

인지과학으로 시작하는 인공지능(AI)

2022. 11. 21

정 준 수 Ph.D.

- | | | |
|------------------------------------|--------------------------|-----------------|
| 1. Human Vision vs Computer Vision | 10. Yolov5 custom object | 18. GAN |
| 2. Vision tasks | 11. AWS cli | 19. Ae-gan |
| 3. AWS Rekognition | 12. AWS SDK | 20. Dc-gan |
| | | 21. Transformer |
| 4. Color Space | 13. OpenCV | |
| 5. RGB vs HSV | 14. Edge detect | |
| 6. HSV 응용 | 15. Gray scale | |
| 7. Pretrained model | 16. CNN | |
| 8. yolov5 | 17. Transfer learning | |
| 9. Yolov5와 color space 응용 | | |

학습목표

인간능력 증강원리 기반의 인지과학기술을 응용하여 근본적이며 창의적인 해결방안 도출을 목표로 하는 문제해결 컨설팅 능력을 제고하며, 상상을 뛰어넘는 인지컴퓨팅(AI) 기술을 응용한 무인서비스, 자율주행 기술, 사물 인식기술을 포함한 표정 및 목소리 등 비정형 정보를 종합한 감정이나 상태 추론까지 가능한 로봇(챗봇) 및 추상적 개념을 이해하고 일반화 학습을 통한 예술적 감각 및 창의적 업무 수행을 기반으로 문제 해결 능력 향상이 목표임.

Evolution's Big Bang



[This image is licensed under CC-BY 2.5](#)



[This image is licensed under CC-BY 2.5](#)



[This image is licensed under CC-BY 3.0](#)

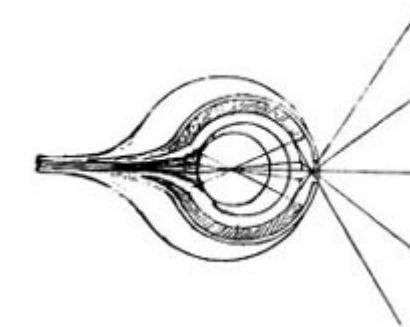
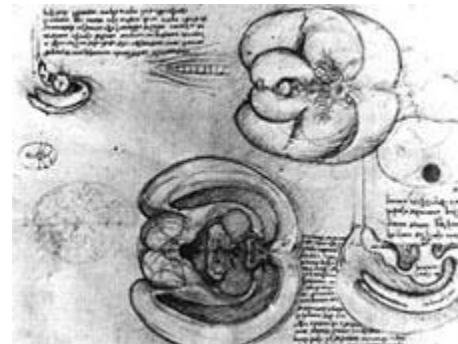
543 million years, B.C.

Men of vision

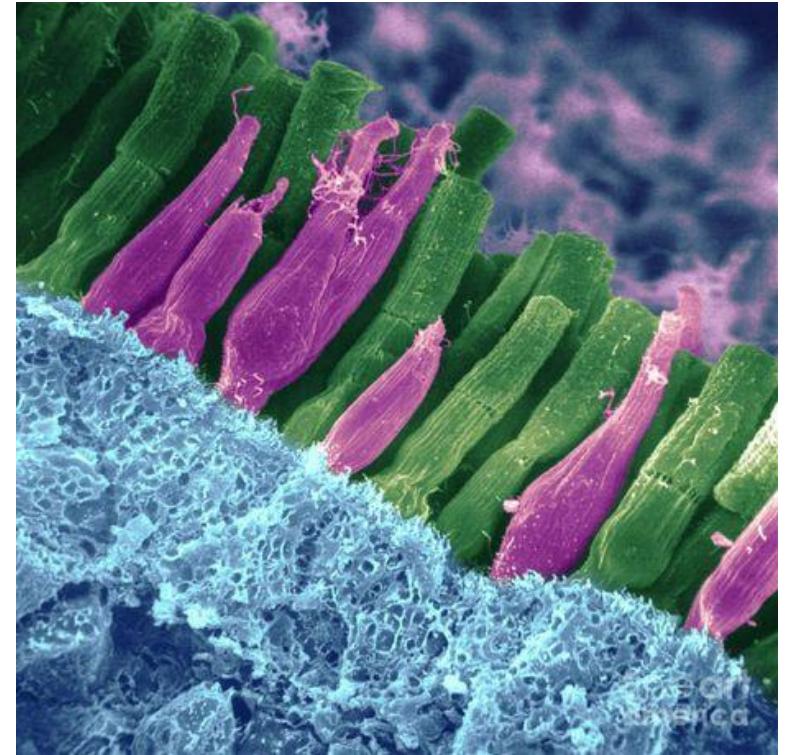
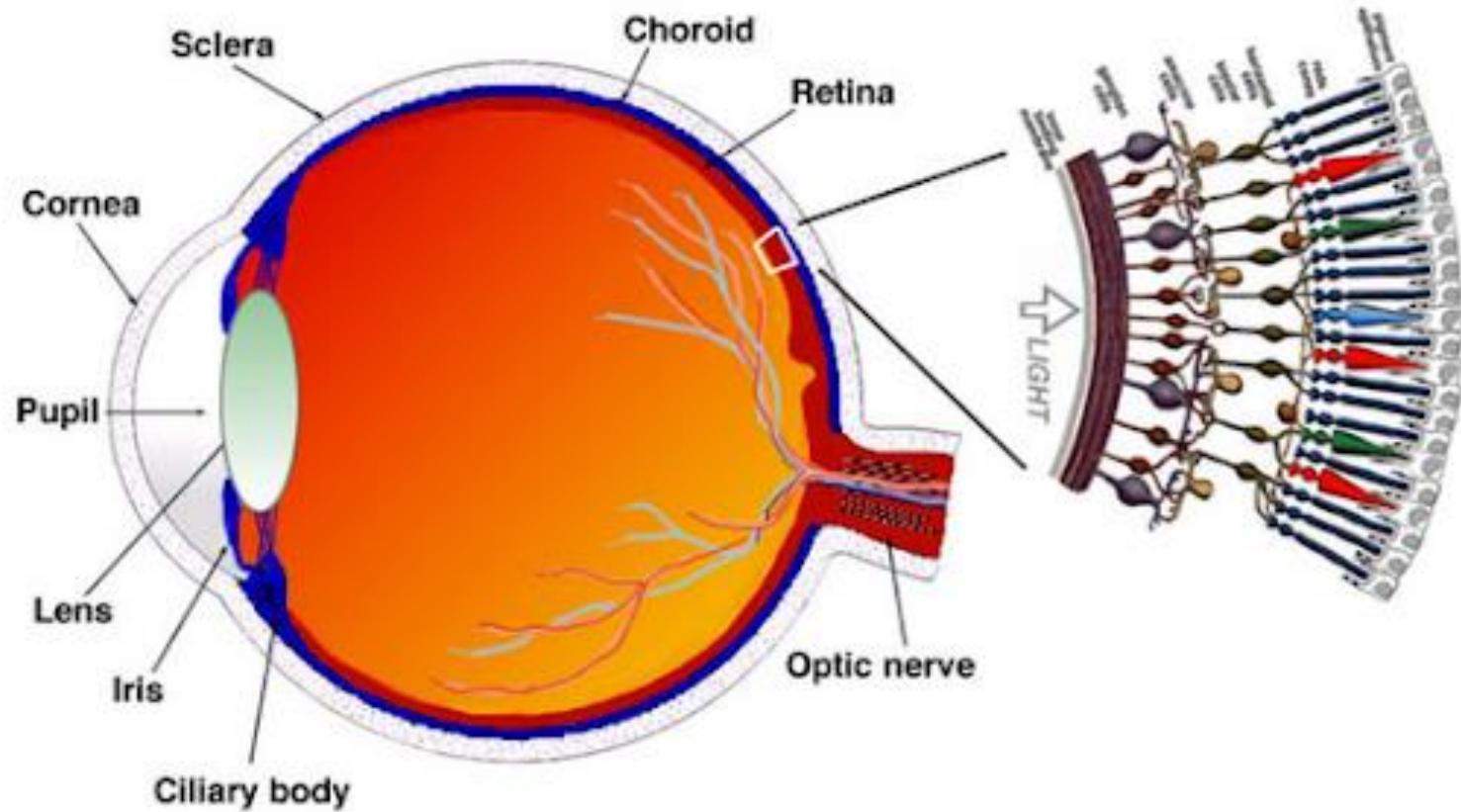
The study of vision has attracted many great thinkers over the centuries, not all of them scientists in the strictest sense, but we can call them all 'visionaries'

The Early Modern Period

Leonardo da Vinci proposed a theory that the inverted retinal image was re-inverted inside the brain. Below you can see two of Leonardo's drawings, dating from 1490 and 1506 respectively. the third image demonstrates his theory of the re-inversion of an inverted retinal image.



[출처] <https://www.college-optometrists.org/the-college/museum/online-exhibitions/virtual-eye-and-vision-gallery/men-of-vision.html>

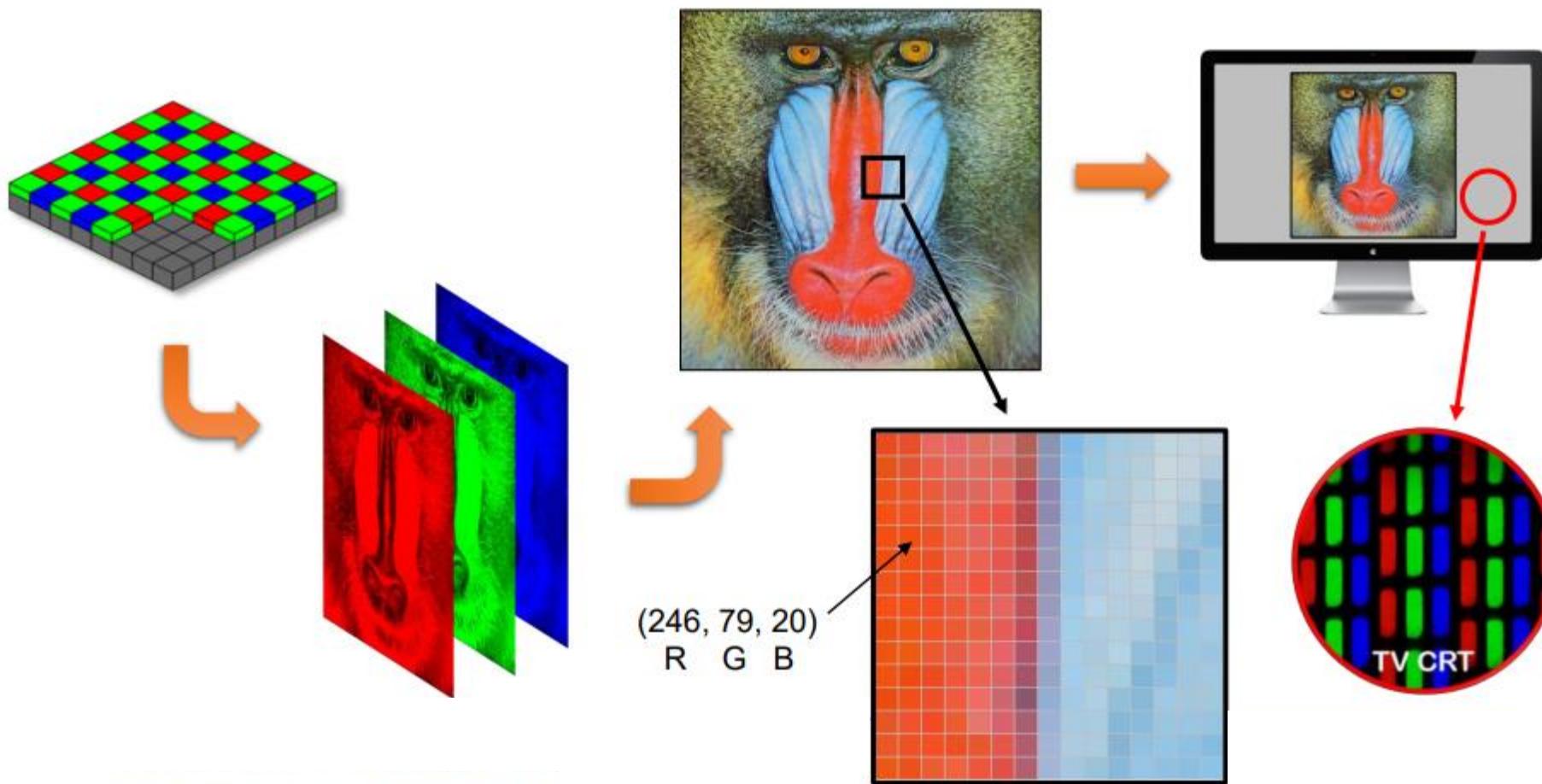


- 원추체 (색상)
- 간상체 (어둡고 밝) 물체의 모양

영상표현 방법

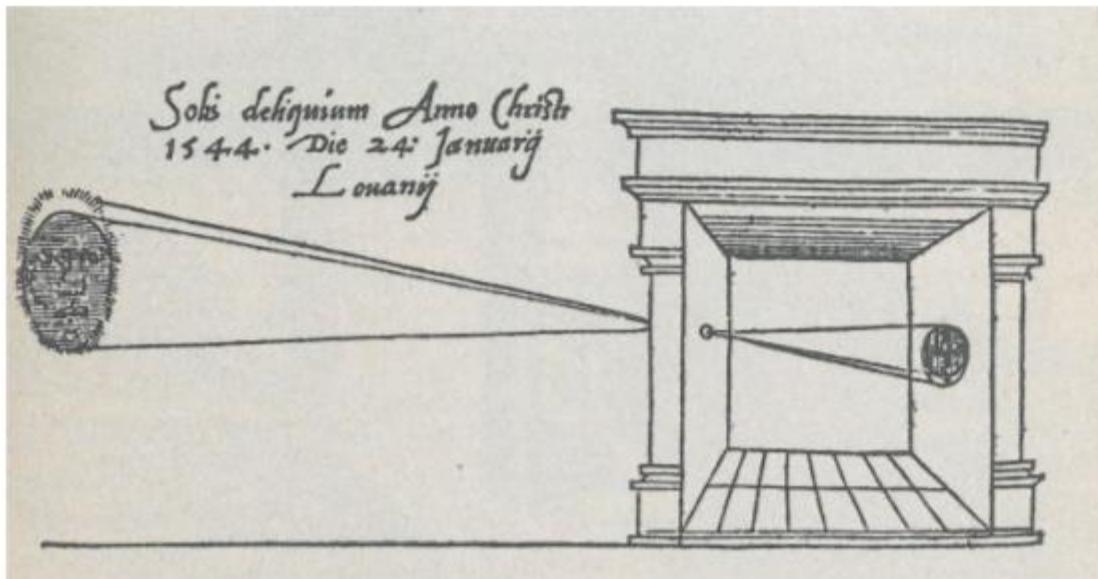
- 영상(image)이란?

- 픽셀(pixel)이 바둑판 모양의 격자에 나열되어 있는 형태 (2차원 행렬)

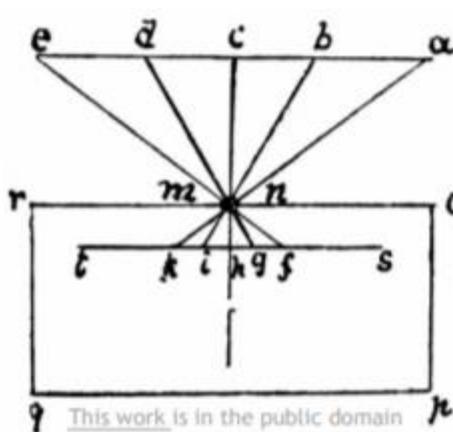
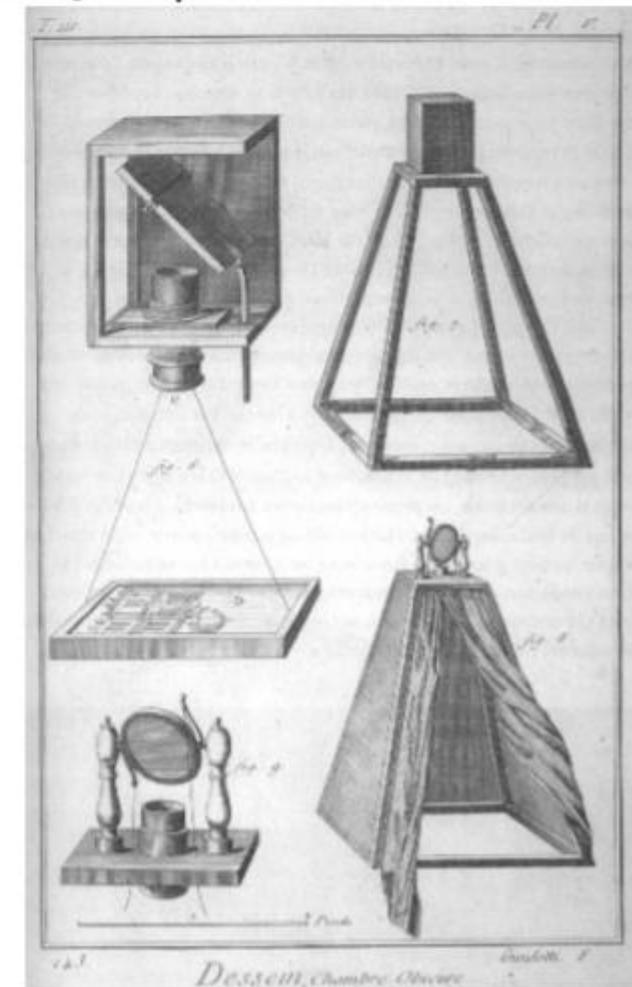


Camera Obscura

Gemma Frisius, 1545

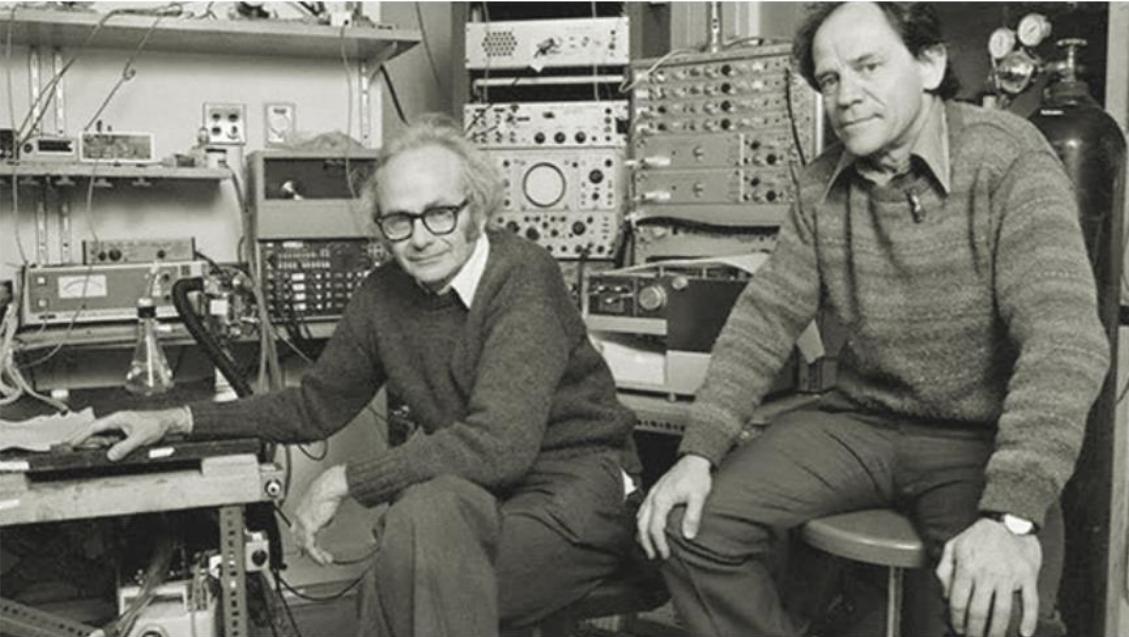


Encyclopedie, 18th Century



Leonardo da Vinci,
16th Century AD

A Nobel Partnership: Hubel & Wiesel

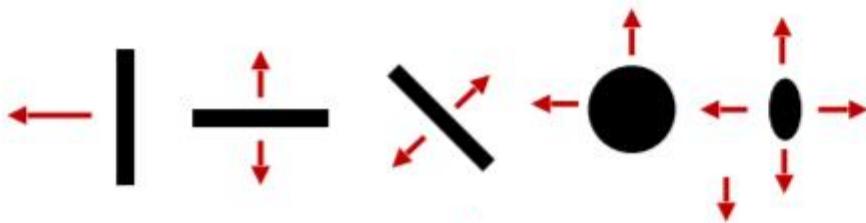


Hubel & Wiesel Come to Harvard

David Hubel and Torsten Wiesel came to Harvard from Johns Hopkins University with Steven Kuffler in the early 1960s to establish the Department of Neurobiology at Harvard Medical School. Their breakthrough discoveries about the visual system and visual processing earned them the Nobel Prize for Physiology or Medicine in 1981.

Hubel and Wiesel recorded electrical activity from individual neurons in the brains of cats. They used a slide projector to show specific patterns to the cats and noted that specific patterns stimulated activity in specific parts of the brain. Such single-neuron recordings were an innovation at the time, enabled by Hubel's earlier invention of a special recording electrode. They systematically created a map of the visual cortex with these experiments. The original film projector, light filters and slides, are held at the Warren Anatomical Museum at the Countway Library of Medicine.

<https://braintour.harvard.edu/archives/portfolio-items/hubel-and-wiesel>



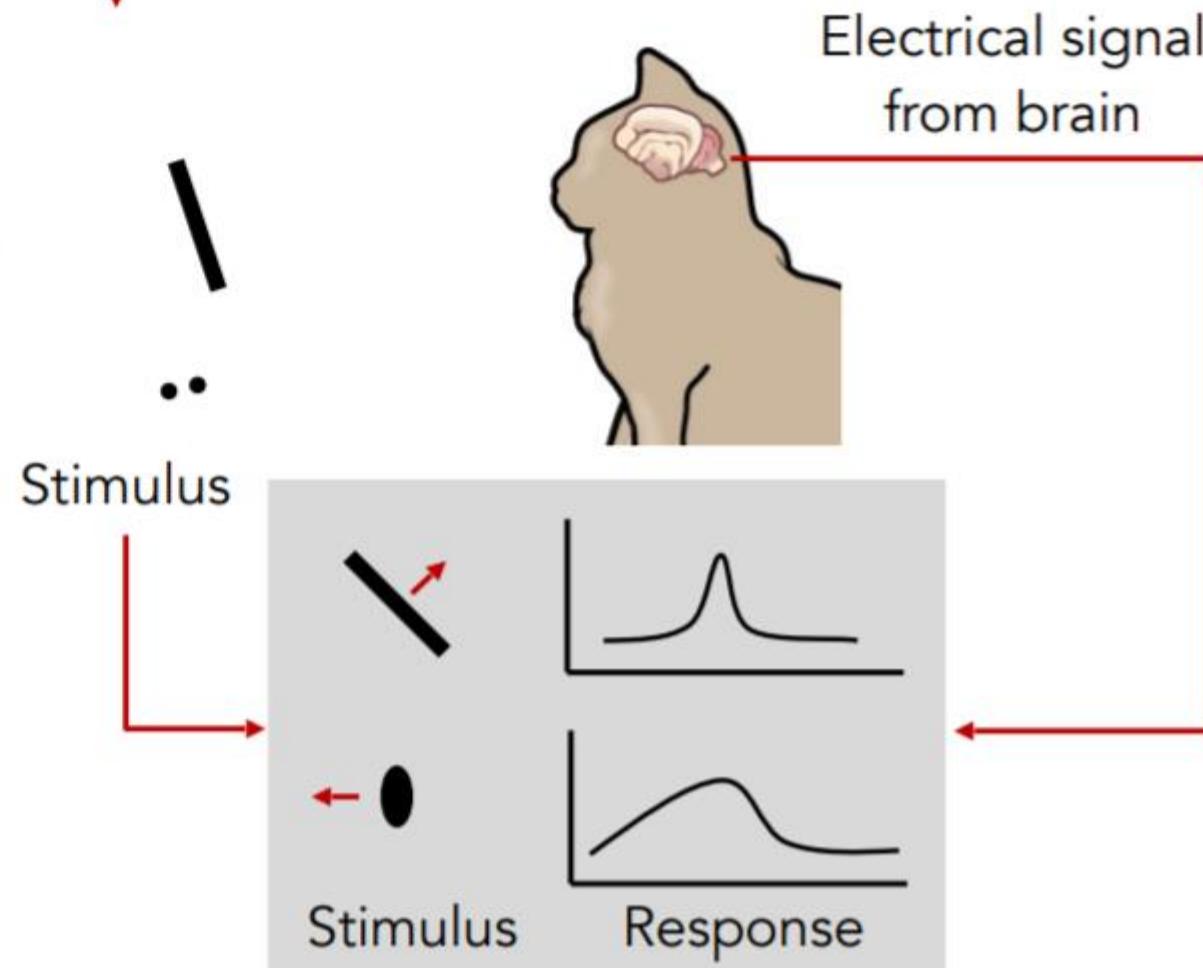
Simple cells:
Response to light orientation

