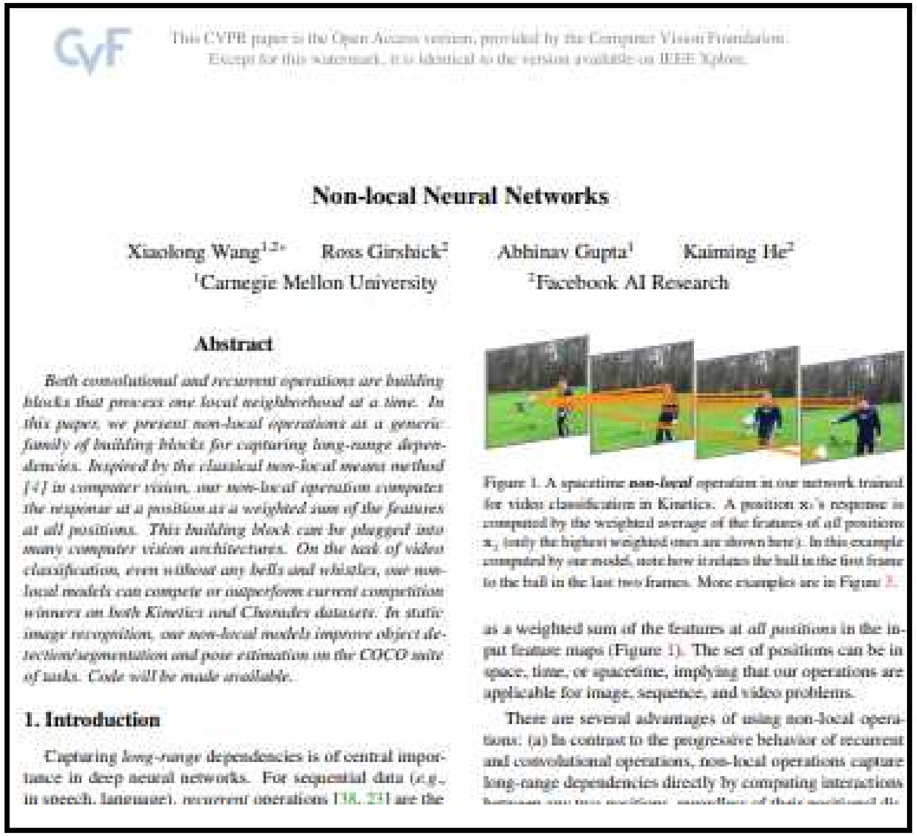
Turing Lecture

Section 3. Recent Advances
(Present : 2017 ~ 2021)

Additional Information

- "Non-local Neural Networks" by Kaiming He
- Attention구조를 수학적으로 일반 뉴럴넷 형태로 정의
  - Ex) Sequence 모델 (RNN,LSTM,GRU, . . .)
- 모든 데이터형태에 상관없이 적용가능한 구조





Non-local Neural Networks

Xiaolong Wang[1,2*] Ross Girshick[2] Abhinav Gupta[1] Kaiming He[2]
[1]Carnegie Mellon University [2]Facebook AI Research

Abstract

Both convolutional and recurrent operations are building blocks that process one local neighborhood at a time. In this paper, we present non-local operations as a generic family of building blocks for capturing long-range dependencies. Inspired by the classical non-local means method [4] in computer vision, our non-local operation computes the response at a position as a weighted sum of the features at all positions. This building block can be plugged into many computer vision architectures. On the task of video classification, even without any bells and whistles, our non-local models can compete or outperform current competition winners on both Kinetics and Charades datasets. In static image recognition, our non-local models improve object detection/segmentation and pose estimation on the COCO suite of tasks. Code will be made available.

1. Introduction

Capturing *long-range* dependencies is of central importance in deep neural networks. For sequential data (*e.g.*, in speech, language), recurrent operations [38, 23] are the

Figure 1. A spacetime ***non-local*** operation in our network trained for video classification in Kinetics. A position $\mathbf{x}_i$'s response is computed by the weighted average of the features of *all* positions $\mathbf{x}_j$ (only the highest weighted ones are shown here). In this example computed by our model, note how it relates the ball in the first frame to the ball in the last two frames. More examples are in Figure 3.

as a weighted sum of the features at *all positions* in the input feature maps (Figure 1). The set of positions can be in space, time, or spacetime, implying that our operations are applicable for image, sequence, and video problems.

There are several advantages of using non-local operations: (a) In contrast to the progressive behavior of recurrent and convolutional operations, non-local operations capture long-range dependencies directly by computing interactions

Turing Lecture

Section 3. Recent Advances
(Present : 2017 ~ 2021)

3) Representation Learning

정의: 데이터의 representation을 주어진 task를 해결하는데 적절한 형태로 Mapping하는 방법론

종류

- Supervised Learning
  - Labeled 데이터를 갖고, 내부 패턴 representation을 학습하는 방법론
  - 많은 Task에서 매우 성공적인 결과를 생산
  - 그러나, 많은 양의 human-labeled data를 필요로 함
- Reinforcement Learning
  - 주어진 환경에서 수행해야할 액션의 representation을 학습하는 방법론
  - RL은 사전에 정의된 Reward를 필요로 하며, 매우 많은 interaction을 필요로 함

Turing Lecture

Section 3. Recent Advances
(Present : 2017 ~ 2021)

3) Representation Learning

정의: 데이터의 representation을 주어진 task를 해결하는데 적절한 형태로 Mapping하는 방법론

종류

- Supervised Learning
  - 많은 Task에서 매우 성공적인 결과를 생산
  - 그러나, 많은 양의 human-labeled data를 필요로 함
- Reinforcement Learning
  - 주어진 환경에서 수행해야할 액션의 representation을 학습하는 방법론
  - RL은 사전에 정의된 Reward를 필요로 하며, 매우 많은 interaction을 필요로 함

공통문제

상기 Learning 방법론은 모두 **Task-Specific, Specialized 시스템을 산출**함

아주 narrow한 **학습된 환경/패턴을 벗어나면 쉽게 고장** 남

Turing Lecture

Section 3. Recent Advances
(Present : 2017 ~ 2021)

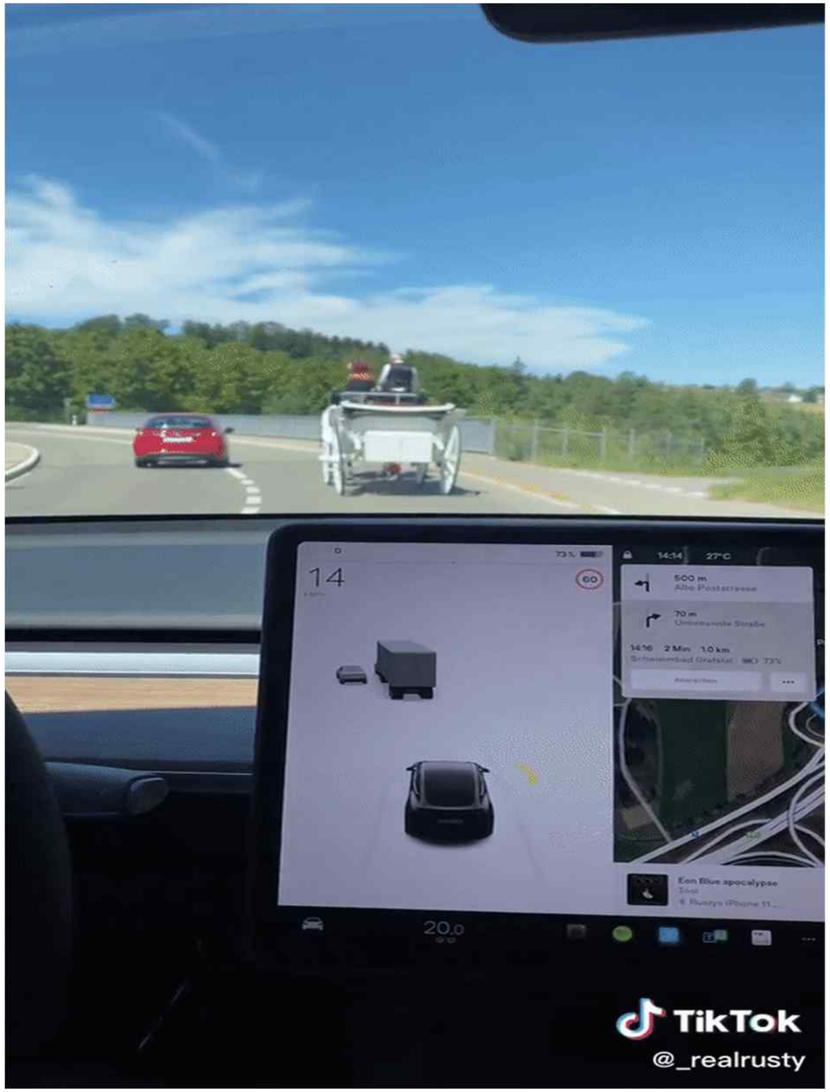
3) Representation Learning

테슬라 AI를 통해 문제점 확인

- 상황
  - ➢ 앞에 [마차 + 사람]이 있음
  - ➢ 테슬라 AI는 마차 + 사람 이라는 데이터를 거의 본 적이 없음. 실제로도 매우 적음
- 마차 + 사람에 대한 인식/추적 모두 고장 남