Complex cells:
Response to light orientation and movement

Hypercomplex cells:
Response to movement with end point

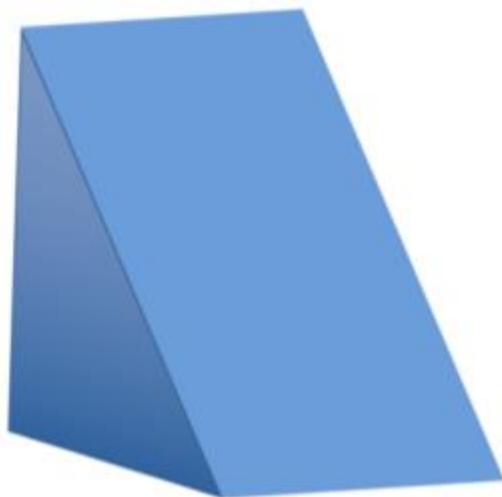


Hubel & Wiesel, 1959

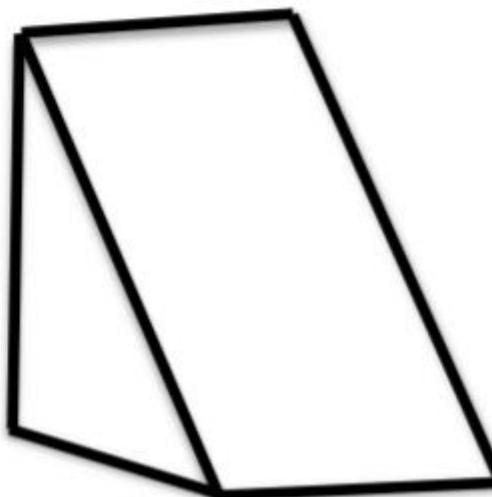


Block world

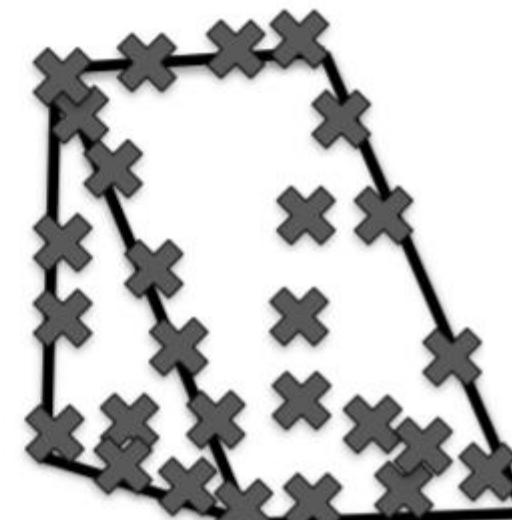
Larry Roberts, 1963



(a) Original picture



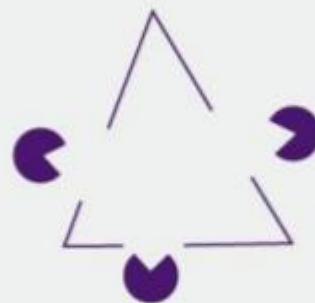
(b) Differentiated picture



(c) Feature points selected

Copyrighted Material

VISION



David Marr

FOREWORD BY
Shimon Ullman

AFTERWORD BY
Tomaso Poggio

Copyrighted Material

David Marr, 1970s

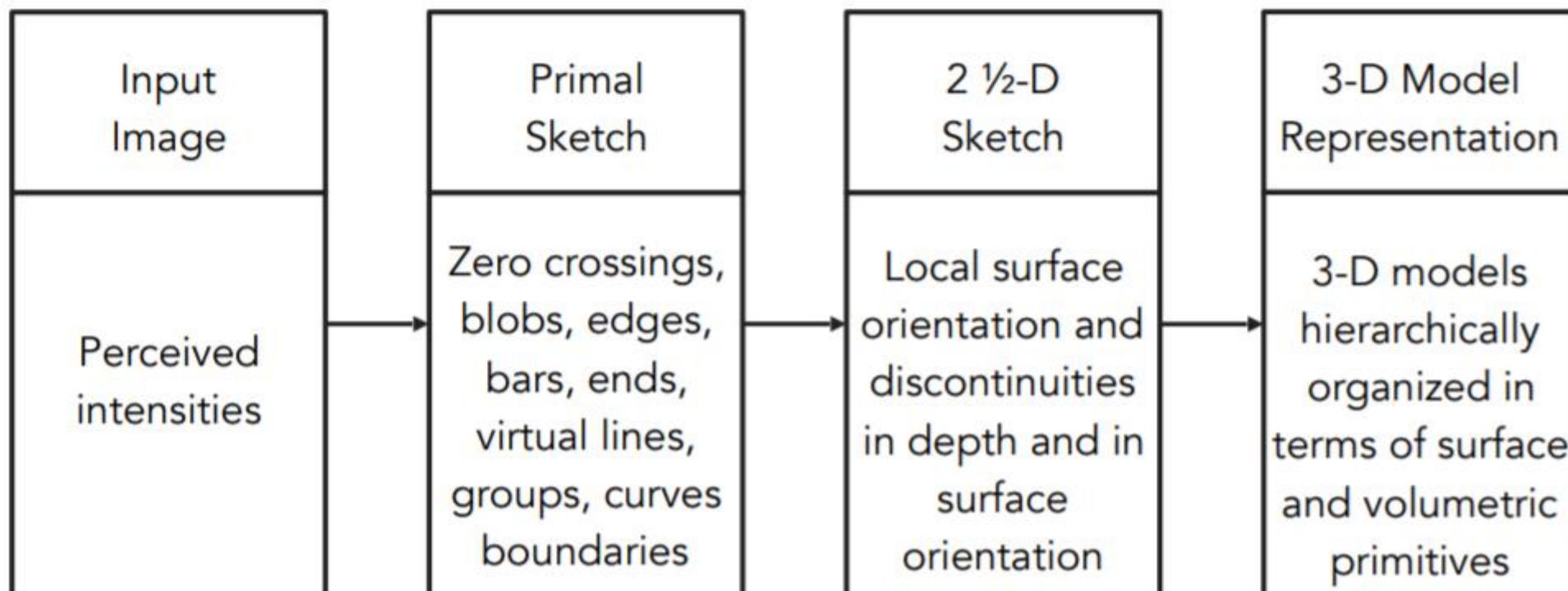
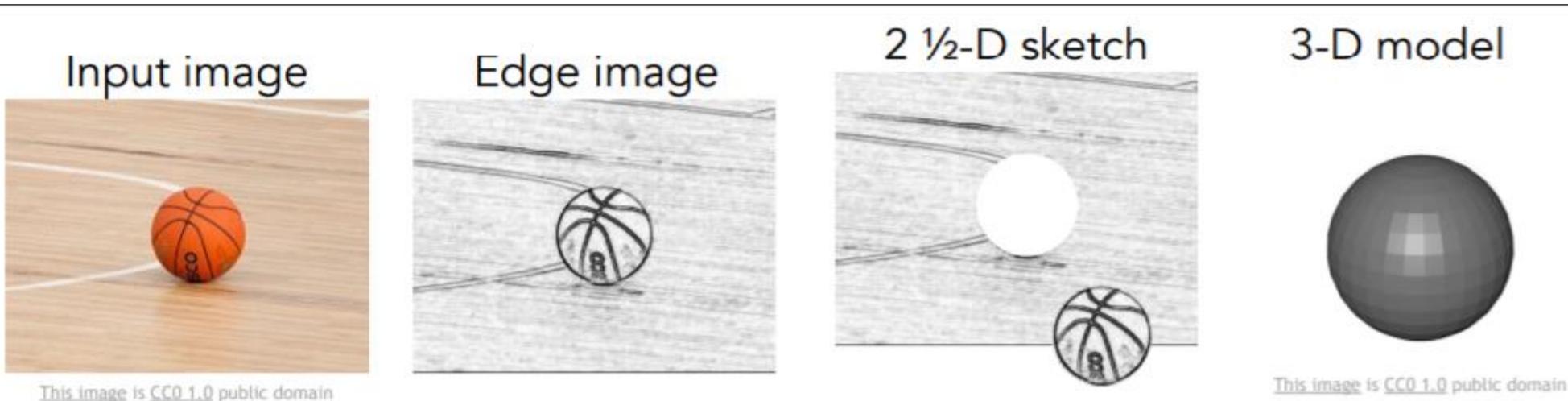
David Courtenay Marr (19 January 1945 – 17 November 1980) was a British neuroscientist and physiologist. Marr integrated results from psychology, artificial intelligence, and neurophysiology into new models of visual processing. His work was very influential in computational neuroscience and led to a resurgence of interest in the discipline.

인지심리학적 접근을 통해 새로운 시각적 대상재인 모형을 제시(Ellis & Young. 1996)

Marr(1982)의 시지각 이론을 바탕으로,

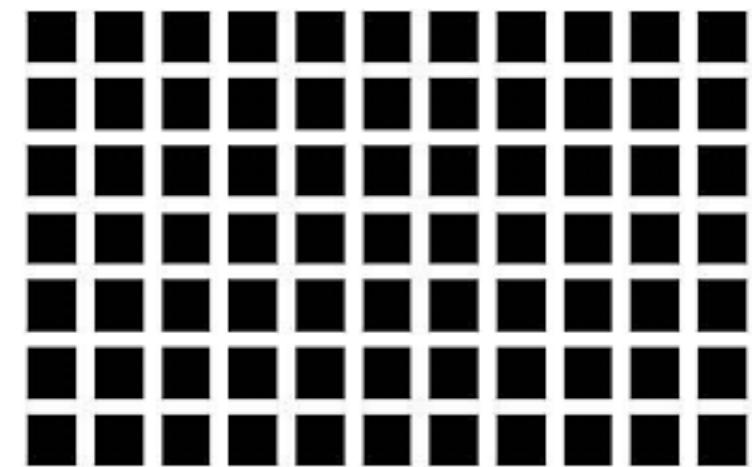
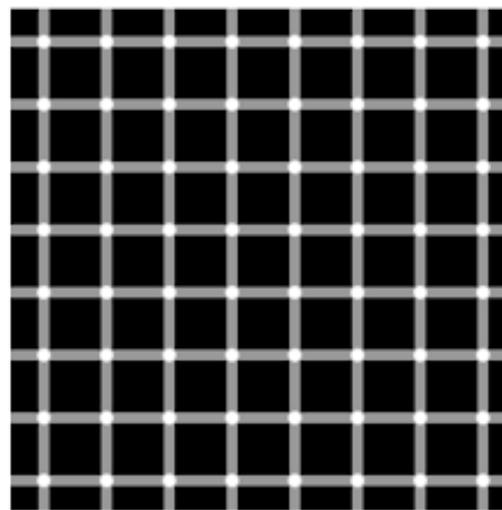
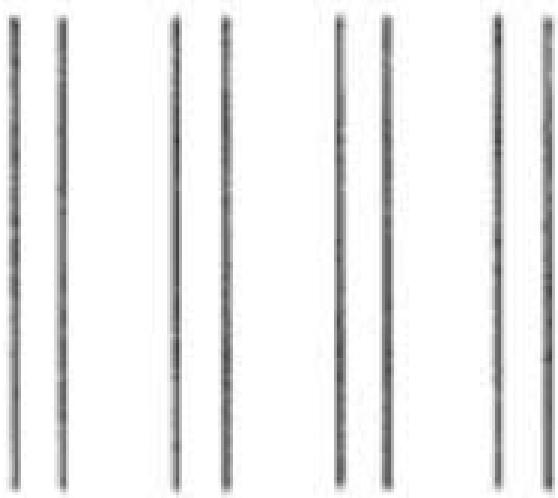
- 1) 망막에 맺힌 2D 영상으로부터 통합된 일차 스케치 생성(경계선, 윤곽, 명암 등)
- 2) 관찰자 중심의 관점에서 전체적인 형태, 윤곽 정보를 포함한 2.5D 영상의 생성
- 3) 대상항등성과 지각적 범주화를 포함한 3D표상의 생성(관찰자의 위치와 무관한 진정한 대상재인)

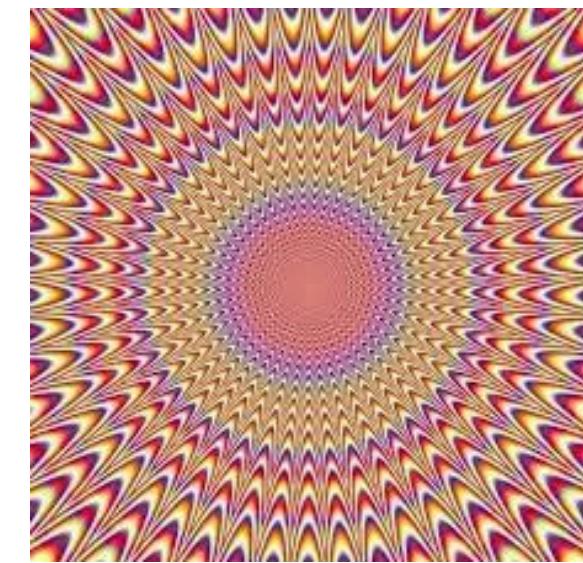
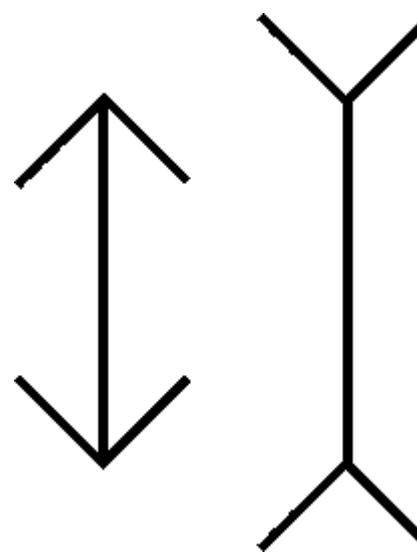
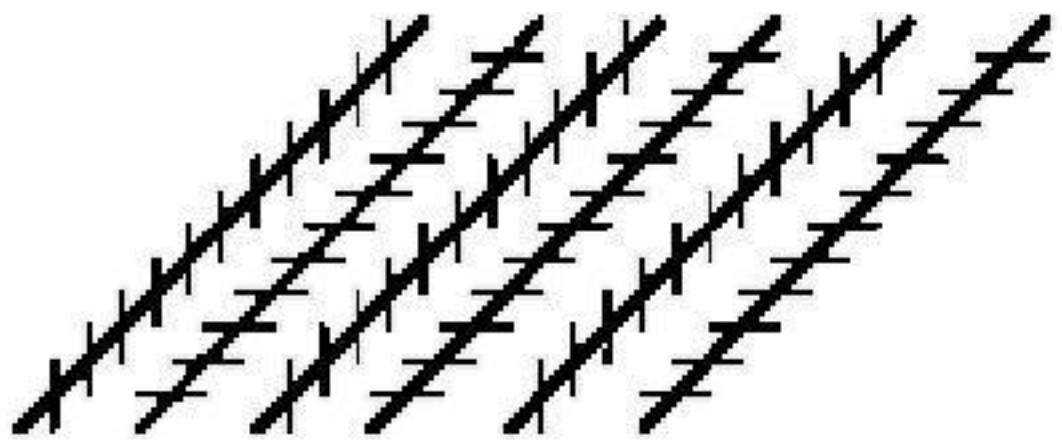
<http://www.aistudy.com/pioneer/Marr.D.htm->



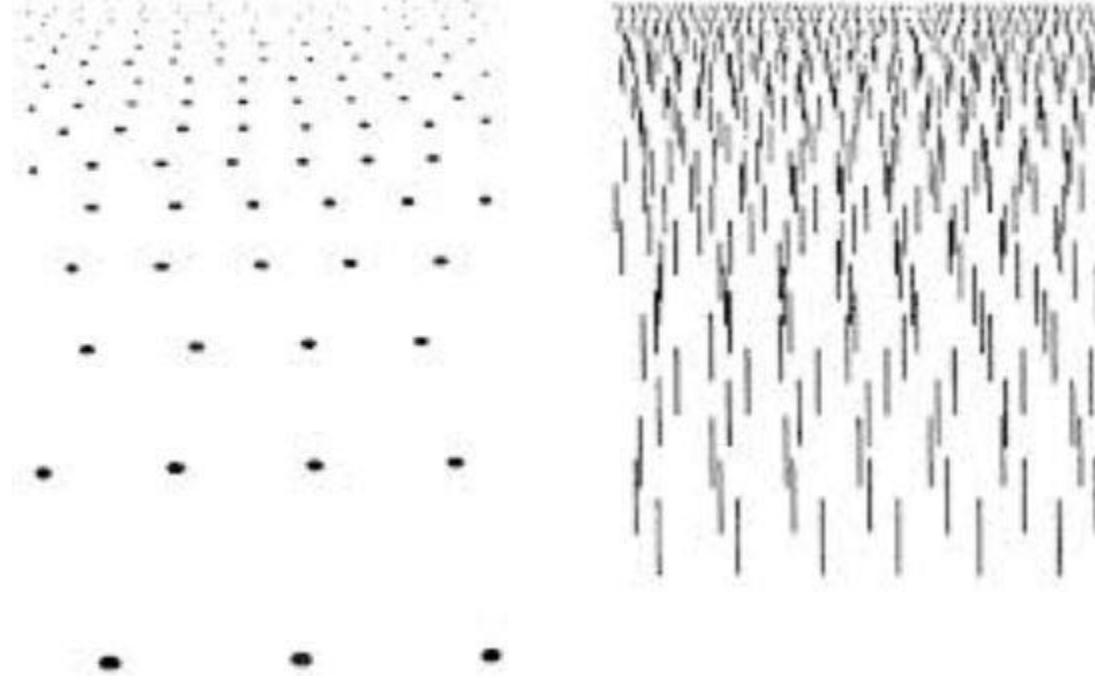
Stages of Visual Representation, David Marr, 1970s

착시 현상의 원인은?





깊이와 표면 지각



결기울기는 지각된 깊이에 대한 강력한 결정요인이며, 결기울기에서의 불연속은 다양한 결 표면들간의 공간적 관계에 대한 또 다른 정보를 제공한다. 따라서 왼쪽 그림의 결밀도 변화는 위로 올라가는 경사면의 느낌을 주고 오른쪽 그림에서의 변화는 갑작스러운 낙하, 즉 "시각절벽"의 느낌을 준다.

매우 강력한 일련의 깊이단서가 결기울기에 의해 제공되는데, 지각심리학에서 매우 영향력 있는 이론가인 James J. Gibson(1904~1979)은 결기울기의 중요성을 강조했다. 결기울기는 궁극적으로 조망에 의해 생성되며, 우리가 길에 깔린 자갈들을 보거나 초원에 널린 풀무더기들을 볼 때 어떤 것이 눈에 들어올까 생각해 보면 이해가 빠르다. Gibson은 이런 물체들이 망막에 투사된 상은 관계된 표면들의 공간적 배열에 따른 연속적인 변화 즉 결기울기를 반드시 보여야 한다고 지적함.

Yann André LeCun

(/la'kvn/ French pronunciation: born July 8, 1960) is a French computer scientist working primarily in the fields of machine learning, computer vision, mobile robotics, and computational neuroscience.

He is well known for his work on optical character recognition and computer vision using convolutional neural networks (CNN), and is a founding father of convolutional nets.

548

LeCun, Boser, Denker, Henderson, Howard, Hubbard, and Jackel

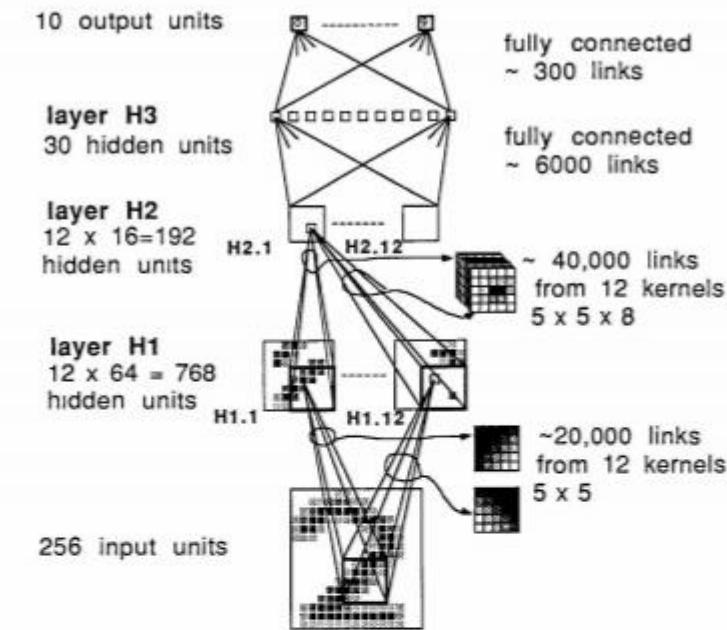


Figure 3 Log mean squared error (MSE) (top) and raw error rate (bottom) versus number of training passes

training set, 8.1% misclassifications on the test set, and 19.4% rejections for 1% error rate on the remaining test patterns. A full comparative study will be described in another paper

5.1 Comparison with Other Work. The first several stages of processing in our previous system (described in Denker *et al.* 1989) involved convolutions in which the coefficients had been laboriously hand designed. In the present system, the first two layers of the network are constrained to be convolutional, but the system automatically learns the coefficients that make up the kernels. This "constrained backpropagation" is the key to success of the present system: it not only builds in shift-invariance, but vastly reduces the entropy, the Vapnik-Chervonenkis dimensionality, and the number of free parameters, thereby proportionately reducing the amount of training data required to achieve a given level

시각작업기억의 처리 과정에는 기억 표상이 형성되고 유지되는 과정 뿐 아니라 기억된 항목에 대한 재인과 관련된 비교 과정이 포함된다(Hyun, Woodman, Vogel et al., 2006)

초기 시각 정보 처리

- 시상
- 물체지각
- 형태주의 체계화 원리 근접성의 원리 연속성의 원리 – 외측 슬상핵
- 2차원 그림 만들기
- 3차원 (깊이와 표면지각) 단안단서 결기울기 입체시 운동시차
- 비더만(Biederman) 물체재인 32개 components
- 형판맞추기

표상 (REPRESENTATION)

원래의 것과 같은 인상을 주는 이미지 또는 형상

- 정신적 표상은 정신 안에서 비교적 일관되게 재생산되는 의미 있는 사물이나 대상에 대한 지각을 일컫는다.
- 관념적 표상은 사고나 생각의 토대를 제공하는 정신적 표상으로서, 실질적으로는 정신적 표상과 동일하다.
- 본능적 표상은 자기 표상 안에 존재하는 개인의 욕동(원본능) 측면들을 말한다.

표상은 자아의 하부 구조를 구성하며 자아 내용물의 일부로 간주된다.

표상 학습(Representation learning)

표상 학습(representation learning) 혹은 특성 학습(feature learning)은 직접적인 데이터 대신, 유용한 정보를 더 쉽게 추출할 수 있게 만들어진 데이터의 표상(representation)을 통해 분류기나 다른 예측 기계를 학습시키는 것을 뜻한다.

확률적으로 좋은 표상은 입력을 설명할 수 있는 내재적인 설명 요인들의 posterior 분포를 포함할 때가 많다. 대화 인식, 신호 처리, 물체 인식, 자연어 처리와 같은 분야들에서 표상 학습 방법론은 실증적인 성공을 이루어 냈으며, 다양한 확률 모델과 인공 신경망, 그리고 딥러닝이 자동으로 표상 학습을 위한 특성 추출에 이용된다.

정보처리적 패러다임의 인지과학은 마음에 대한 보는 틀을 이와 같이 상정하고 나서, 정보처리체계로서의 마음의 작용을 감각, 지각, 학습, 기억, 언어, 사고, 정서 등의 여러 과정으로 나눈 다음, 각 과정에서 어떠한 정보처리가 일어나는가, 각 과정들은 어떻게 상호 작용 하는가를 묻고, 각 과정에서 어떠한 정보(지식)구조, 즉 표상(표현)구조가 관련 되는가를 규명하려 한다. 따라서 마음의 현상, 즉 심리적 사건은 정보의 내용 및 정보를 처리하는 사건으로 개념화 되어지는 것이다.

표상(표현: representation)의 강조 (Fodor, 1975)

인간과 컴퓨터가 자극 정보를 어떠한 상징으로 기억에 저장한다는 것은 자극 자체를 저장하는 것이 아니라 자극에 대한 표상(표현)을 저장하는 것이며 이는 마음과 컴퓨터 모두가 자극의 정보를 내적 기호(상징)로 변화시켜 기억에 보유한다는 것이다. 따라서 무엇을 안다는 것은 이들 **표상간의 연관을 찾거나 새로운 관계성을 만들어 낸다는 것을 의미한다.** 따라서 암의 과정에 대한 연구는 자극들이 어떻게 상징(기호) 표상들로 전환되고 또 활용 되는가를 연구하는 것이라 하겠다. 즉 **인지과학의 핵심 연구주제는 마음이나 컴퓨터에서의 표상의 처리과정(계산)과, 표상의 본질 및 그 구조적 특성의 연구라고 할 수 있다.**

핵심어 -> 연관성과 추상화

게슈탈트 심리학([독일어](#): Gestaltpsychologie, Gestalt 心理學)은 [심리학](#)의 한 학파이다. 인간의 정신 현상을 개개의 감각적 부분이나 요소의 집합이 아니라 하나의 그 자체로서 전체성으로 구성된 구조나 갖고 있는 특질에 중점을 두고 이를 파악한다. 이 전체성을 가진 정리된 구조를 독일어로 게슈탈트(Gestalt)라고 부른다. 박은정 교수에 따르면 이 단어의 의미는 전체 형태의 모양, 배열인데 지금 이 순간의 경험이다. 한편 형태심리학(形態心理學)을 나타내는 독일어이다.

HUMAN COGNITIVE NEUROPSYCHOLOGY

A Textbook with Readings

Andrew W. Ellis
and
Andrew W. Young

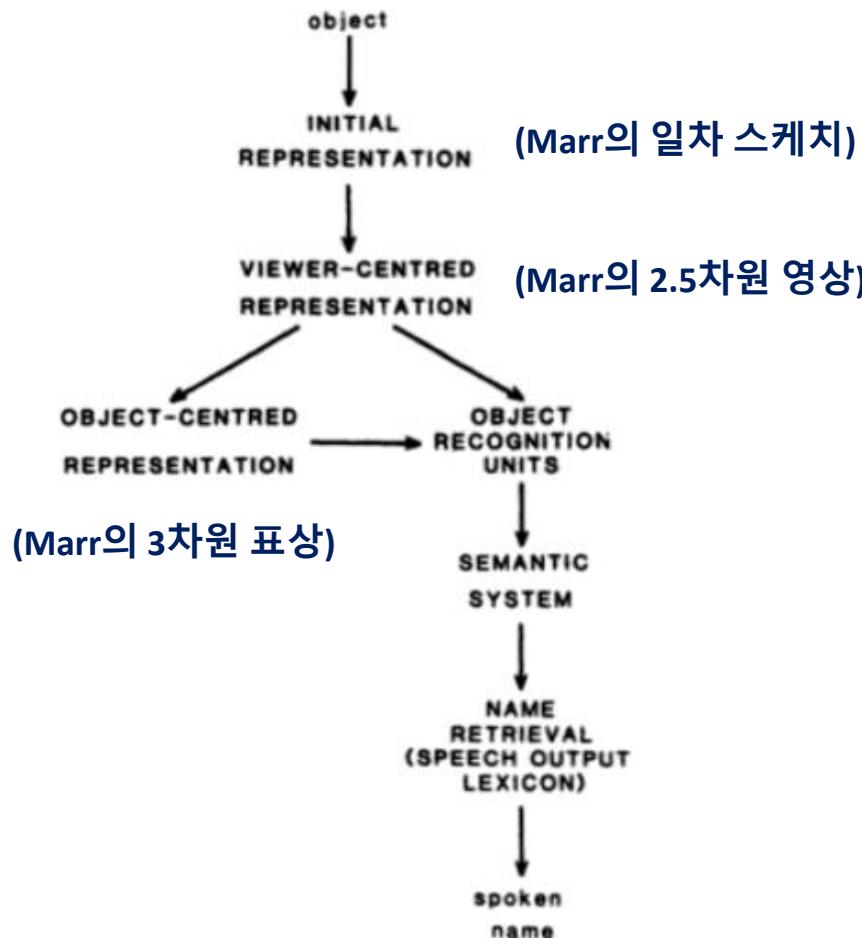


FIG. 2.1. Functional model for object recognition.

James Jerome Gibson (/ˈgɪbsən/; January 27, 1904 – December 11, 1979), was an [American psychologist](#) and one of the most important contributors to the field of [visual perception](#). Gibson challenged the idea that the nervous system actively constructs conscious visual perception, and instead promoted [ecological psychology](#), in which the mind directly perceives environmental stimuli without additional cognitive construction or processing.

생태학적 지각이론

생태학적 지각이론(Ecological Theory of Perception)은 김슨(James J. Gibson)에 의해 체계적으로 발전된 이론으로, 지각을 주변환경 안에서 일어나는 생존을 위한 적극적 행위로 인식한다.

즉 인간의 감각기관들을 외부의 자극에 대하여 수동적이기만한 수용기(受容器)로 보지 않으며, 정보탐색을 위한 적극적인 지각 시스템으로 본다.

이 이론은 인간과 환경이 서로 깊은 영향을 주고받으며, 상호보완하고 의존하는 모든 환경 내의 지각대상물들을 토탈 시스템의 차원에서 인식하는 시스템적 접근방식을 취한다.

AARON BEN-ZEEV*

J. J. GIBSON AND THE ECOLOGICAL
APPROACH TO PERCEPTION

"Whom he loves he rebukes" (*Proverbs*, 3, 12)

THE CONTRIBUTION of J. J. Gibson to psychology spans half a century; his first article was published in 1929, and his last book appeared in 1979. In this paper, I deal with his mature theory as it is expressed in his two last books: *The Senses Considered as Perceptual Systems* (1966), and *The Ecological Approach to Visual Perception* (1979); occasionally, earlier claims will be mentioned (especially from his book *The Perception of the Visual World* [1950]).¹ The aim of this paper is to examine the conceptual foundations, rather than the empirical claims of Gibson's view. Concerning the empirical level, Gibson's contributions to the understanding of perceptual processes are extremely significant, but this is not the place to describe them. His conceptual contribution has been less often evaluated, and there has been little agreement over such evaluation. I attempt, in this paper, to clarify some theoretical issues concerning his view.

Gibson's approach already has advocates who clarify and elaborate different aspects and implications of this approach. Due to limitations of space, I will not deal with their views separately, but merely discuss them when they clarify Gibson's own position.²

In the first section of this paper, I compare Gibson's approach to perception with the traditional ones; I indicate that Gibson suggests solutions to traditionally basic difficulties. Then I deal with some basic conceptual issues in

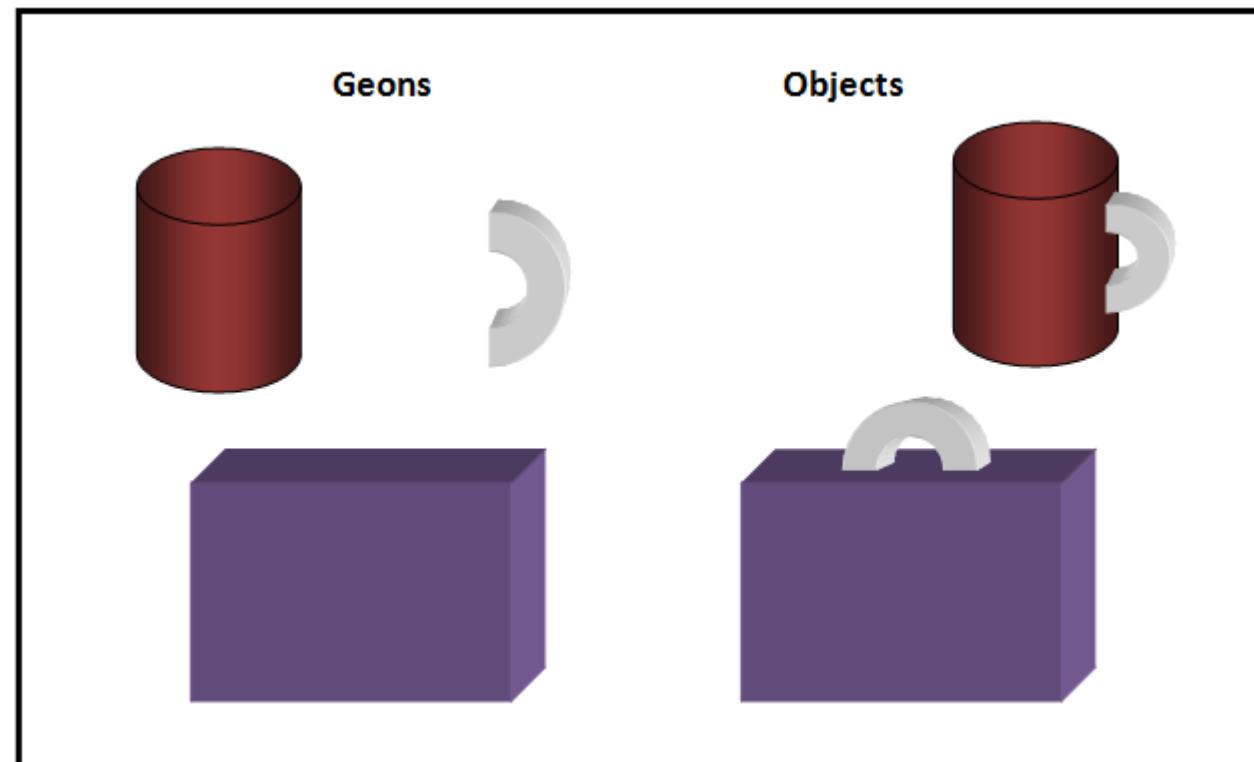
*Department of Philosophy, University of Haifa, Mt. Carmel, Haifa 31999, Israel.

'All these works published by Houghton Mifflin, Boston. These books will be referred to by date in the text. On the development of Gibson's ideas from 1929 to 1973 see T. J. Lombardo, *J. J. Gibson's Ecological Approach to Visual Perception: Its Historical Context and Development* (Unpublished doctoral dissertation, University of Minnesota, 1973).

²The papers most relevant for our discussion were written by R. E. Shaw and his colleagues: R. E. Shaw and J. Bransford, 'Introduction: Psychological Approaches to the Problem of Knowledge', *Perceiving, Acting and Knowing: Toward an Ecological Psychology*, R. E. Shaw and J. Bransford (Eds) (Hillsdale: Lawrence Erlbaum, 1977); R. E. Shaw, M. T. Turvey and W. Mace, 'Ecological Psychology: The Consequence of a Commitment to Realism', *Cognition and Symbolic Processes*, II, W. Weimer and D. Palermo (Eds) (Hillsdale: Lawrence Erlbaum, in press); R. E. Shaw and M. T. Turvey, 'Coalitions as Models for Ecosystems: A Realist Perspective on Perceptual Organization', *Perceptual Organization*, M. Kubovy and J. Pomerantz (Eds) (Hillsdale: Lawrence Erlbaum, in press); M. T. Turvey and R. E. Shaw, 'The Primacy of Perceiving: An Ecological Reformulation of Perception for Understanding Memory', *Perspectives on Memory Research*, L. G. Nilsson (Ed) (Hillsdale: Lawrence Erlbaum, 1979).

지각을 환경에 적응하려는 인간의 행태를 지원하는 기능으로 보며, 인간의 움직임에 따라 주변조건이 변화하는 역동적인 지각현상으로 해석하고, 인간과 환경간의 상호관계를 인간의 지각과 행동의 측면에서 설명한다. 즉, 생태학적 지각이론에서의 지각은 인간이 주변 환경에 적응하여 활동하기 위해 이루어진 환경과의 직접적인 관계이며, 주변 환경을 구성하는 구체적 사건, 사물 혹은 배치상태에 관한 정보를 환경으로부터 직접 습득하는 과정이다. 그리고 이러한 지각을 통해 얻는 정보의 핵심은 환경 내에서의 원활한 활동을 위한 수단을 얻고자 하는 것이며, 주변환경이 지각자에게 제공하는 행태지원성(affordance)이라는 것이다. 따라서 인간은 생태적 환경 내의 다양한 행태지원성을 지각하고 이러한 지각을 바탕으로 환경 내에서의 행위를 결정하게 된다. 즉, 지각은 행태를 지원하며, 행태는 지각에 필요한 정보를 제공하는 상호작용인 것이다. 이는 지각이 주변 환경으로부터 다양한 행태지원성에 대한 정보를 습득하는 것과 아울러 자신의 행동능력과 환경이 가지고 있는 행태지원성 사이에서 적합성을 찾아내려는 시도임을 의미한다. 생태학적 지각이론은 공간디자이너들로 하여금 어떠한 공간을 구성할 때, 그 기저에 시각적 복잡성과 같은 물리적 구성의 차원을 넘어 지원성과 같은 생태적 특성이 존재함을 인식시켰고, 이에 기반하여 인간과 환경 또는 공간의 상호관계에 적합성을 제공할 수 있는 디자인이 수행되어야 함을 이론적으로 뒷받침하였다. 또한 생태학적 지각이론은 인간행태를 일상적인 환경 하에서의 지각현상으로 설명하기 때문에, 환경디자인 실무에 적용 가능한 실제적 이론으로 발전되었다.

The **recognition-by-components theory**, or **RBC theory**,^[1] is a bottom-up process proposed by [Irving Biederman](#) in 1987 to explain [object recognition](#). According to RBC theory, we are able to recognize objects by separating them into **geons** (the object's main component parts). Biederman suggested that **geons** are based on basic 3-dimensional shapes (cylinders, cones, etc.) that can be assembled in various arrangements to form a virtually unlimited number of objects

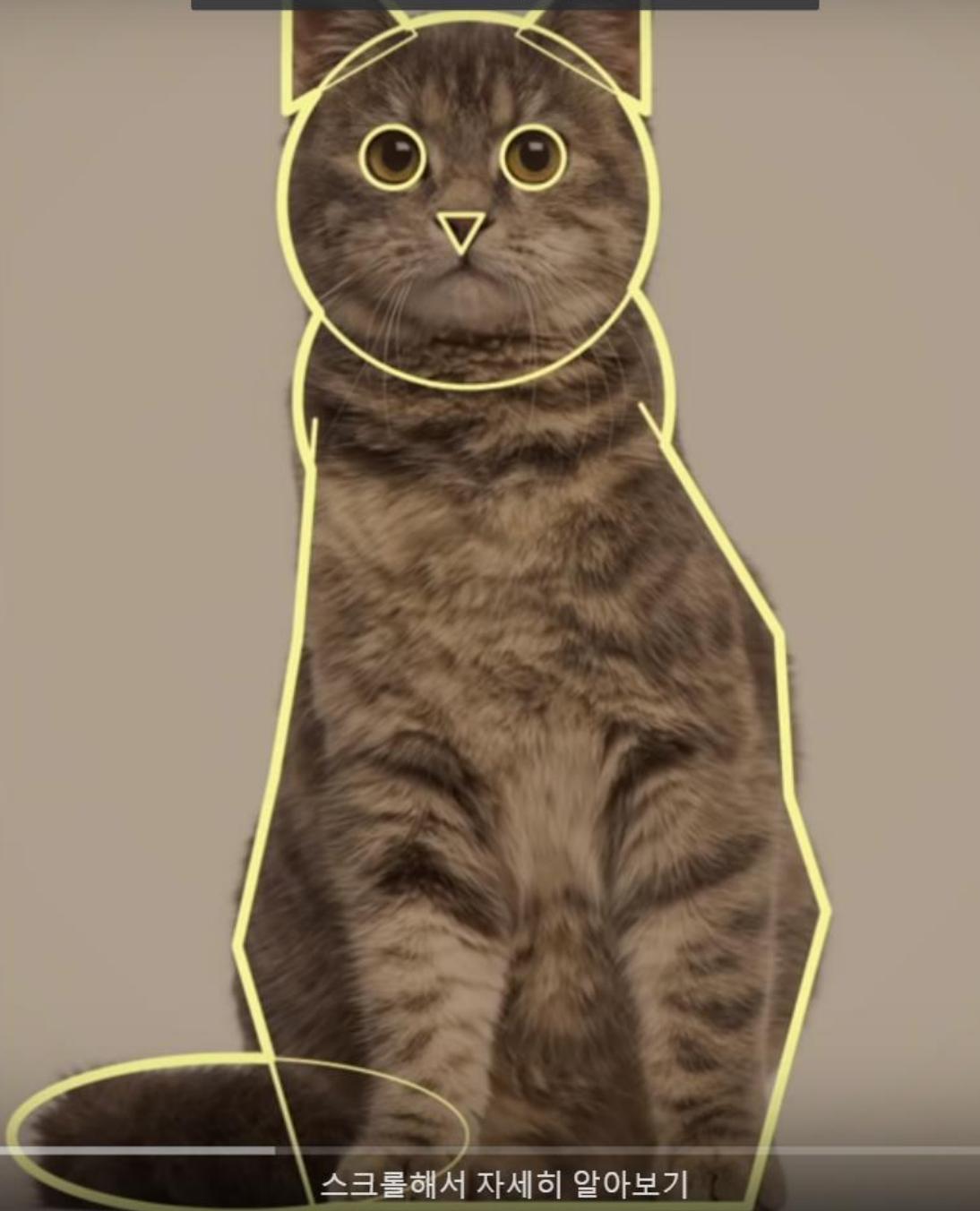


전체화면을 종료하려면 Esc 을(를) 누르세요.



스크롤해서 자세히 알아보기

전체화면을 종료하려면 Esc 을(를) 누르세요.



스크롤해서 자세히 알아보기

형태재인이란?

표상(Representation): 외부의 시각 정보를 받아 들여 어떠한 방식으로 우리 내부에서 재현

형태재인: 표상을 이미 내부에 저장되어 있는 시각적 사물들에 대한 기억이나 기준의 표상들과 대조하는 과정

- 자극 분석, 비교, 기업 탐색, 의사결정 등의 과정을 포함하는 복잡한 과정
- 시각, 청각 등의 여러 감각 정보에 대해 이루어질 수 있으며, 정보 처리도 2차원적인 것과 3차원적인 것 모두 일어날 수 있음

형태재인의 고전적 모형

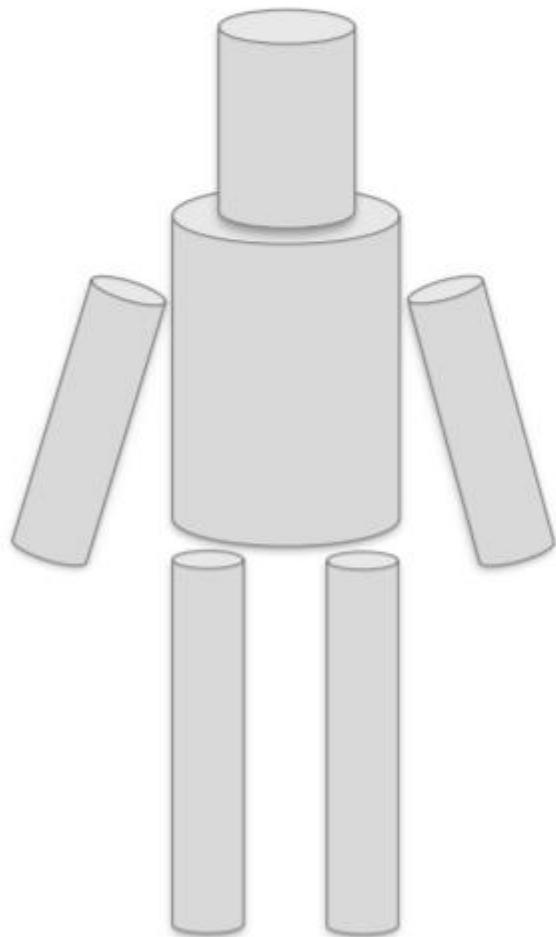
- 형판 맞추기(Template matching)모형
- 세부특징 분석모형(Feature detectors)

형태재인의 현재 모형

- 공간주파수 분석 접근
- 계산적 접근(Computational approach): David Marr의 시지각 계산모형

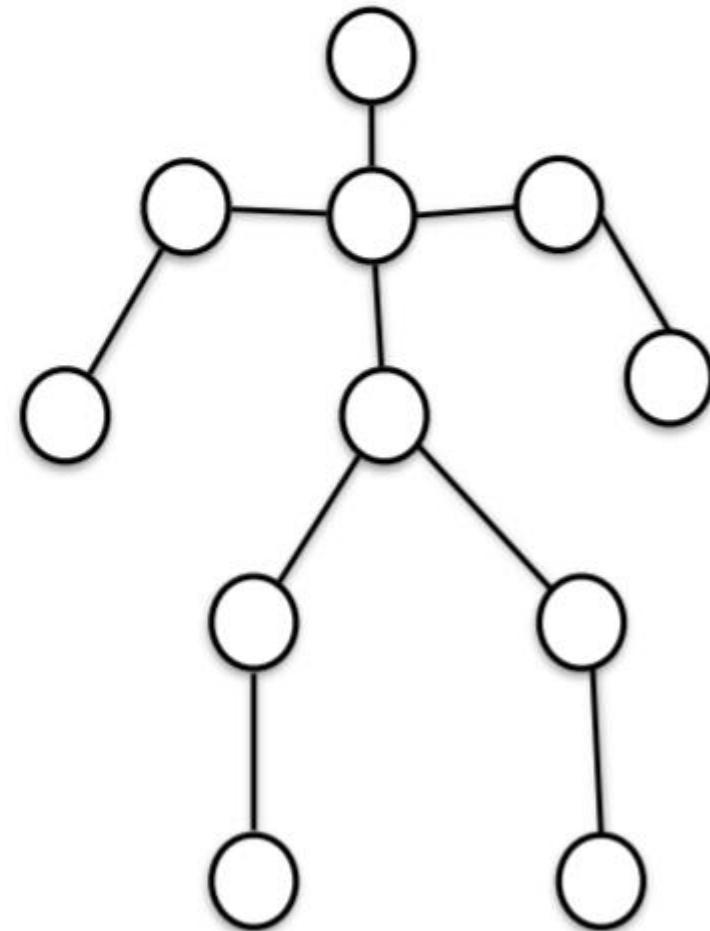
- Generalized Cylinder

Brooks & Binford, 1979

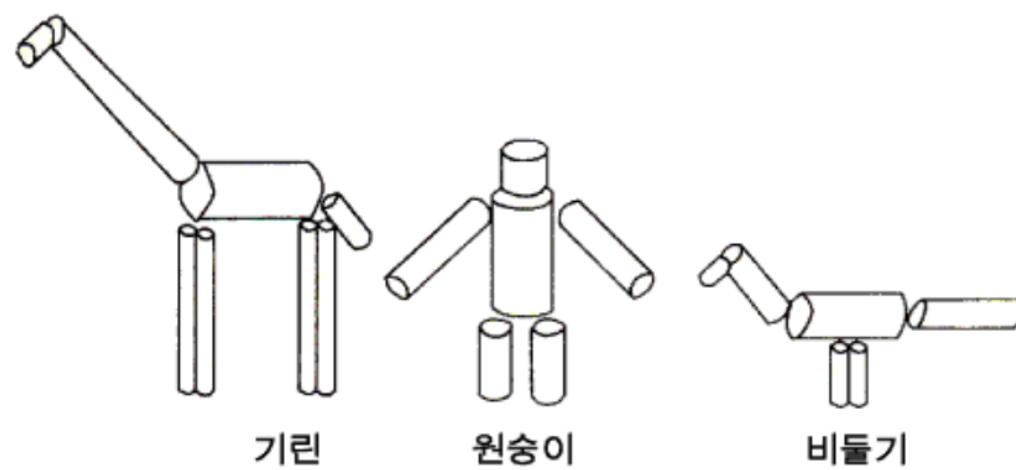
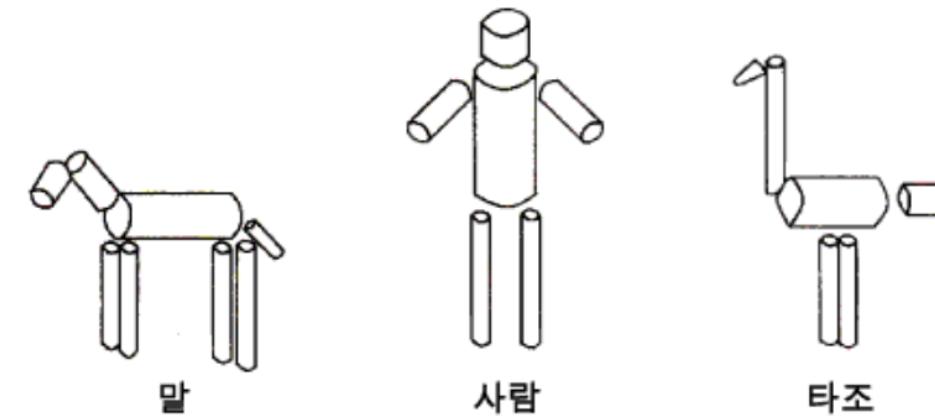