  인식: 트럭, 소형차, 사람, ... 등 인식 불가

  추적: 직진, 우회전 등 인식 불가

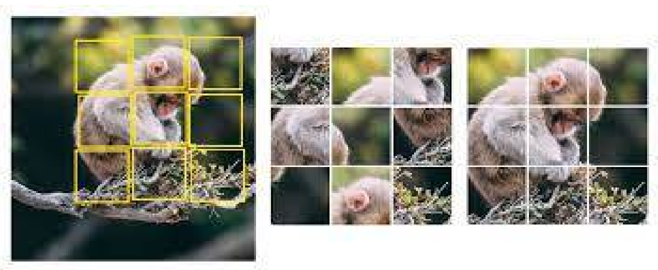


Turing Lecture

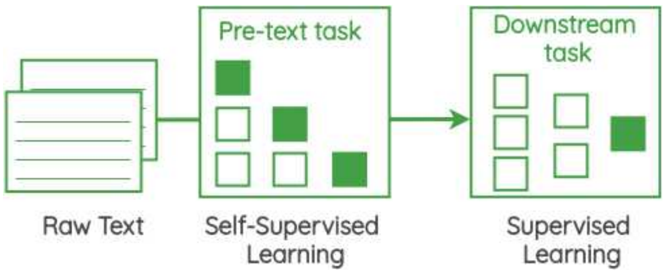
Section 3. Recent Advances
(Present : 2017 ~ 2021)

3) Representation Learning

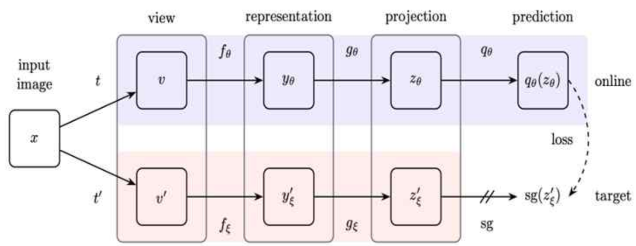
- 상기 언급된 문제를 해결하는 방식으로 발전을 해옴
  - ➢ Unsupervised/Self-Supervised Learning
  - ➢ Contrastive Self-Supervised Learning



Jigsaw self-supervised learning



BERT self-supervised learning



BYOL

Turing Lecture

Section 3. Recent Advances
(Present : 2017 ~ 2021)

3) Representation Learning

Representation Learning 중 SSL이 중요한 이유

1. Human-labeled sample을 줄이자 (less-supervised)
2. 하나의 task를 배우는데 필요한 interaction을 줄이자 (less-interaction)
3. out-of-domain에 robust한 모델을 만드는 것은 실생활적 관점에서 '매우' 중요하다

Section 4. The Future of Deep Learning
('22 ~ Future)

Turing Lecture

Section 4. The Future of Deep Learning (Future)

1) Less Supervised Learning

2) Robustness against distribution changes

3) Beyond perception tasks: Reasoning

Turing Lecture

Section 4. The Future of Deep Learning (Future)

1) Less Supervised Learning

대가들이 생각하는 Less Supervised Learning의 연구 방향

- 어떻게 NNs이 인간처럼 생각하게 할 것인가?
- A key question for the future of AI is how do humans learn so much from observation alone?