- Pictorial Structure

Fischler and Elschlager, 1973

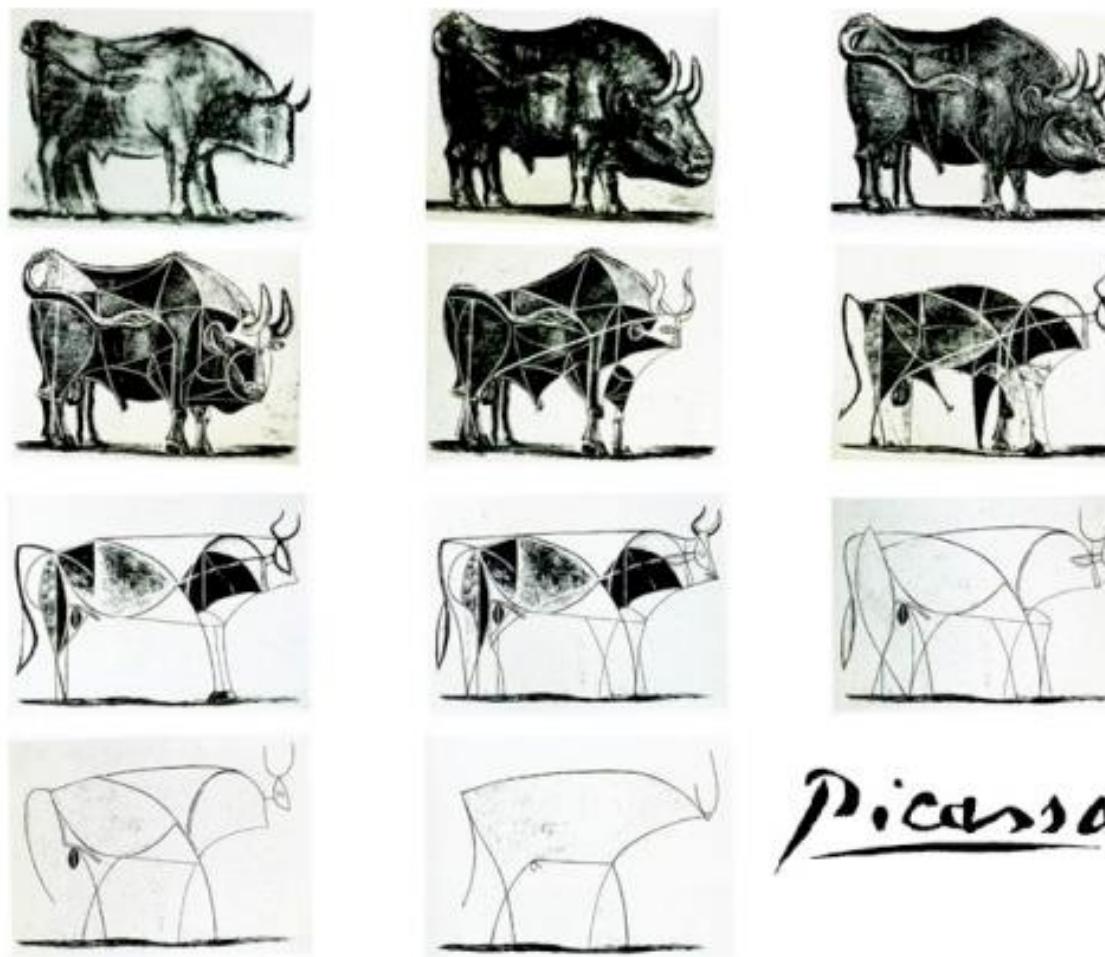


◆ 사람의 물체(시각) 재인 방식



친숙한 물체들을 기본 원통 형태로 분할한 것. (Marr & Nishihara. 1978)

◆ 추상화 (Abstract)



Pablo Picasso, Bull (plates I - XI) 1945



Image is CC BY-SA 4.0

David Lowe, 1987

Normalized Cut (Shi & Malik, 1997)

[Image](#) is CC BY 3.0



[Image](#) is public domain



[Image](#) is CC-BY SA 3.0





[Image is public
domain](#)



[Image](#) is public domain

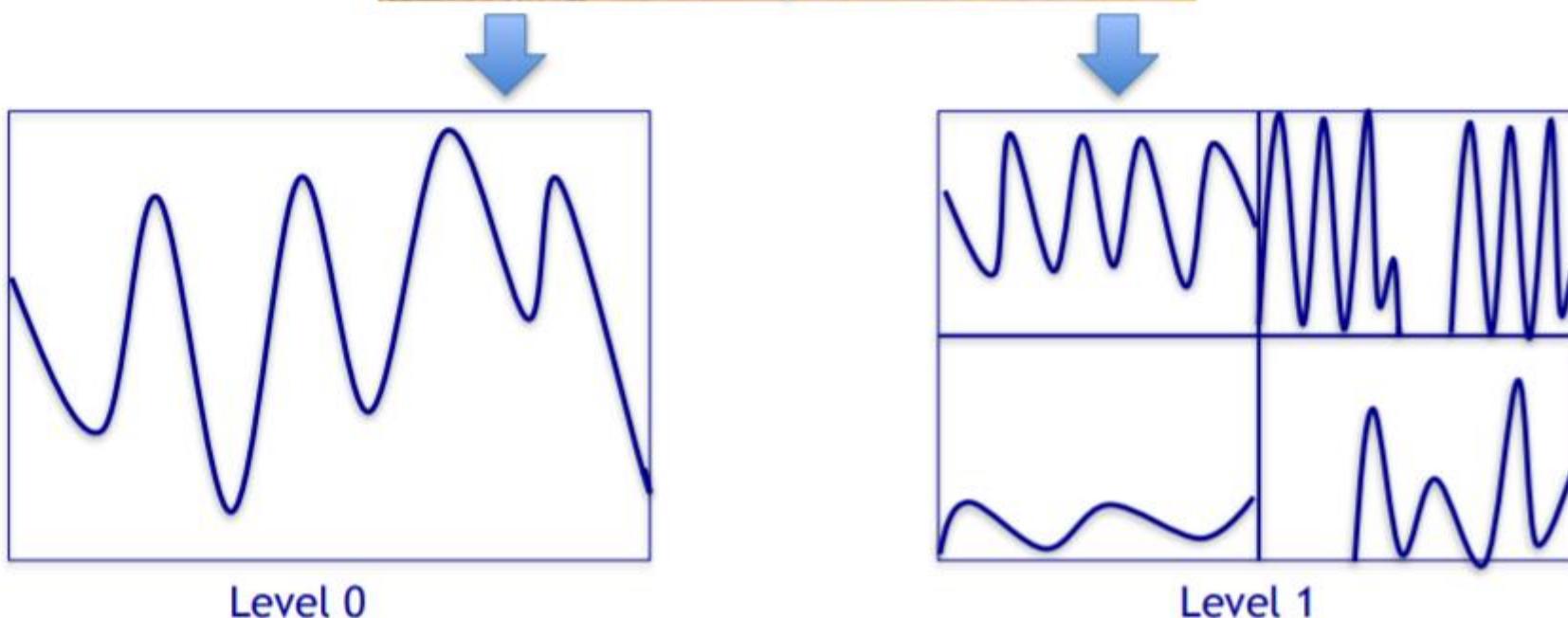


[Image](#) is CC BY-SA 2.0

"SIFT" & Object Recognition, David Lowe, 1999



Image is CC0 1.0 public domain

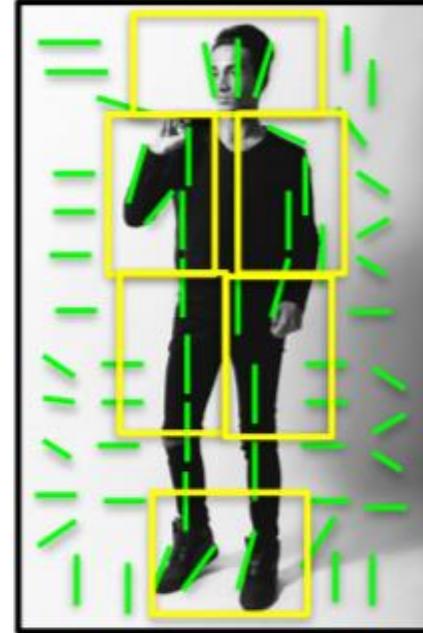
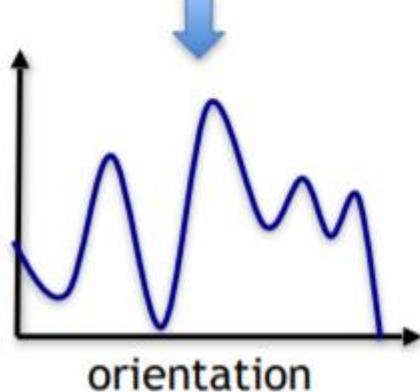


Level 0

Level 1

Spatial Pyramid Matching, Lazebnik, Schmid & Ponce, 2006

<https://www.youtube.com/watch?v=hPCTwxF0qf4>



Deformable Part Model
Felzenswalb, McAllester, Ramanan, 2009

Histogram of Gradients (HoG)
Dalal & Triggs, 2005

PASCAL Visual Object Challenge

(20 object categories)

[Everingham et al. 2006-2012]

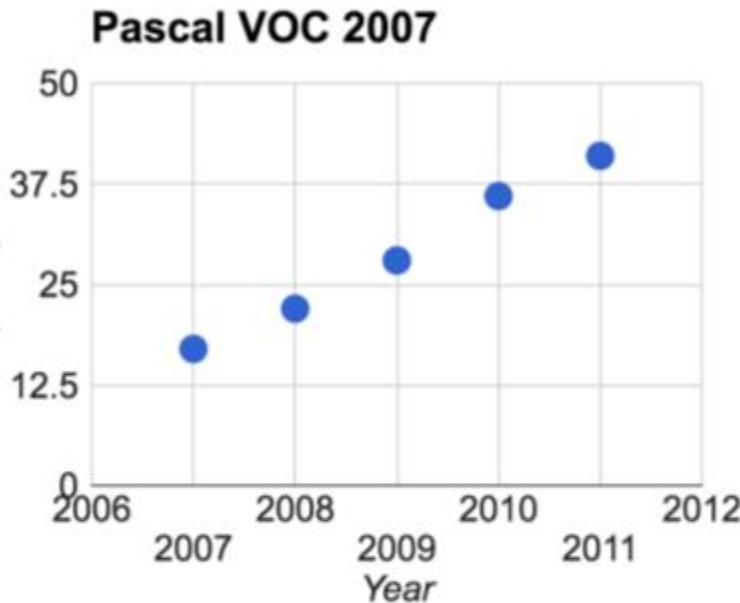
[Image](#) is CC BY-SA 3.0



[Image](#) is CC0 1.0 public domain

[This image](#) is licensed under
CC BY-SA 2.0; changes made

Mean Average Precision
(mAP)





www.image-net.org

22K categories and **14M** images

- Animals
 - Bird
 - Fish
 - Mammal
 - Invertebrate
- Plants
 - Tree
 - Flower
 - Food
 - Materials
- Structures
 - Artifact
 - Tools
 - Appliances
 - Structures
- Person
- Scenes
 - Indoor
 - Geological Formations
- Sport Activities



IMAGENET Large Scale Visual Recognition Challenge

Steel drum

The Image Classification Challenge:

1,000 object classes

1,431,167 images



Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle

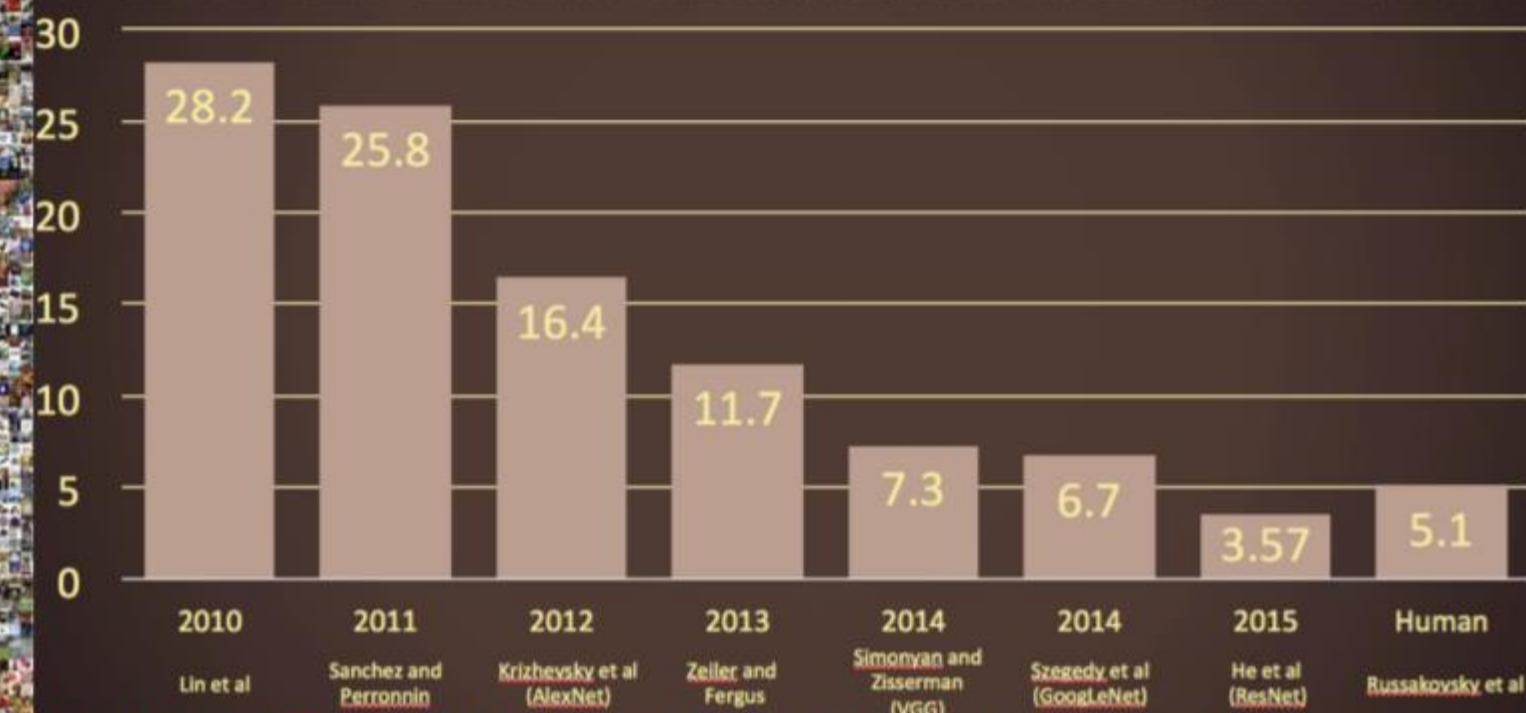


Output:
Scale
T-shirt
Giant panda
Drumstick
Mud turtle



IMAGENET Large Scale Visual Recognition Challenge

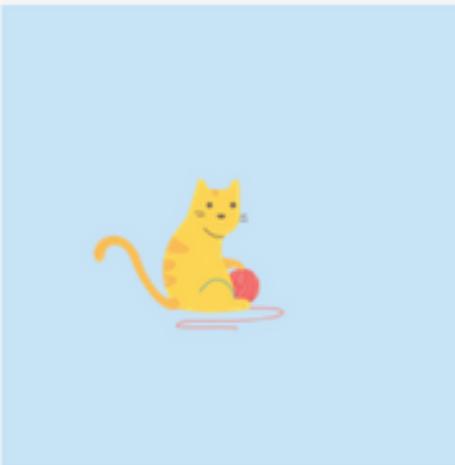
The Image Classification Challenge:
1,000 object classes
1,431,167 images



Vision Tasks

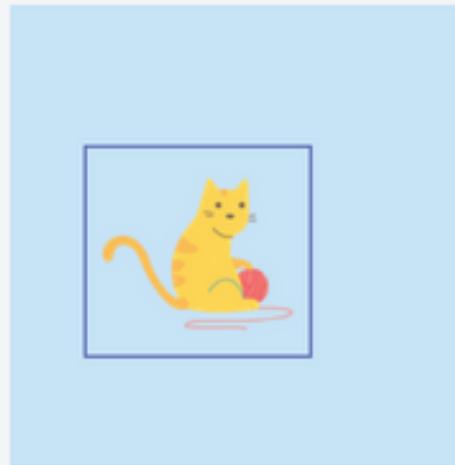
Computer Vision Problem Types

Classification



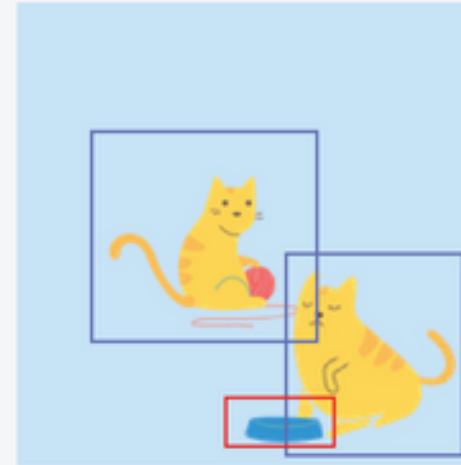
CAT

Classification
+ Localization



CAT

Object Detection



CAT, CAT, BOWL

Single Object

Semantic
Segmentation



CAT, CAT, BOWL

Multiple Objects

roboflow

제조 현장에서의 컴퓨터 비전 적용 사례

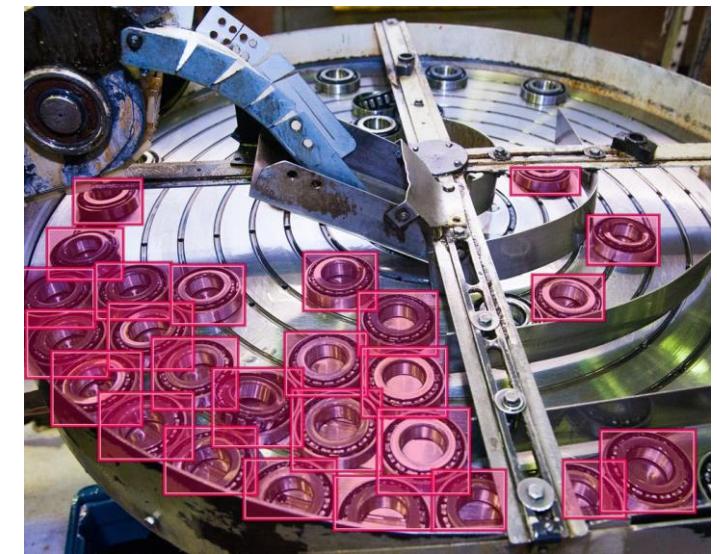
Predictive maintenance



Defect detection



Assembling products and components

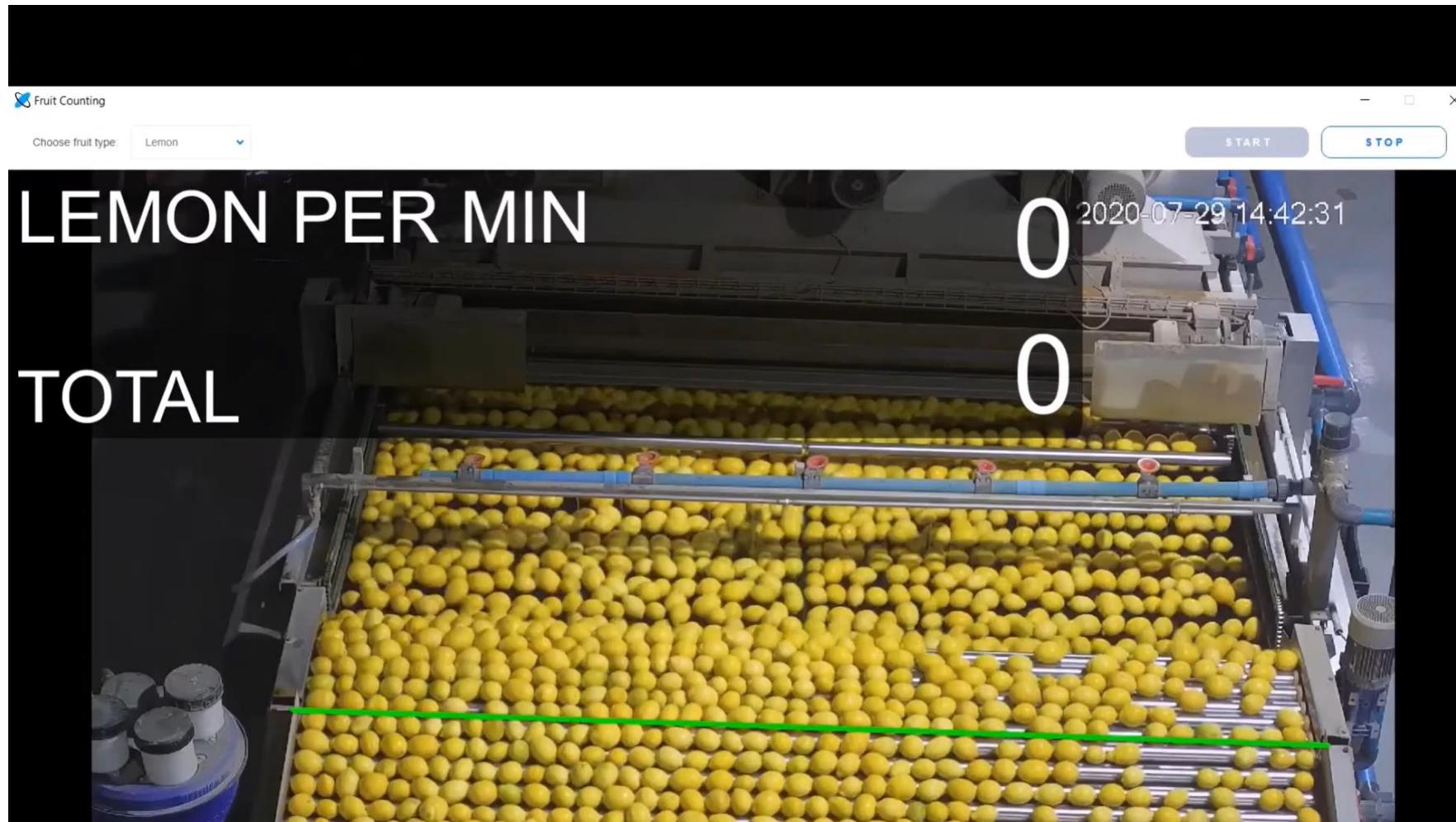


[출처] <https://blog.roboflow.com/computer-vision-use-cases-for-manufacturing/>

Object Counting



Object Counting

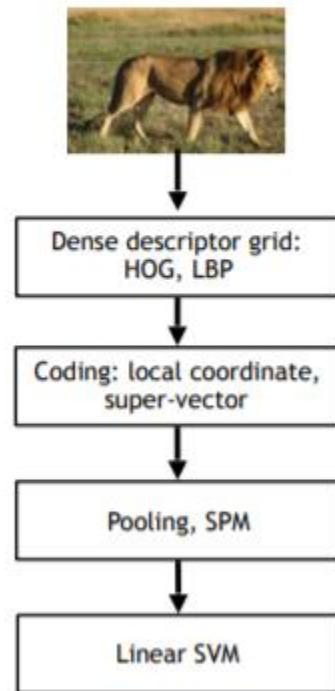


Convolutional Neural Networks (CNN)
were not invented overnight

IMAGENET Large Scale Visual Recognition Challenge

Year 2010

NEC-UIUC

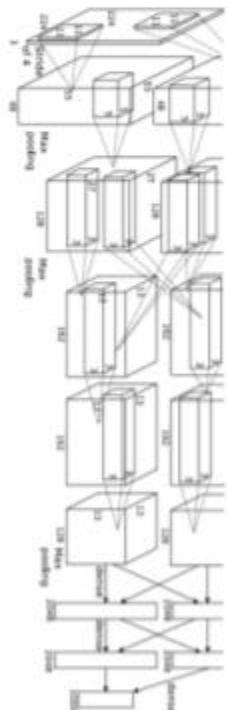


[Lin CVPR 2011]

Lion image by Swissfrog is licensed under CC BY 3.0

Year 2012

SuperVision



[Krizhevsky NIPS 2012]

Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

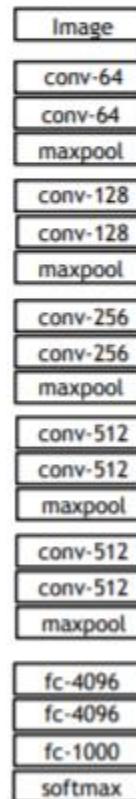
Year 2014

GoogLeNet

- Pooling
- Convolution
- Softmax
- Other



VGG

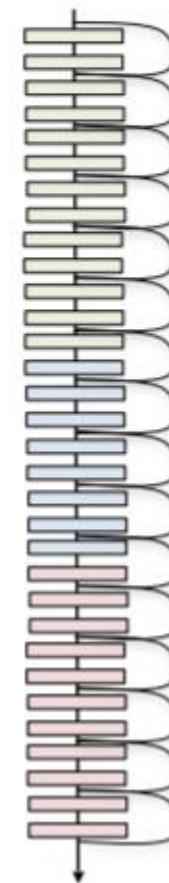


[Szegedy arxiv 2014]

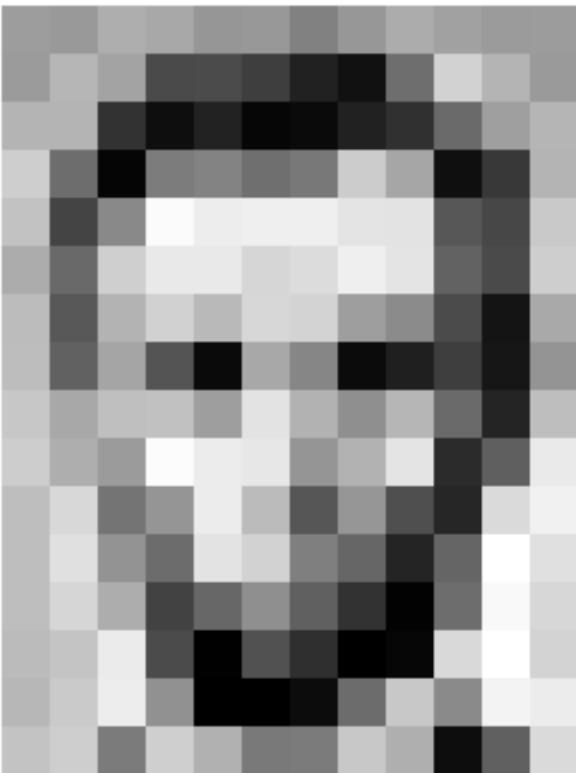
[Simonyan arxiv 2014]

Year 2015

MSRA

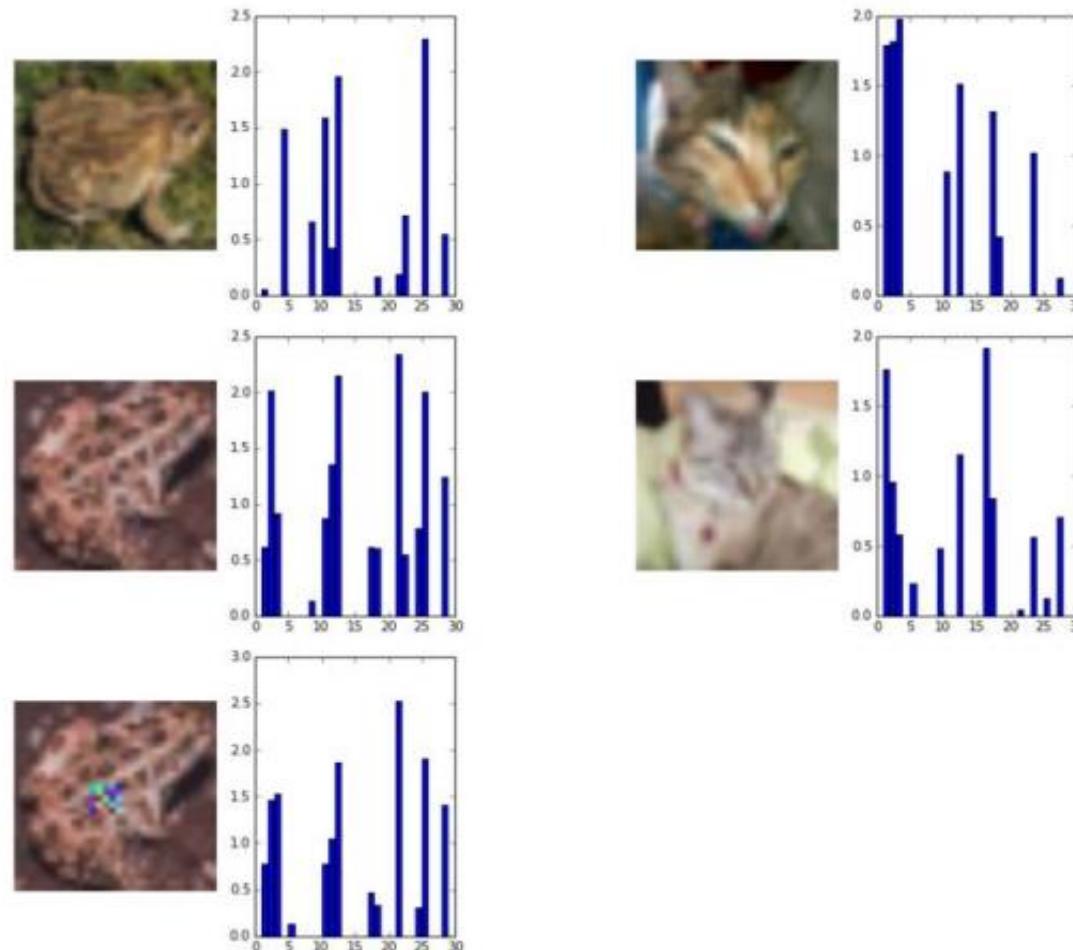


[He ICCV 2015]

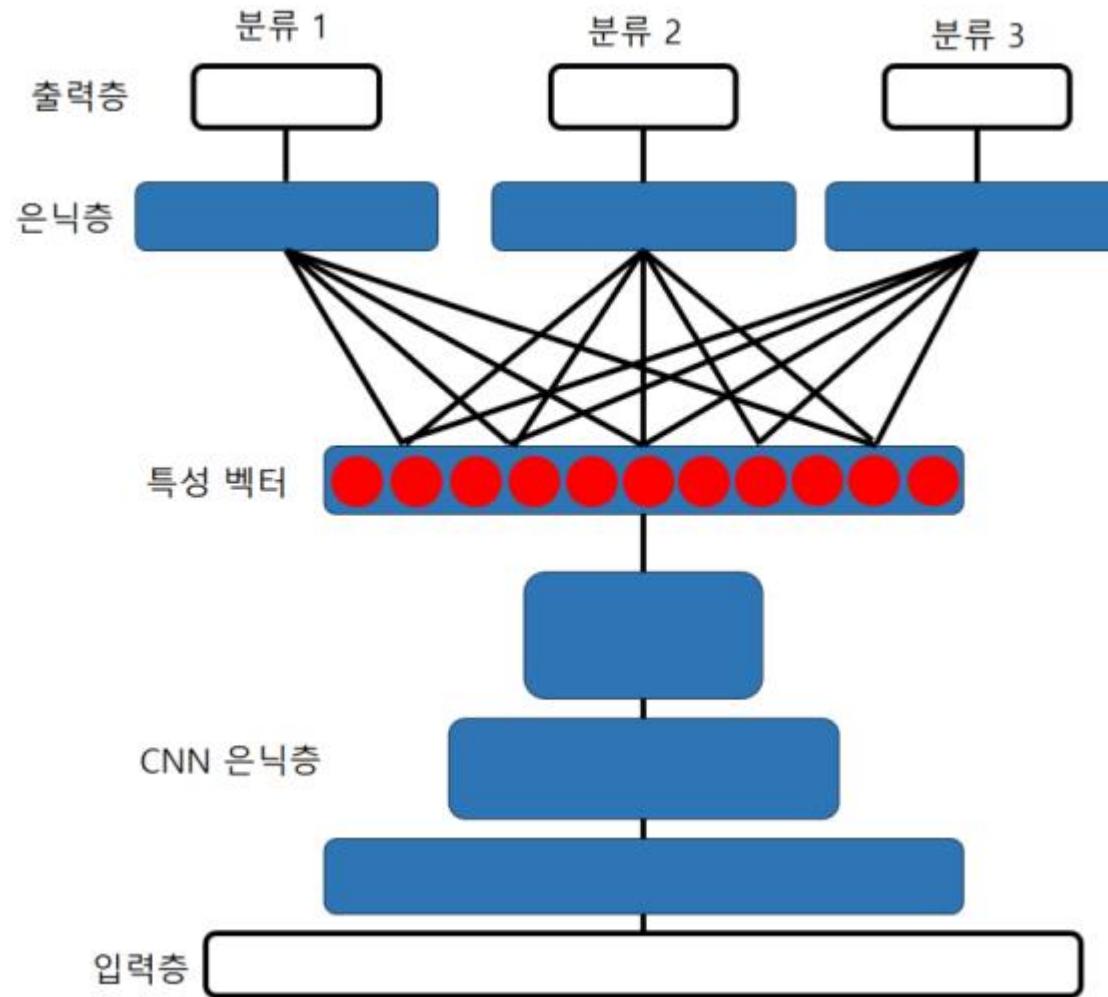


157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	105	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	105	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	105	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218



서로 다른 클래스의 물체들과 딥러닝 방식을 통해 추출한 특성 벡터. 이미지 내의 선, 모양, 무늬들을 감지해 이것을 값들의 집합으로 나타낸다. 딥러닝으로 추출된 벡터는 이미지의 범주나 구도 등에 따라 상이한 결과가 나오며, 여러에 관한 내성이 존재한다.

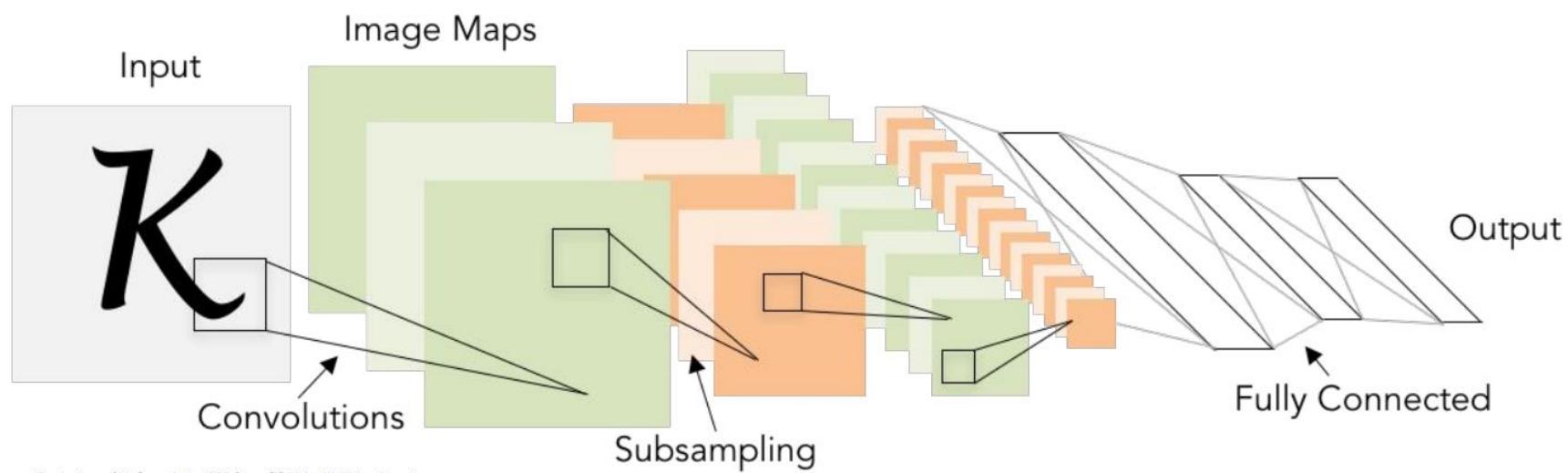


(예시) 다목적 이미지 분류를 위한 전체적인 구조도로서 학습된 딥러닝 구조를 통해 특성 벡터를 추출하고, 특성 벡터를 인공 신경망의 입력으로 삼아 다양한 분류기를 구성한다.

CNN (Convolution neural networks)

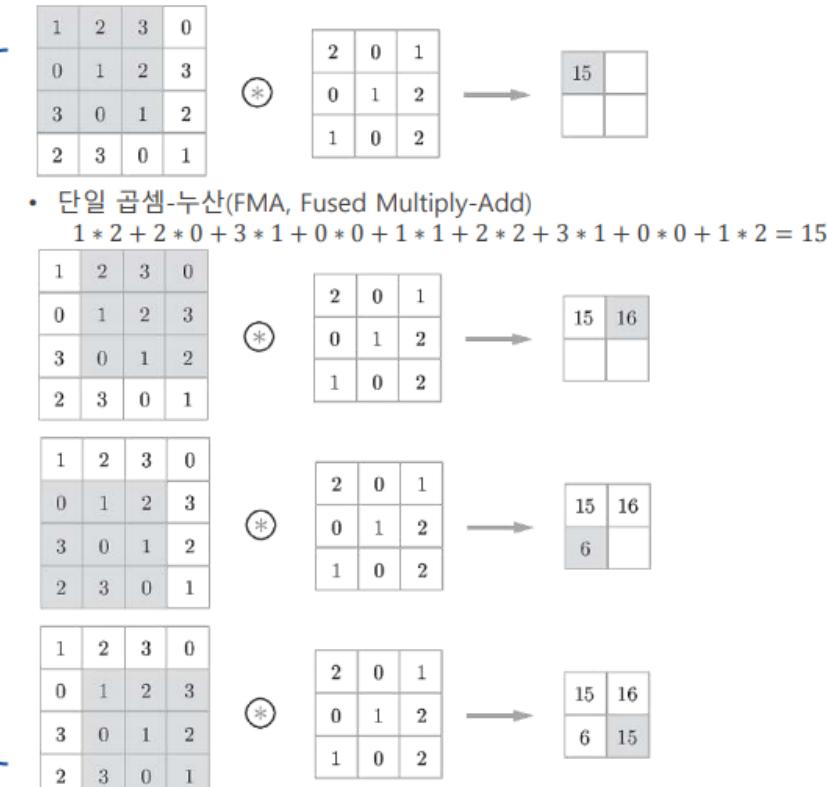
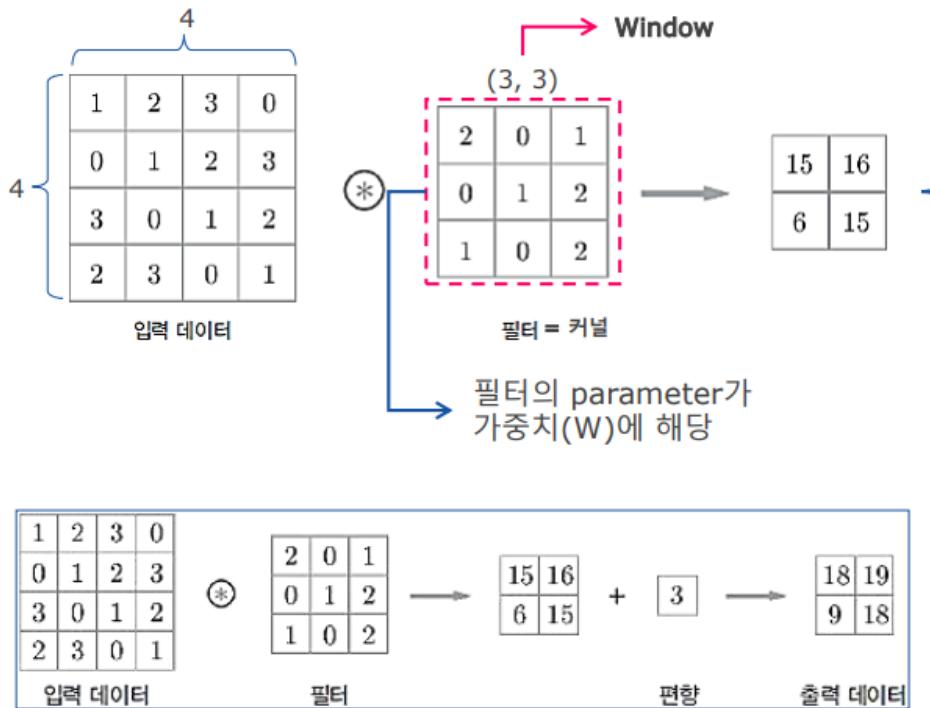
Gradient-based learning applied to document recognition

[LeCun, Bottou, Bengio, Haffner 1998]

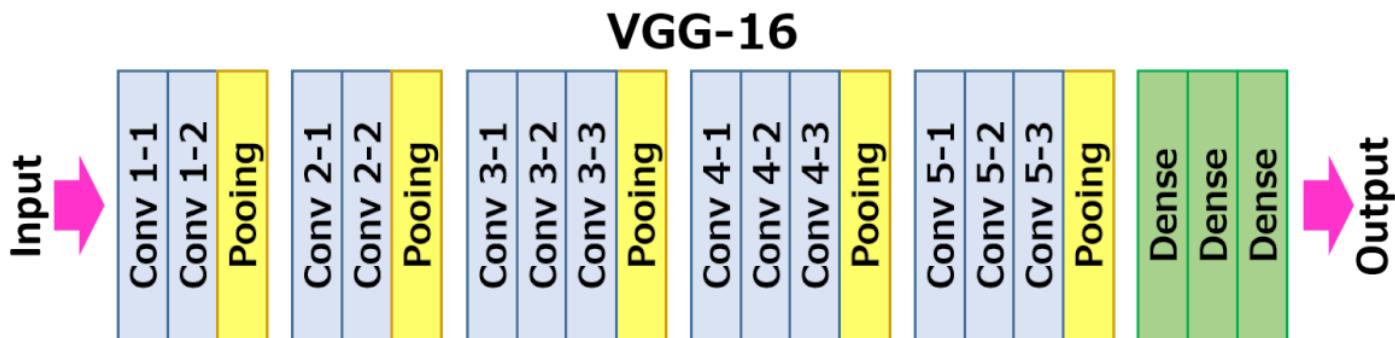
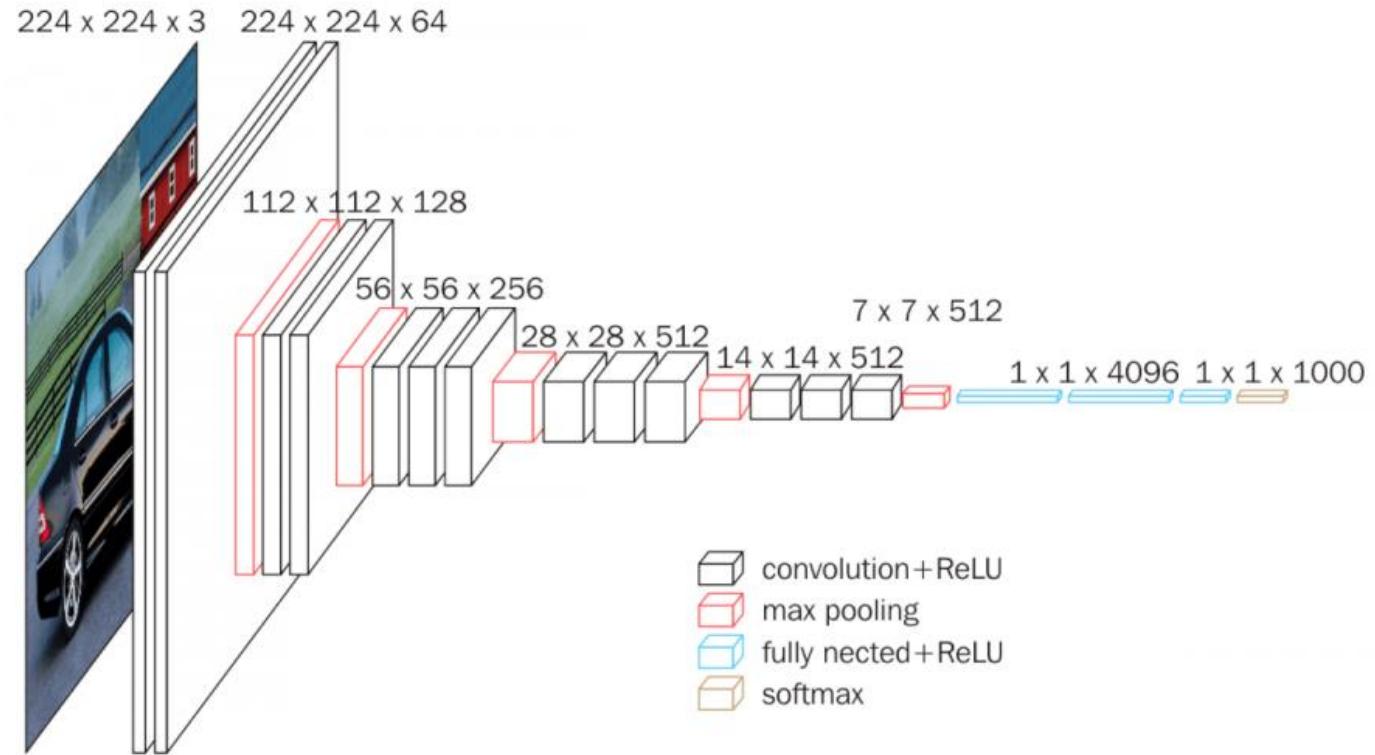


합성곱 계층 - 연산

- ✓ 합성곱 계층에서 연산 수행 → 필터(커널) 연산
- ✓ 데이터와 필터의 형상을 (높이height, 너비width)로 표기
- ✓ 윈도우window 를 일정 간격(Stride)으로 이동하며 계산

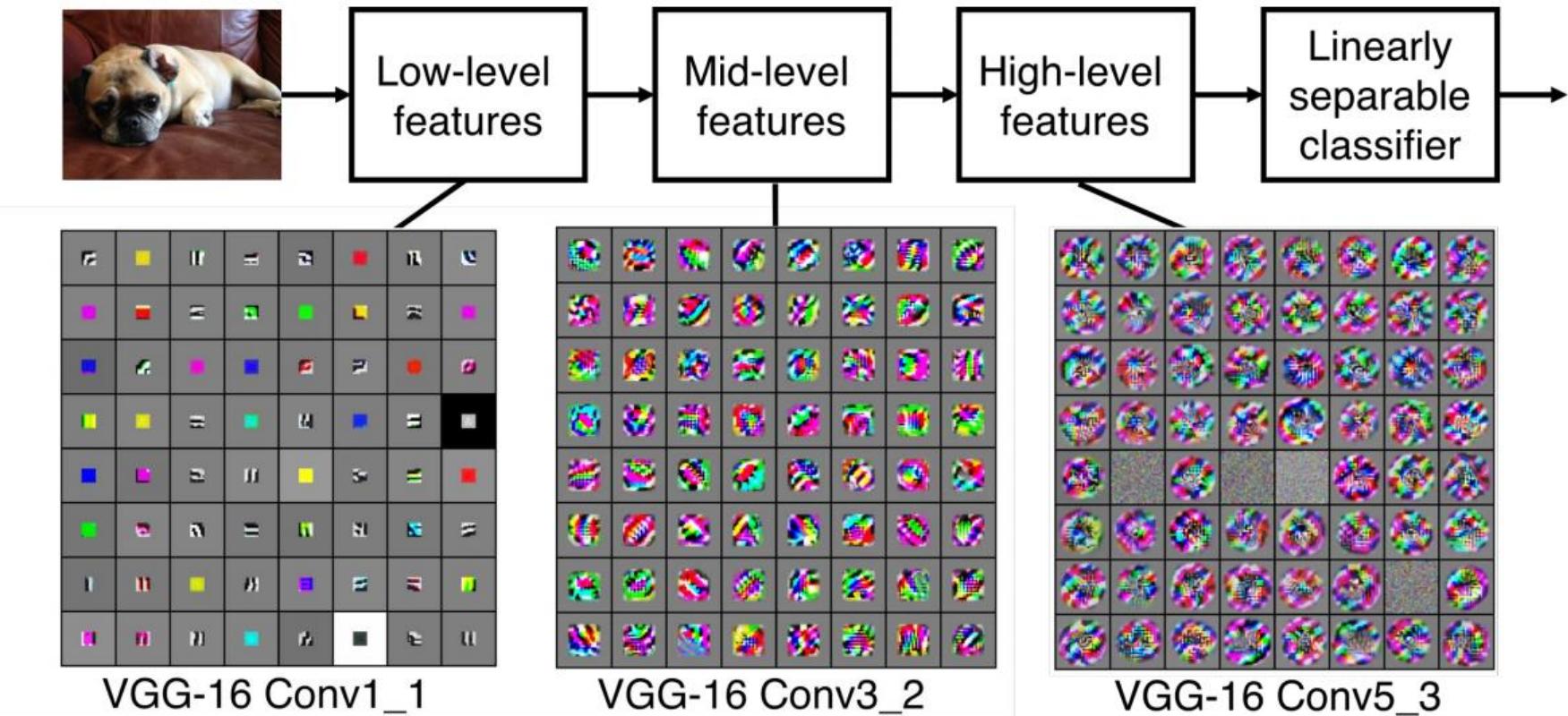


VGG-16 Pre-Trained Model 구성



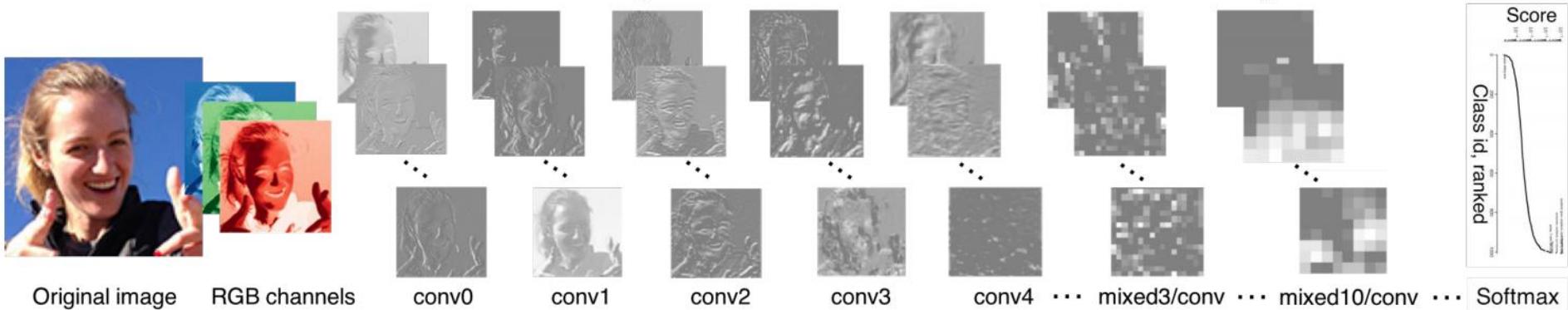
Layer 별 Features

Preview

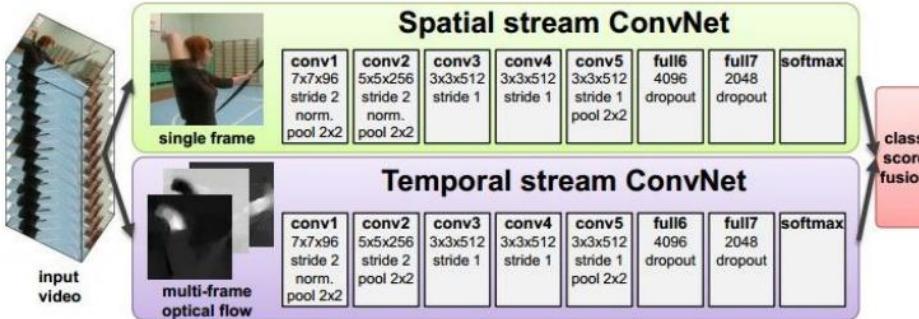


Activation functions

Fast-forward to today: ConvNets are everywhere

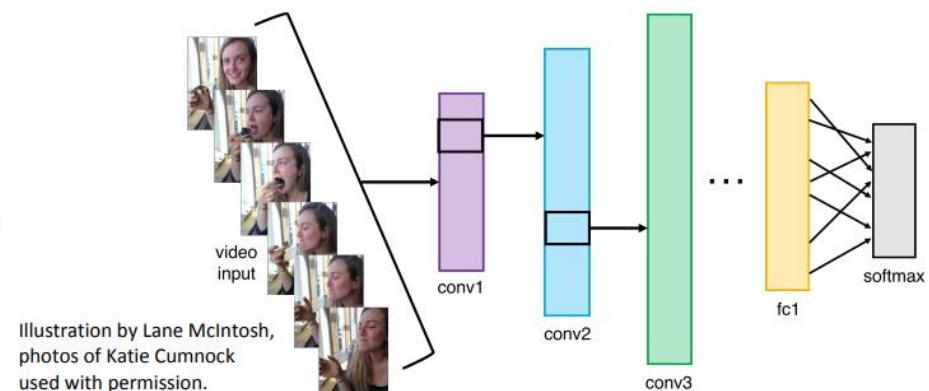


Activations of [inception-v3 architecture](#) [Szegedy et al. 2015] to image of Emma McIntosh, used with permission. Figure and architecture not from Taigman et al. 2014.