  → 이전에는 AI를 도구로서 많은 문제를 해결함. 이제는 인간처럼 생각하게/배우게 만들고 싶은 것

Yann LeCun and Yoshua Bengio: Self-supervised learning is the key to human-level intelligence

- Venture Beat Interview (2020)

2) Robustness against distribution changes

- 학습 데이터의 분포와 테스트 데이터의 분포가 달라도 잘 동작하는 우수한 모델
- Domain Adaptation or Generalization

얼굴탐지

보행자 탐지

Turing Lecture

Section 4. The Future of Deep Learning (Future)

2) Robustness against distribution changes

- 여러 Task를 배워도 성능이 우수한 모델 연구
  - ➢ Task1, Task2, Task3 을 학습한 이후 Task4 에 대해서 잘 할 수 있는지

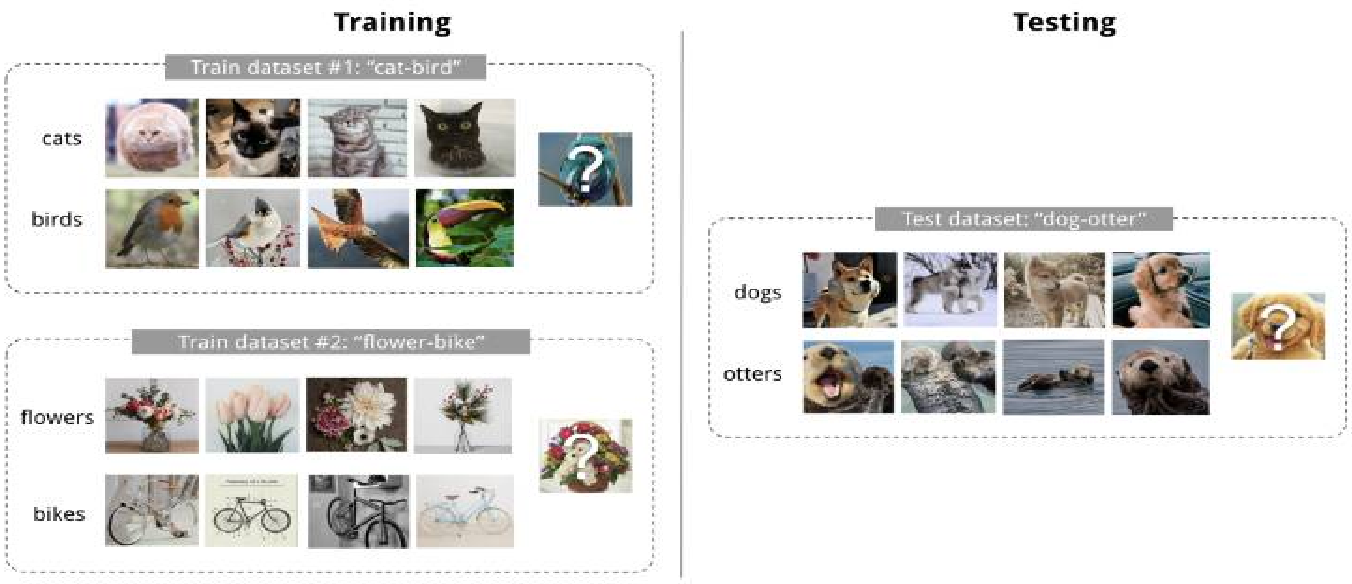


Fig. 1. An example of 4-shot 2-class image classification. (Image thumbnails are from Pinterest)

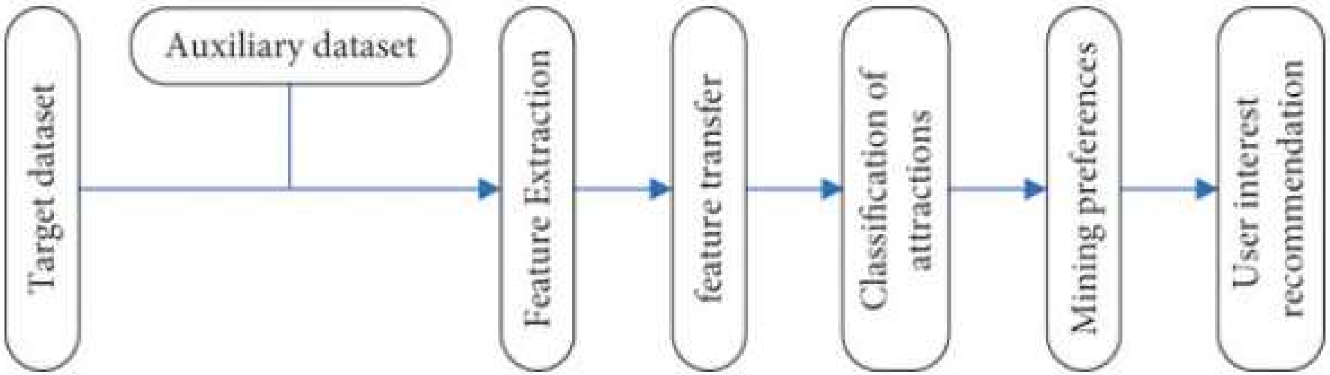
Turing Lecture

Section 4. The Future of Deep Learning (Future)

2) Robustness against distribution changes

- Recommendation을 포함한 모든 부분에서 생각해 볼 필요가 있음

➢ 'Personalized recommendation of attractions based on domain adaptation'



Turing Lecture

Section 4. The Future of Deep Learning (Future)

3) Beyond perception tasks: Reasoning

Perception vs. Reasoning

- Perception: 주어진 Task내에서 객체/정보를 인식 or 분류하는 것
- Reasoning: 주어진 Task에 대해서 (Event) 에 대한 해석, 논리적인 추론

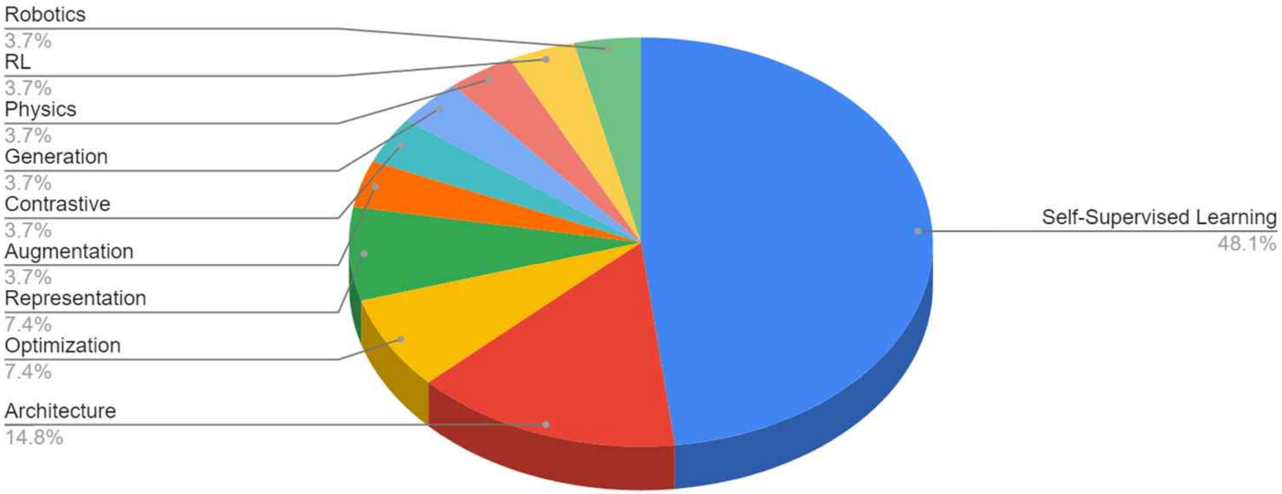
Ex) 미적분학 풀기, Choice of Plausible Alternatives (COPA), Symbolic AI research

- Premise: 남자의 발가락이 부러졌다. 원인이 무엇인가?
  Alternative 1: 양말에 구멍이 났다
  Alternative 2: 망치를 그의 발등 위에 떨어뜨렸다

Turing Lecture Additional Information

Yann LeCun

Total	27
Self-Supervised	13
Architecture	4
Optimization	2
Representation	2
Augmentation	1
Contrastive	1
Generation	1
Physics	1
RL	1
Robotics	1