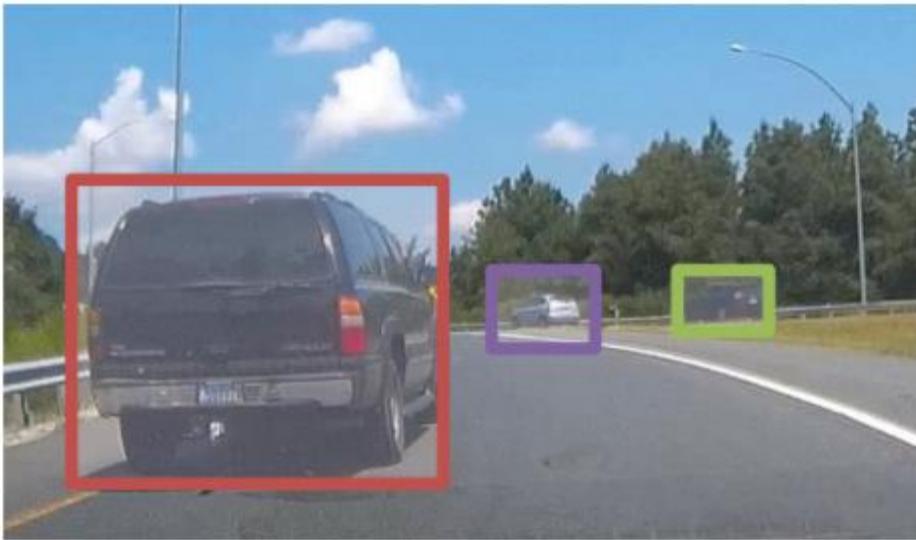
[Simonyan et al. 2014]

Figures copyright Simonyan et al., 2014.
Reproduced with permission.

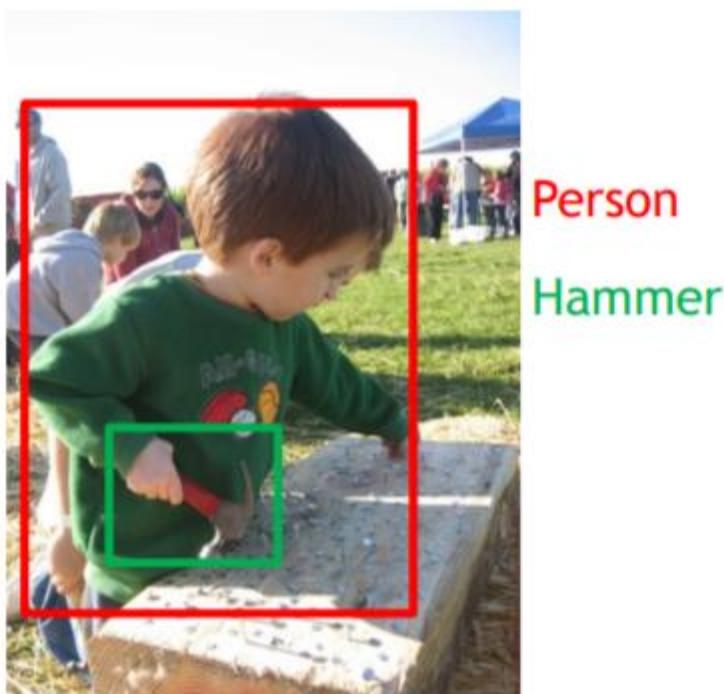


The quest for visual intelligence
goes far beyond object recognition...

Convolutional Neural Networks (CNN) have become an important tool for object recognition



[This image](#) is licensed under CC BY-NC-SA 2.0; changes made

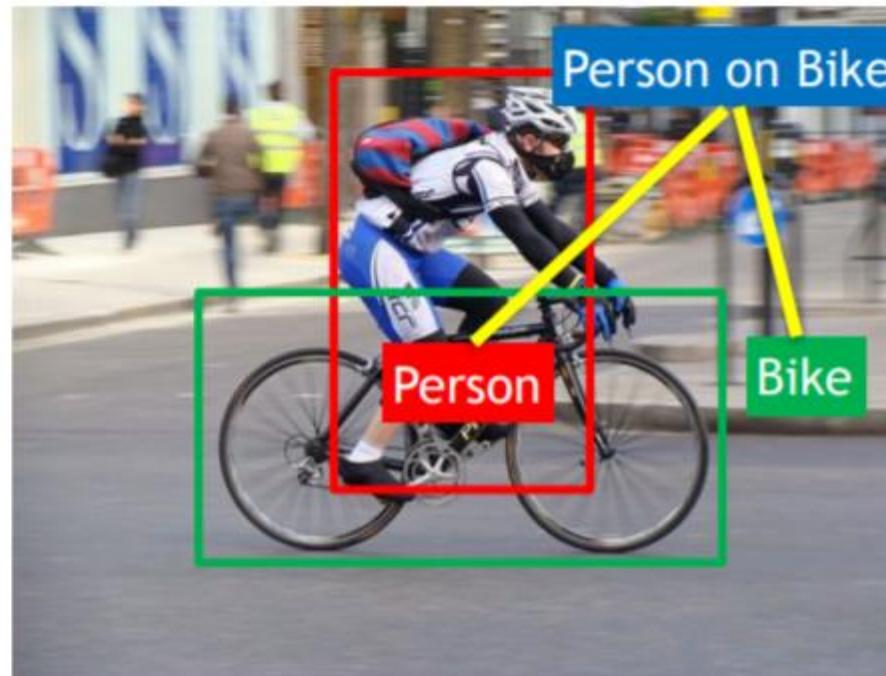


Person

Hammer

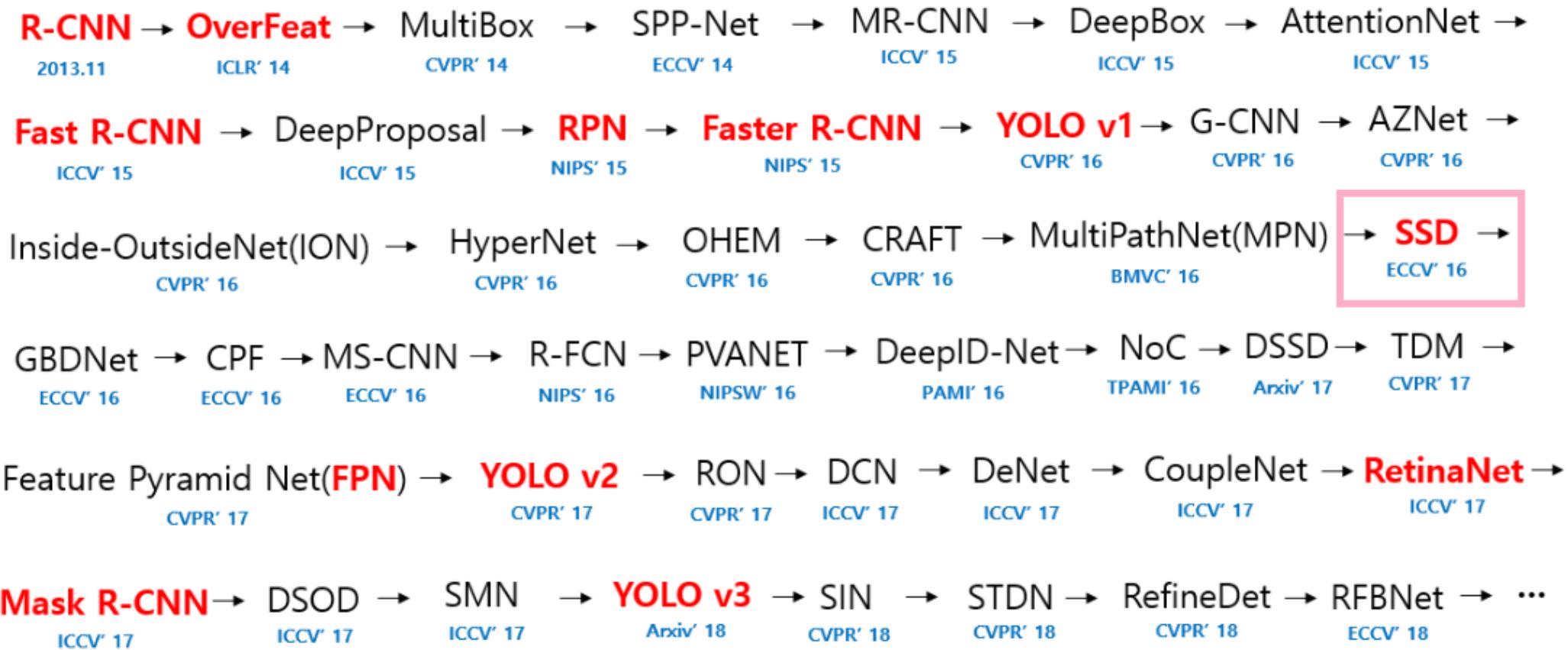
[This image](#) is licensed under CC BY-SA 2.0; changes made

- Object detection
- Action classification
- Image captioning
- ...



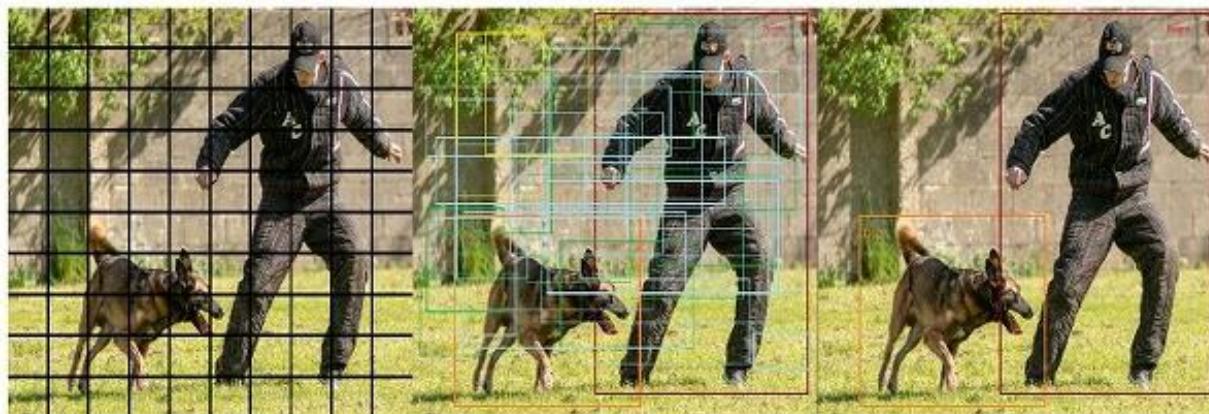
[This image](#) is licensed under CC BY-SA 3.0; changes made

Object Detection Models



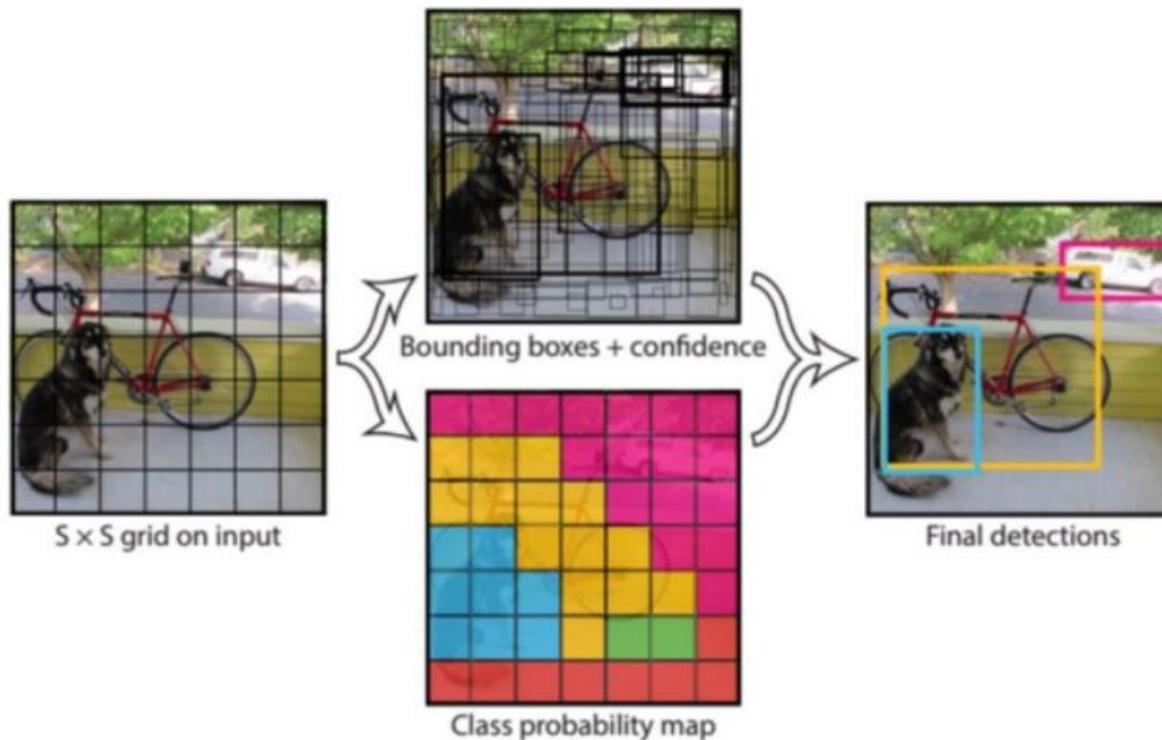
YOLO - 단일 단계 방식의 객체 탐지 알고리즘

- YOLO(You Only Look Once)는 대표적인 단일 단계 방식의 객체 탐지 알고리즘입니다. YOLO 알고리즘은 원본 이미지를 동일한 크기의 그리드로 나눕니다.
- 각 그리드에 대해 그리드 중앙을 중심으로 미리 정의된 형태(predefined shape)으로 지정된 경계박스의 개수를 예측하고 이를 기반으로 신뢰도를 계산합니다.
- 이미지에 객체가 포함되어 있는지, 또는 배경만 단독으로 있는지에 대한 여부가 포함되겠죠. 높은 객체 신뢰도를 가진 위치를 선택해 객체 카테고리를 파악합니다.



YOLO - 객체 탐지 방식

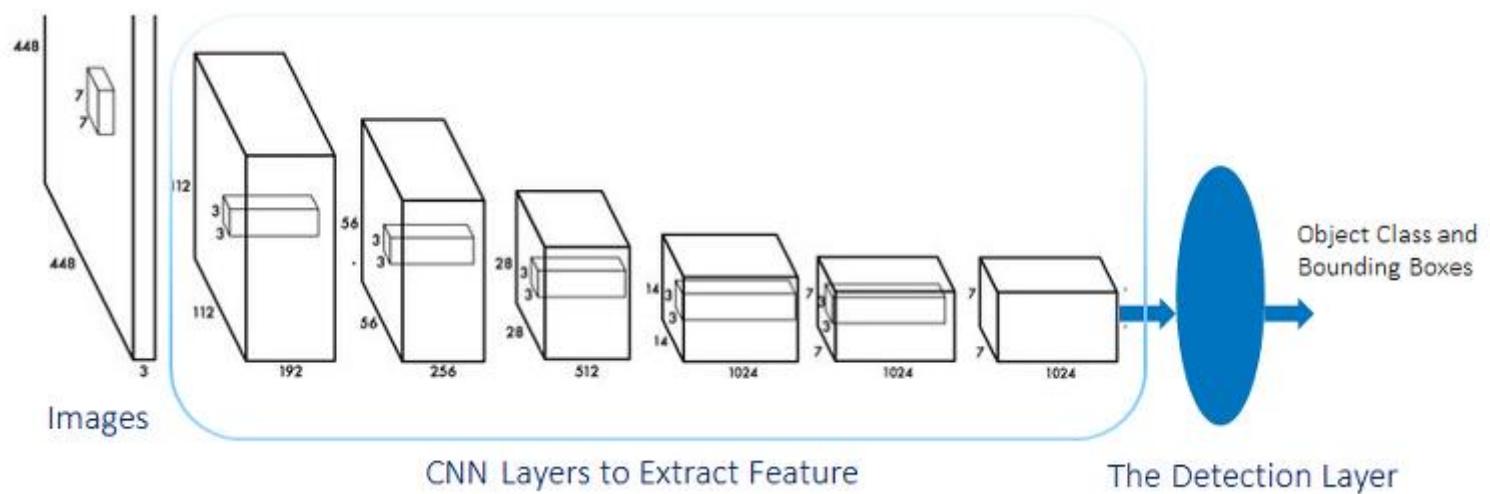
- 영역의 박스와 해당 박스의 분류를 동시에 복수개를 출력하는 네트워크
- 모두 98개($7 \times 7 \times 2$)가 제안되고, 중복되지 않고 확신도가 높은 것만 선택



모델 구조는 입력 이미지를 $S \times S$ 그리드로 나누고, 객체의 중심이 그리드 셀에 들어가면 해당 그리드 셀이 해당 객체를 감지함.



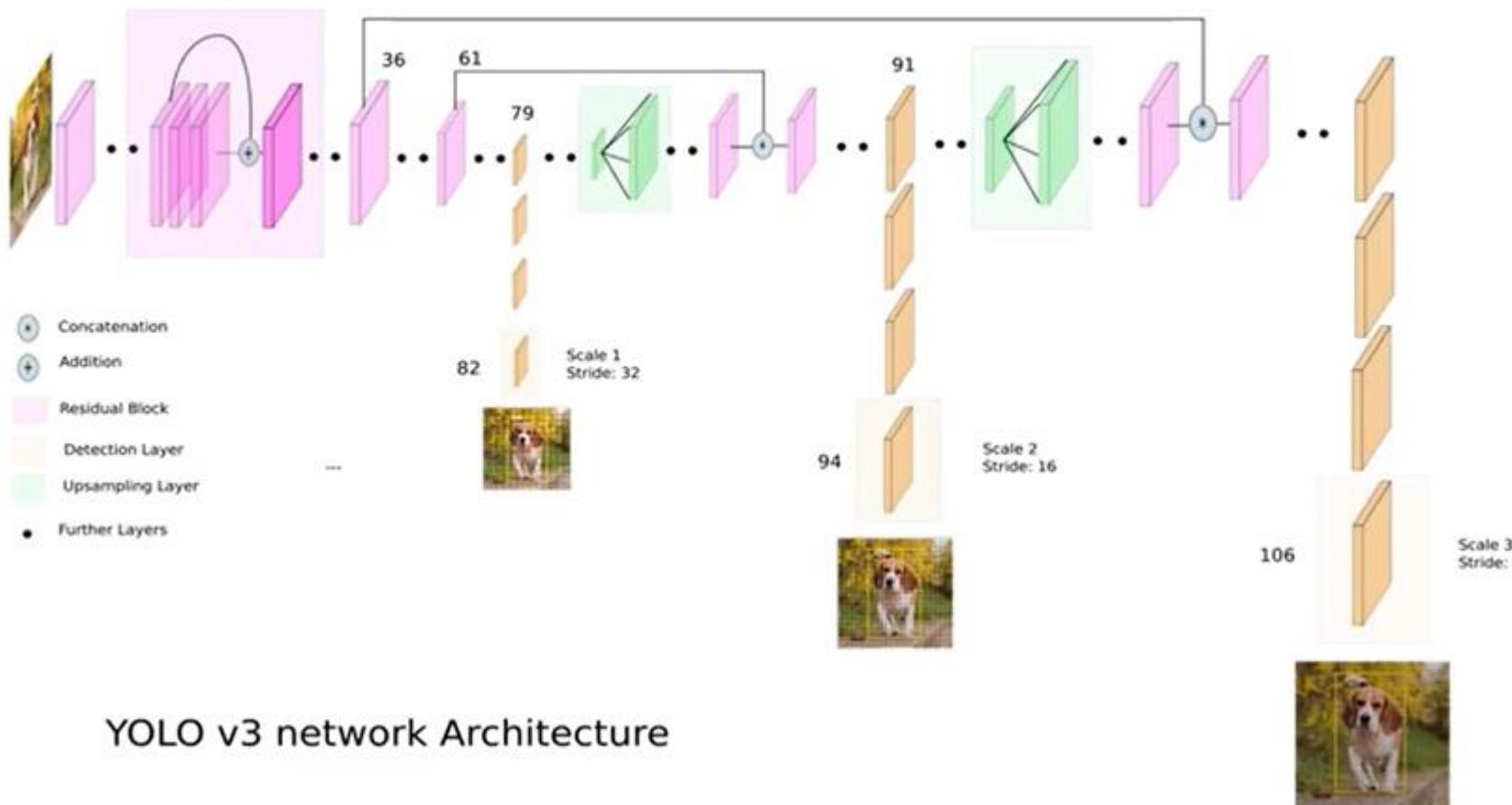
[그림 1] 앵커 박스(이미지 출처: https://www.flickr.com/photos/762_photo/16751321393/)



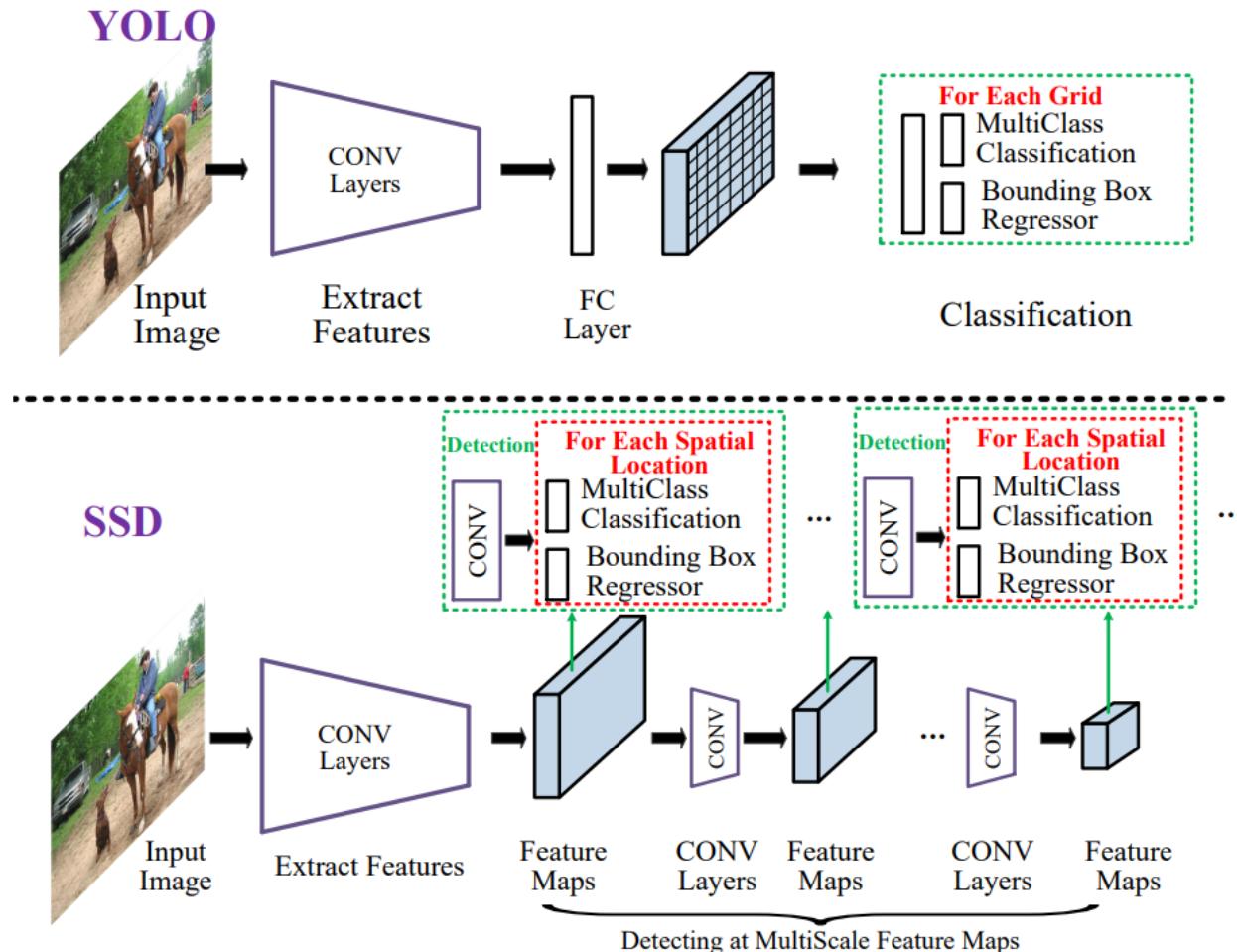
[그림 2] YOLO 네트워크 아키텍처

YOLO3 – Architecture

YOLO3

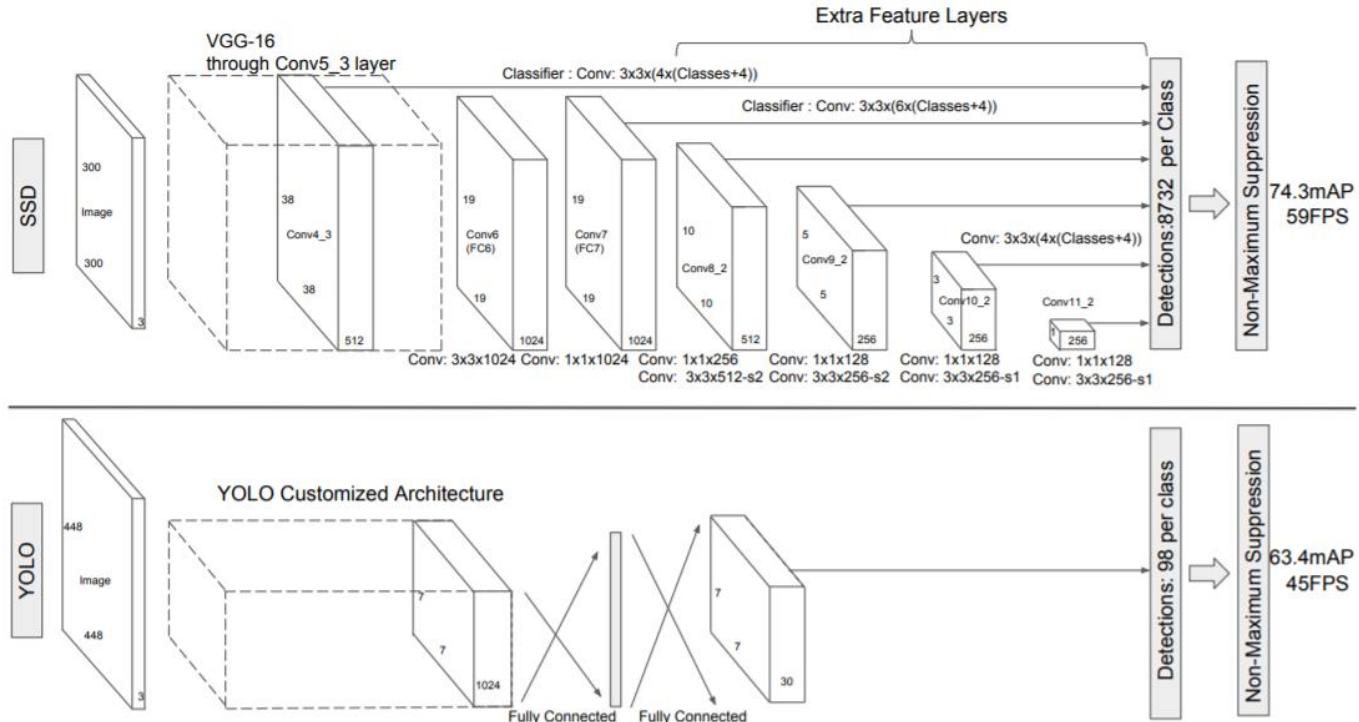


YOLO와 SSD 모델의 차이 비교



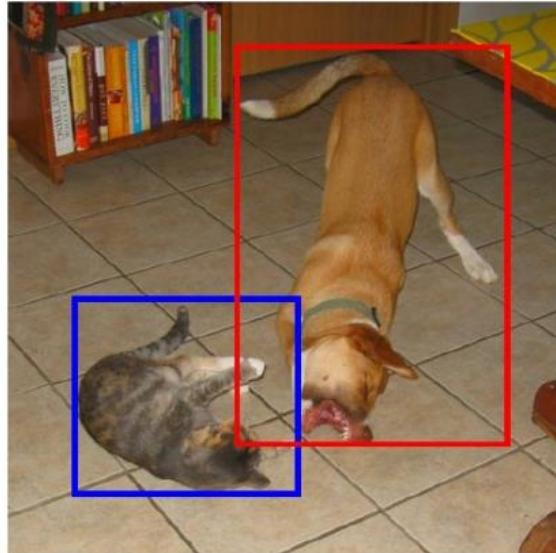
경계 상자 가설 (1 단계)에 대한 픽셀 또는 특징을 리샘플링하지 않고 접근 방식만큼 정확한 최초의 딥 네트워크 기반 객체 감지함.

Multifull Feature Map의 활용

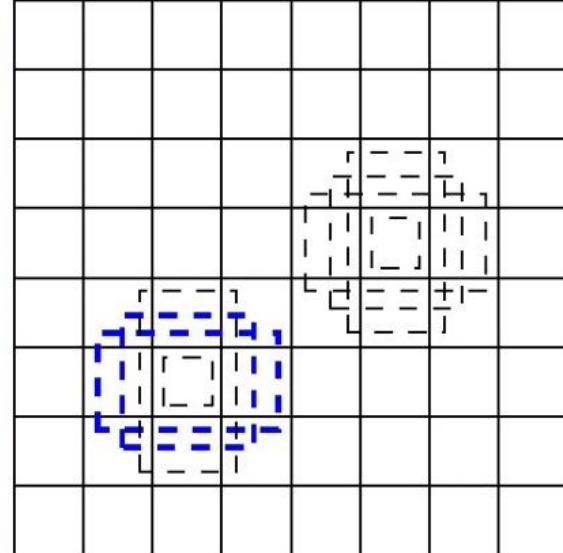


6개의 기본 컨볼루션 네트워크의 끝에서 피처를 가져오며, 이러한 Layer의 크기는 점진적으로 줄어들고 여러 스케일에서 탐지하여 예측 가능.

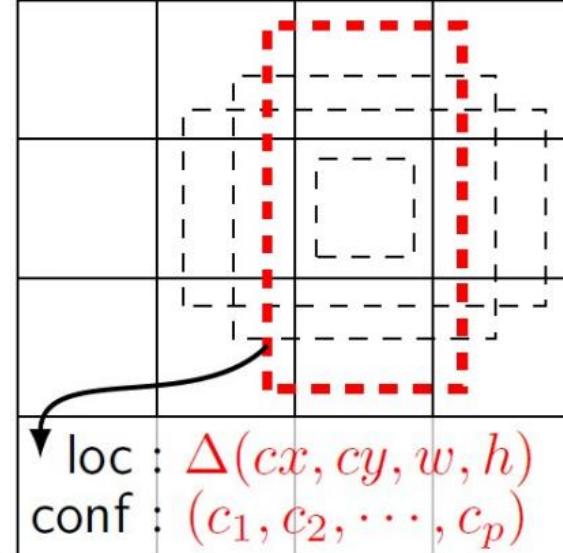
다양한 배율 Convolution Layer에서 종횡 비율에 따른 선택



(a) Image with GT boxes



(b) 8×8 feature map

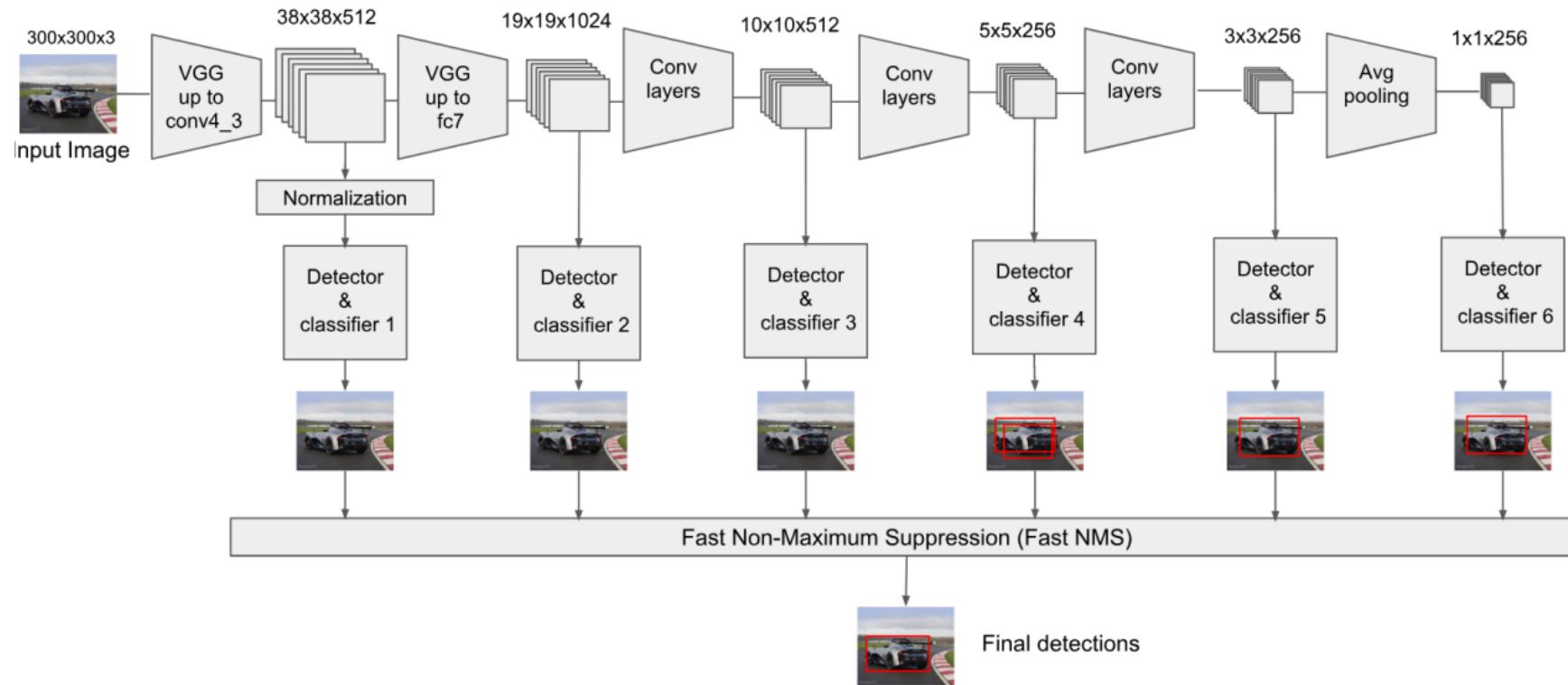


▼ loc : $\Delta(cx, cy, w, h)$
conf : (c_1, c_2, \dots, c_p)

(c) 4×4 feature map

다른 레이어의 피처 맵을 사용하여 스케일 차이를 처리하며, 특정 기능 맵
위치는 이미지의 특정 영역과 개체의 특정 배율에 반응하는 법을 학습함.

Anchor Box에 대한 배율 및 종횡비에 따른 선택



상대적으로 큰 물체는 종단 Layer에서 Object를 선택함.

YOLO – Anchor Boxes

- 미리 정의된 형태를 가진 경계박스 수를 '앵커 박스(Anchor Boxes)'라고 하는데요. 앵커 박스는 K-평균 알고리즘에 의한 데이터로부터 생성되며, 데이터 세트의 객체 크기와 형태에 대한 사전 정보를 확보합니다.
- 각각의 앵커는 각기 다른 크기와 형태의 객체를 탐지하도록 설계되어 있습니다. 앞장의 [그림 2]를 보면 한 장소에 3개의 앵커가 있는데요. 이 중 붉은색 앵커 박스는 가운데에 있는 사람을 탐지합니다.
- 이 알고리즘은 앵커 박스와 유사한 크기의 객체를 탐지한다는 뜻인데요. 최종 예측은 앵커의 위치나 크기와는 차이가 있습니다. 이미지의 피쳐(Feature) 맵에서 확보한 최적화된 오프셋이 앵커 위치나 크기에 추가됩니다.

앵커 박스?

이미지에서 다양한 물체를 예측하고 현지화하기 위해 EfficientDet 및 YOLO 모델과 같은 대부분의 최신 물체 감지 모델은 이전과 같이 앵커 상자로 시작하여 거기에서 조정됩니다.

최근 물체 감지 모델은 일반적으로 다음 순서로 경계 상자(Bounding Box)를 사용합니다.

이미지 주위에 수천 개의 후보 앵커 박스 형성

- 각 앵커 박스에 대해 해당 박스에서 일부 오프셋을 후보 상자로 예측
- Ground Truth 예를 기반으로 손실 함수 계산
- 주어진 오프셋 상자가 실제 물체와 겹칠 확률 계산
- 해당 확률이 0.5보다 크면 예측을 손실 함수로 인수합니다.
- 예측된 박스를 보상하고 페널티를 줌으로써 실제 객체 만 현지화 하는 쪽으로 모델을 천천히 이끌어 냅니다.

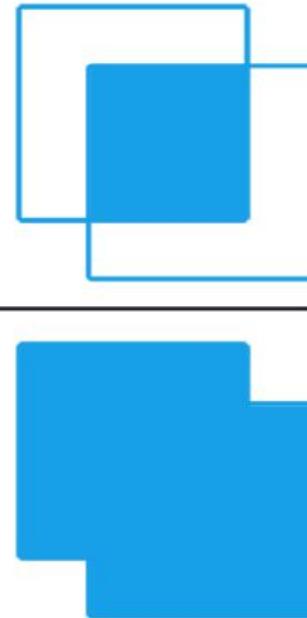
그렇기 때문에 모델을 약간만 훈련했을 때 모든 곳에 예측 상자가 표시되는 것을 볼 수 있습니다.

Anchor sorting and filtering



학습방법 – 매칭 전략

$$\text{Jaccard Overlap} = \text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

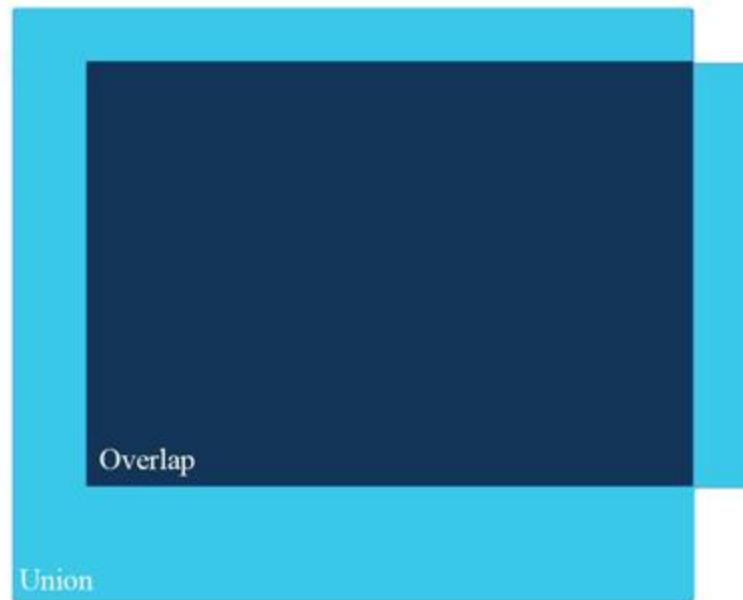


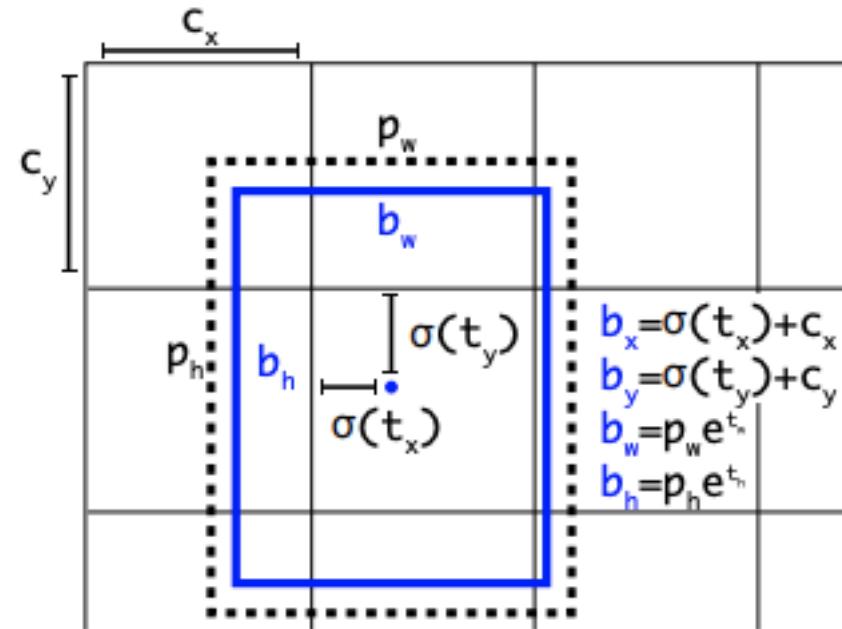
모든 Ground Truth 상자에는 최소한 1 개의 대응을 시키며, jaccard 겹침이 가장 큰 Default Box에 각 원본 영상과 일치시킴.
그런 다음 jaccard 오버랩이 임계 값 (0.5)보다 높은 모든 Ground Truth에 Default Box에 일치시킴.



- Ground truth
- Prediction

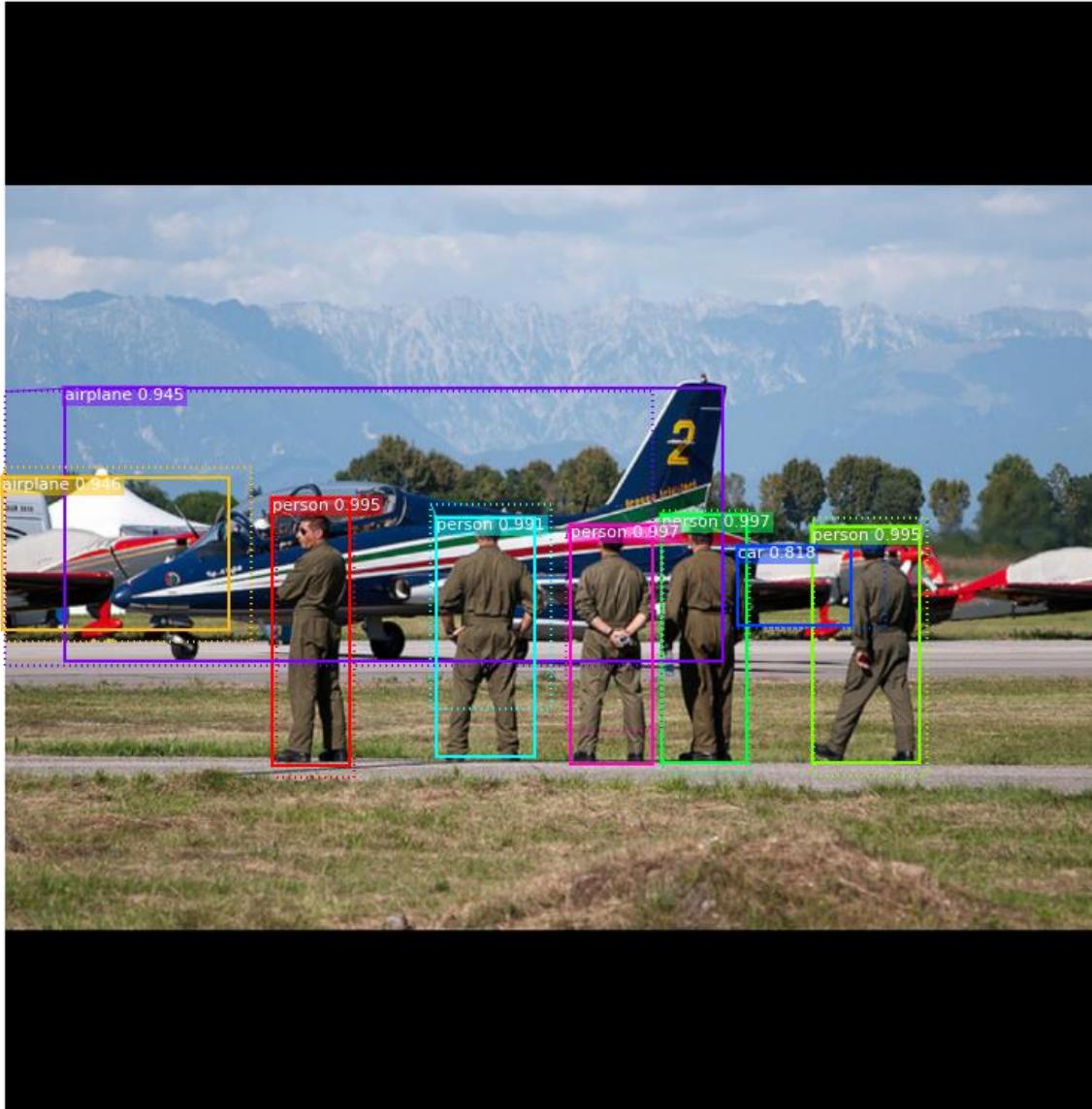
$$IoU = \frac{\text{area of overlap}}{\text{area of union}}$$





A picture of bounding box prediction based on an anchor box

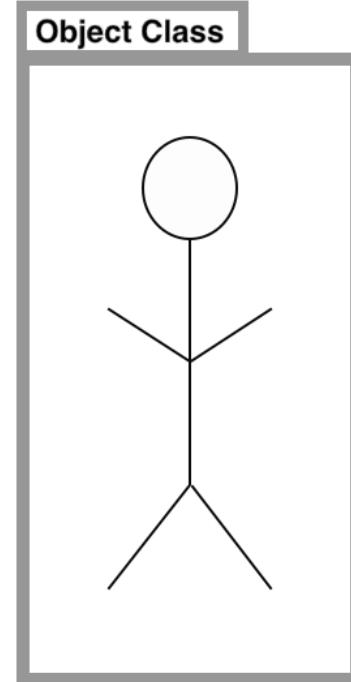
Bounding Box Refinement(Detection after NMS)