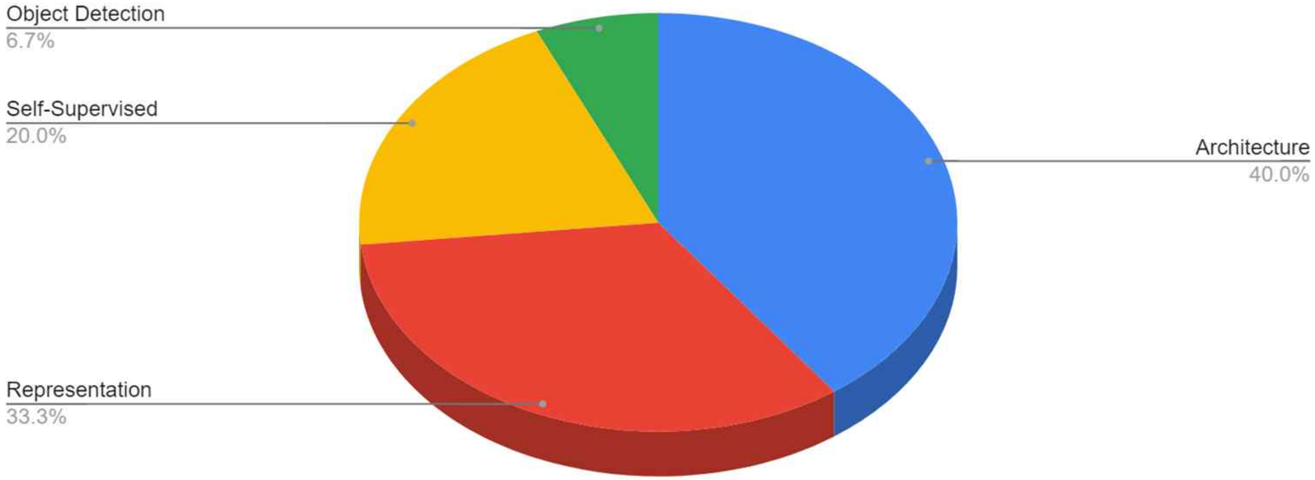


Google Scholar 기준 Yann LeCun의 '21~'22 년도 논문 통계

Turing Lecture Additional Information

Jeffery Hinton

Total	15
Architecture	6
Representation	5
Self-Supervised	3
Object Detection	1



Google Scholar 기준 Jeffery Hinton의 '21~'22 년도 논문 통계

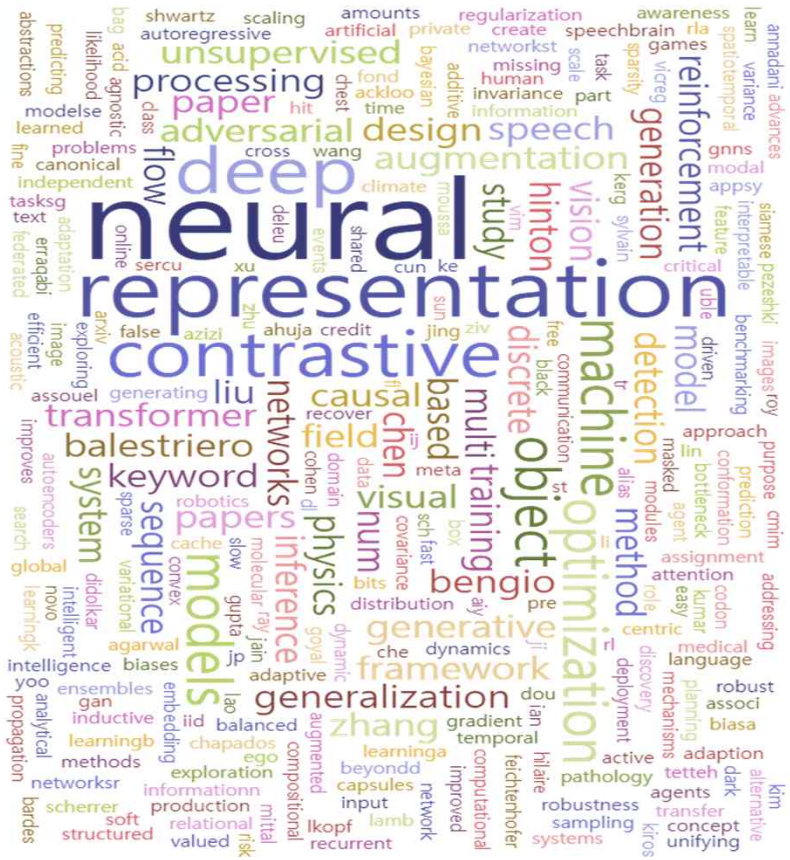
Turing Lecture Additional Information

Yoshua Bengio

- Total: 110 papers
- Tracking이 불가능해, 논문 제목 Word Cloud 수행

"Right now, the way we're teaching machines to be intelligent is that we have to tell the computer what is an image, even at the pixel level. For autonomous driving, humans label huge numbers of images of cars to show which parts are pedestrians or roads. It's not at all how humans learn, and it's not how animals learn. We're missing something big."

- Yoshua Bengio



Yoshua Bengio의 '21~'22 년도 논문 통계

The Core of the Turing Lecture

Conclusion

- 자신의 연구주제에 대해 [Less Supervised, Robustness, Reasoning] 관점에서 문제를 찾아보자
- 찾았다면 해결하고, 실생활에 적용하자

AI systems tends to take a hit when they go from the lab to the field

- Bengio, Hinton, Lecun