Y2



Object Class



Y1

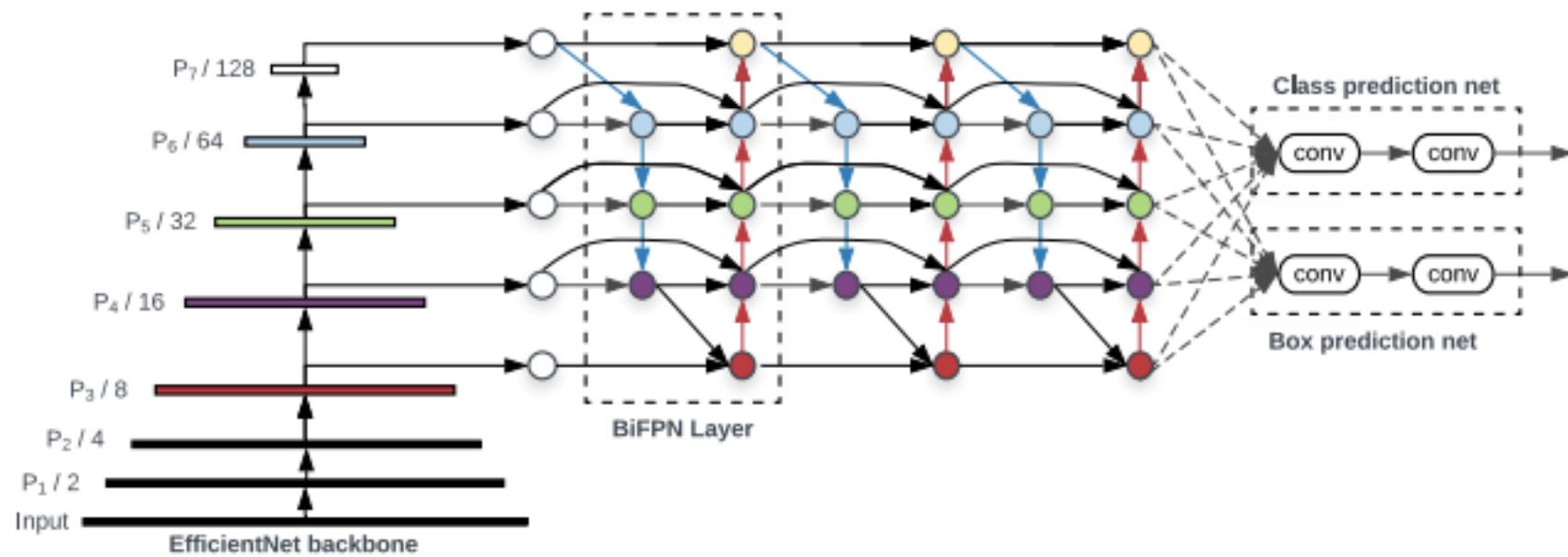


X1

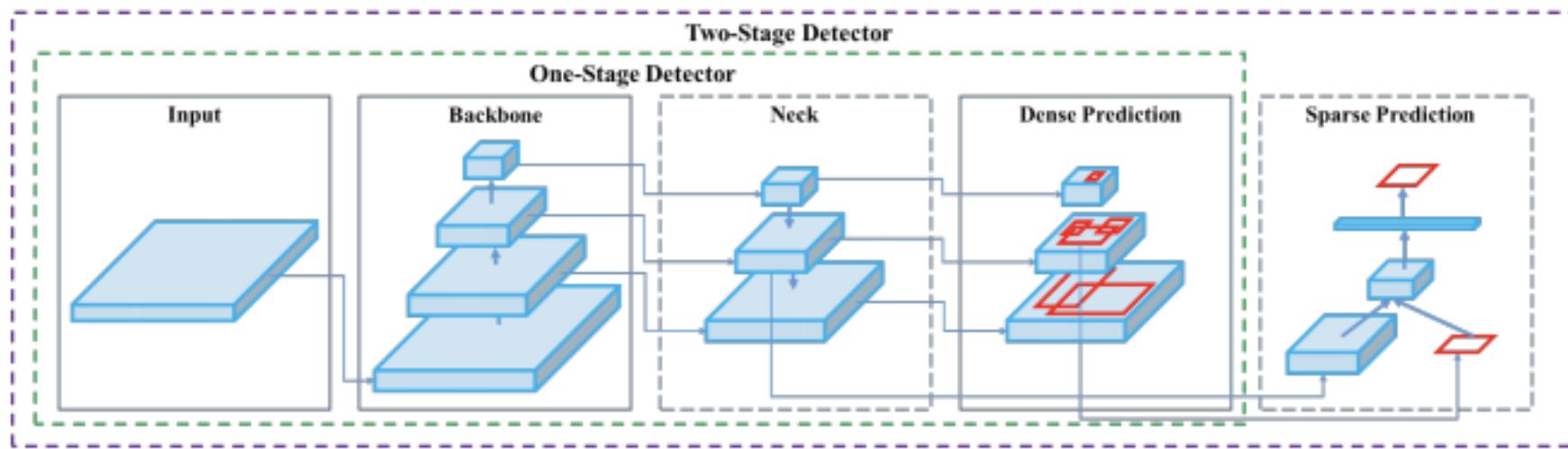
X2



The anatomy of an object detector ([citation](#))



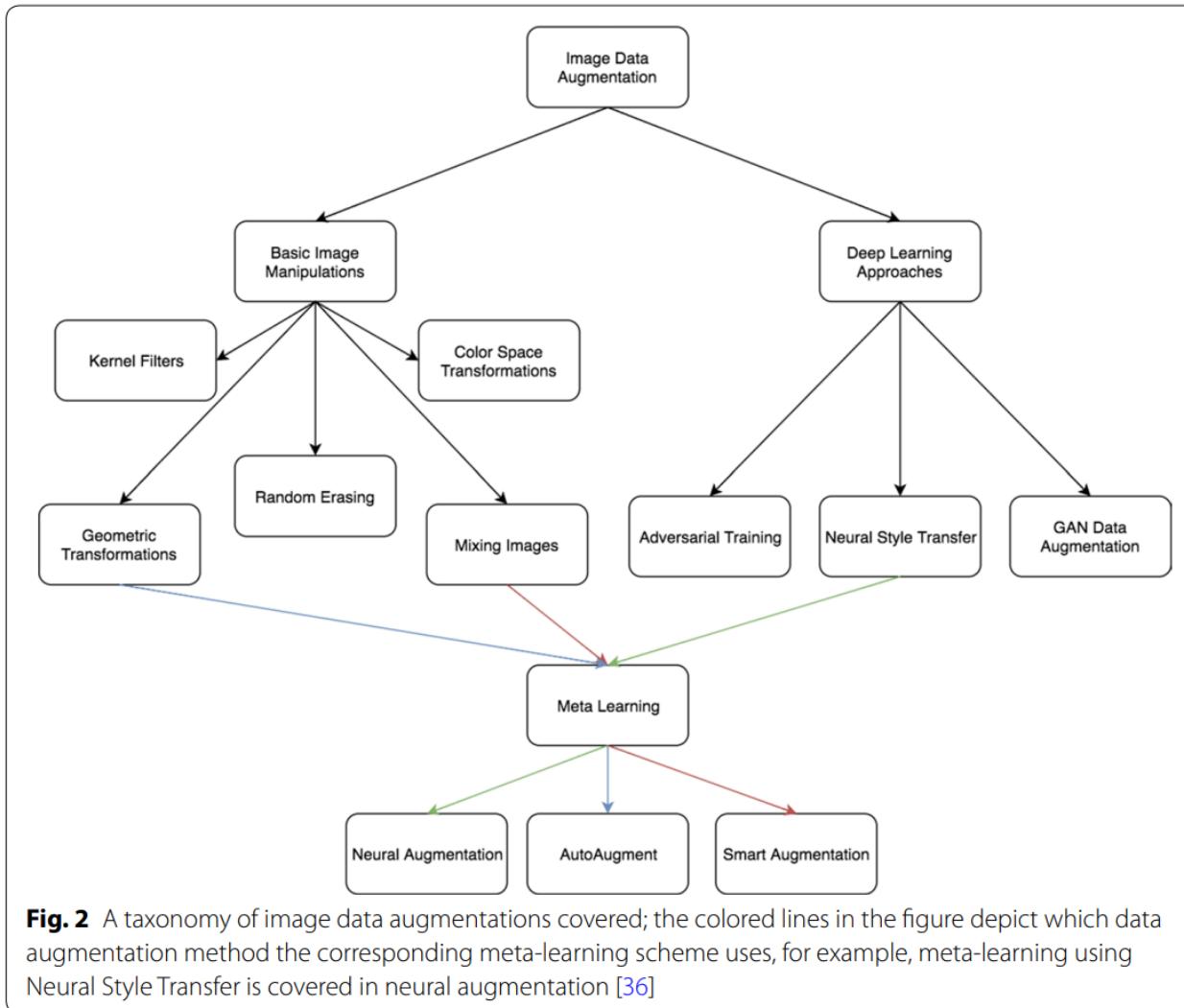
Another picture of the object detection process ([citation](#) from YOLOv4)



```
# parameters nc: 80
# number of classes depth_multiple: 0.33
# model depth multiple width_multiple: 0.50 #
layer channel multiple
# anchors
anchors:
    - [116,90, 156,198, 373,326] # P5/32
    - [30,61, 62,45, 59,119] # P4/16
    - [10,13, 16,30, 33,23] # P3/8
# YOLOv5 backbone
```

The anchors in the YOLOv5 config file are now auto learned based on training data.

Augmentation



Attention is all you need

Attention Is All You Need

Ashish Vaswani^{*}
Google Brain
avaswani@google.com

Noam Shazeer^{*}
Google Brain
noam@google.com

Niki Parmar^{*}
Google Research
nikip@google.com

Jakob Uszkoreit^{*}
Google Research
usz@google.com

Llion Jones^{*}
Google Research
llion@google.com

Aidan N. Gomez^{* †}
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser^{*}
Google Brain
lukasz.kaiser@google.com

Illia Polosukhin^{* ‡}
illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

1 Introduction

Recurrent neural networks, long short-term memory [13] and gated recurrent [7] neural networks in particular, have been firmly established as state of the art approaches in sequence modeling and

^{*}Equal contribution. Listing order is random. Jakob proposed replacing RNNs with self-attention and started the effort to evaluate this idea. Ashish, with Illia, designed and implemented the first Transformer models and has been crucially involved in every aspect of this work. Noam proposed scaled dot-product attention, multi-head attention and the parameter-free position representation and became the other person involved in nearly every detail. Niki designed, implemented, tuned and evaluated countless model variants in our original codebase and tensor2tensor. Llion also experimented with novel model variants, was responsible for our initial codebase, and efficient inference and visualizations. Lukasz and Aidan spent countless long days designing various parts of and implementing tensor2tensor, replacing our earlier codebase, greatly improving results and massively accelerating our research.

[†]Work performed while at Google Brain.
[‡]Work performed while at Google Research.

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

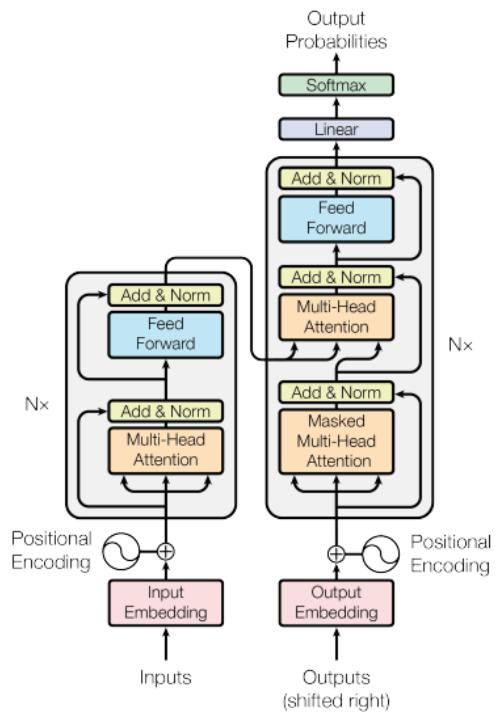
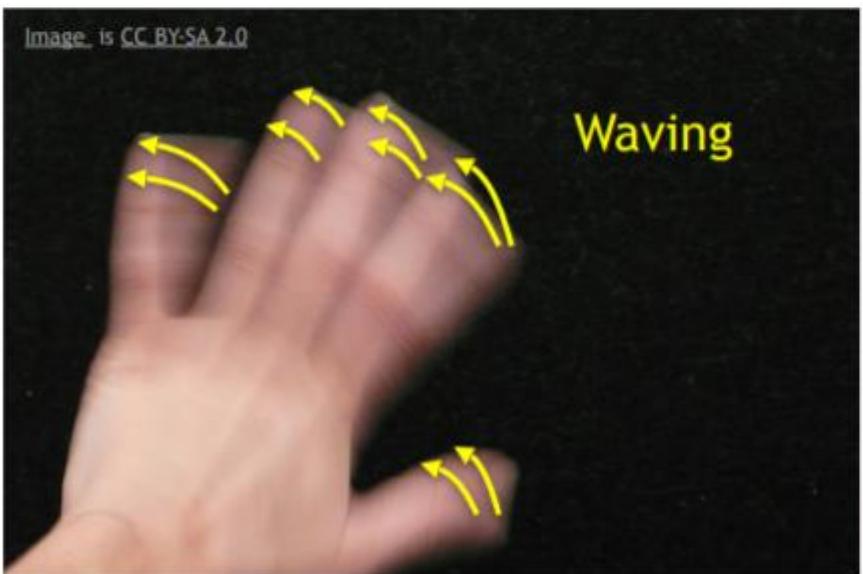
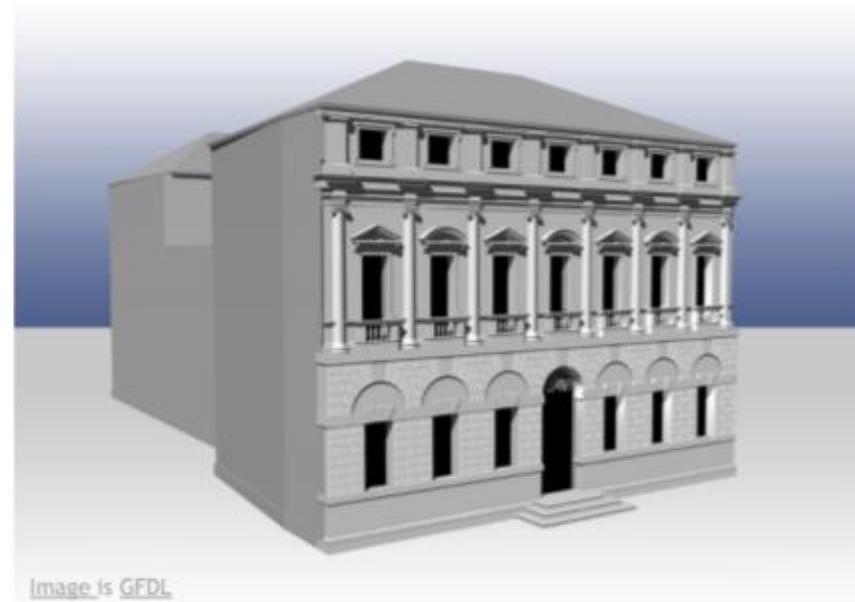
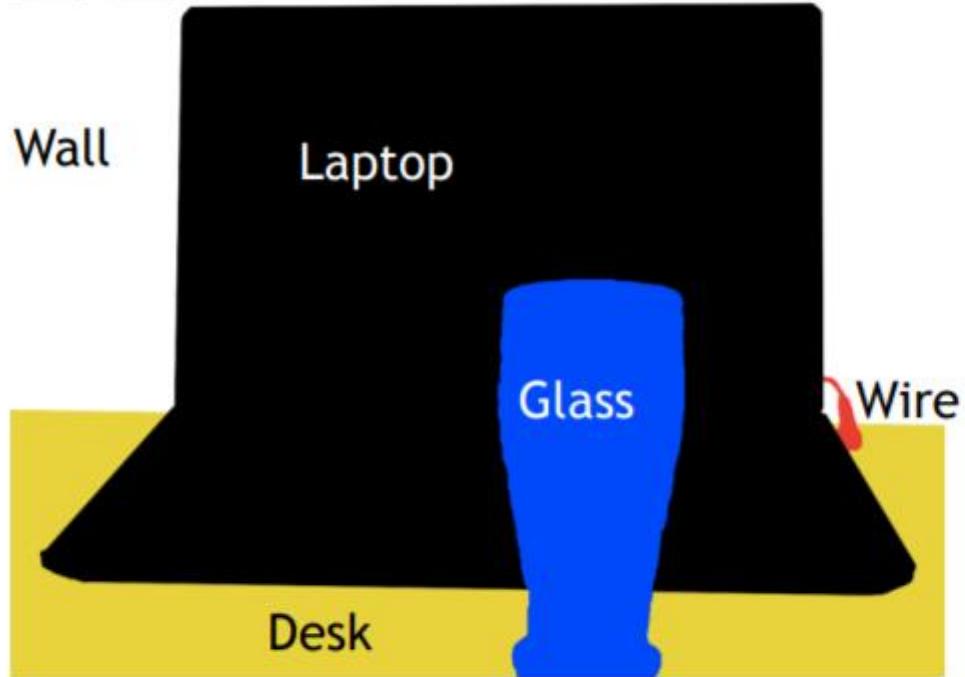
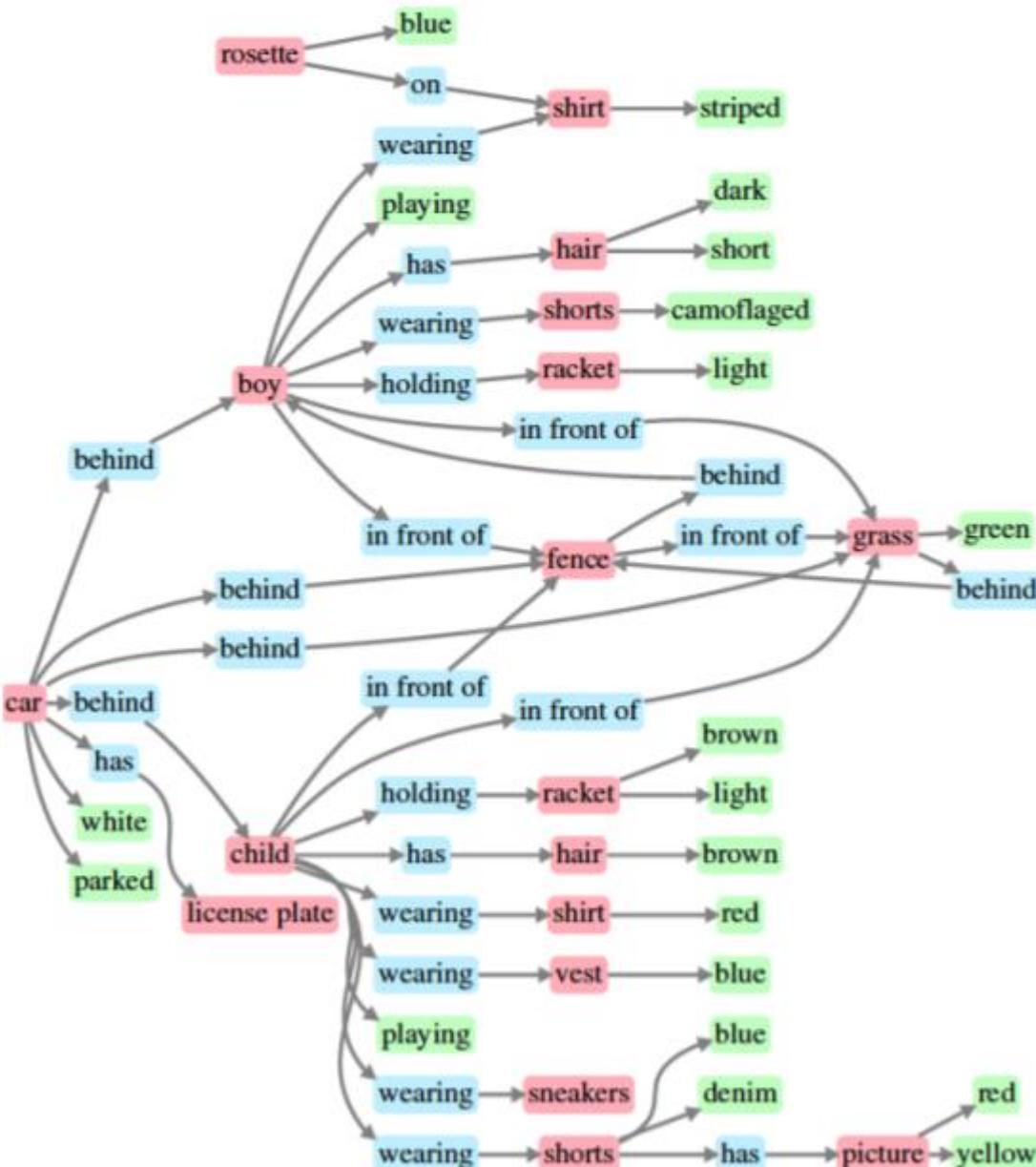
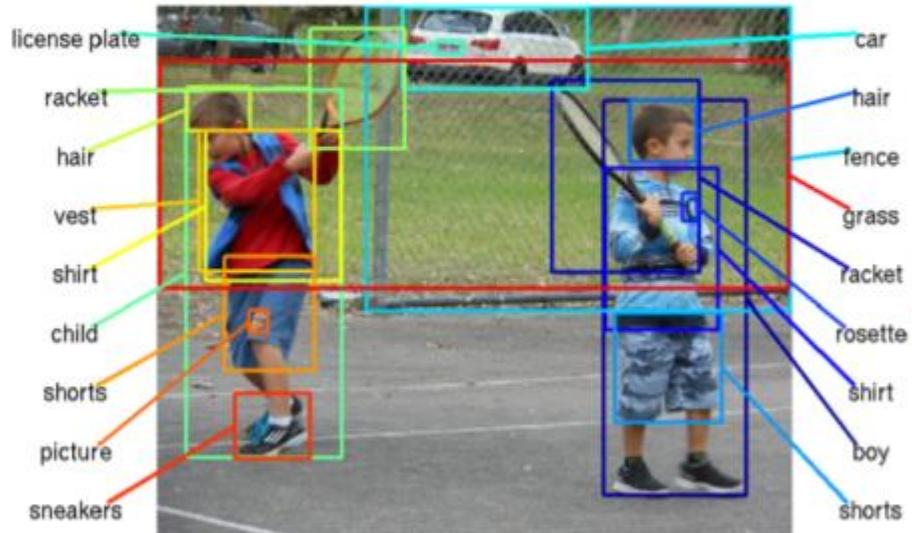


Figure 1: The Transformer - model architecture.

https://keras.io/examples/vision/image_classification_with_vision_transformer/





Johnson *et al.*, “Image Retrieval using Scene Graphs”, CVPR 2015

Figures copyright IEEE, 2015. Reproduced for educational purposes

PT = 500ms



Some kind of game or fight. Two groups of two men? The man on the left is throwing something. Outdoors seemed like because i have an impression of grass and maybe lines on the grass? That would be why I think perhaps a game, rough game though, more like rugby than football because they pairs weren't in pads and helmets, though I did get the impression of similar clothing. maybe some trees? in the background. (Subject: SM)



Outside border images, clockwise, starting from top left:

[Image by Pop Culture Geek](#) is licensed under CC BY 2.0; changes made

[Image by the US Government](#) is in the public domain

[Image by the US Government](#) is in the public domain

[Image by Glogger](#) is licensed under CC BY-SA 3.0; changes made

[Image by Sylenius](#) is licensed under CC BY 3.0; changes made

[Image by US Government](#) is in the public domain

Inside four images, clockwise, starting from top left:

[Image](#) is CCO 1.0 public domain

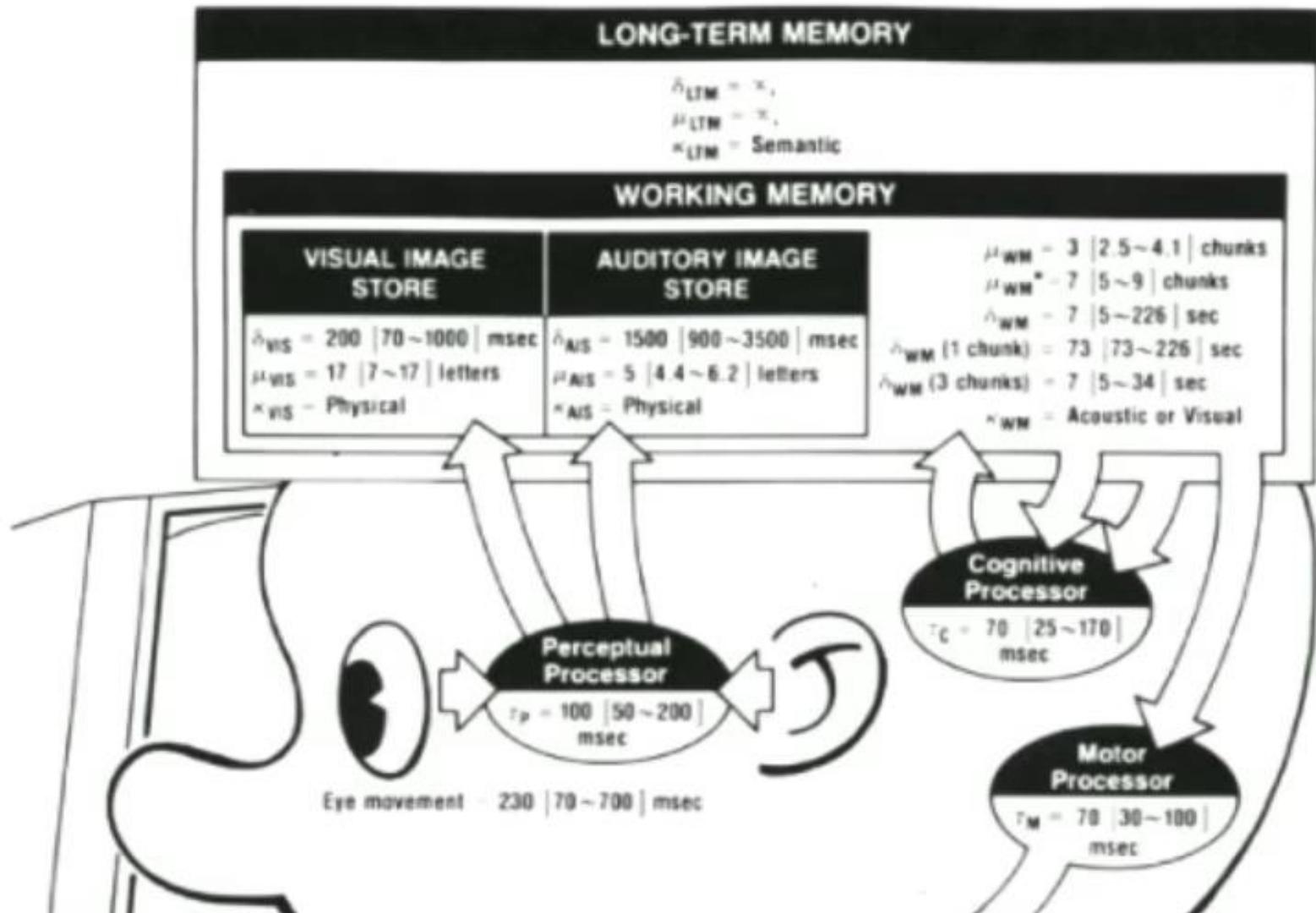
[Image by Tucania](#) is licensed under CC BY-SA 3.0; changes made

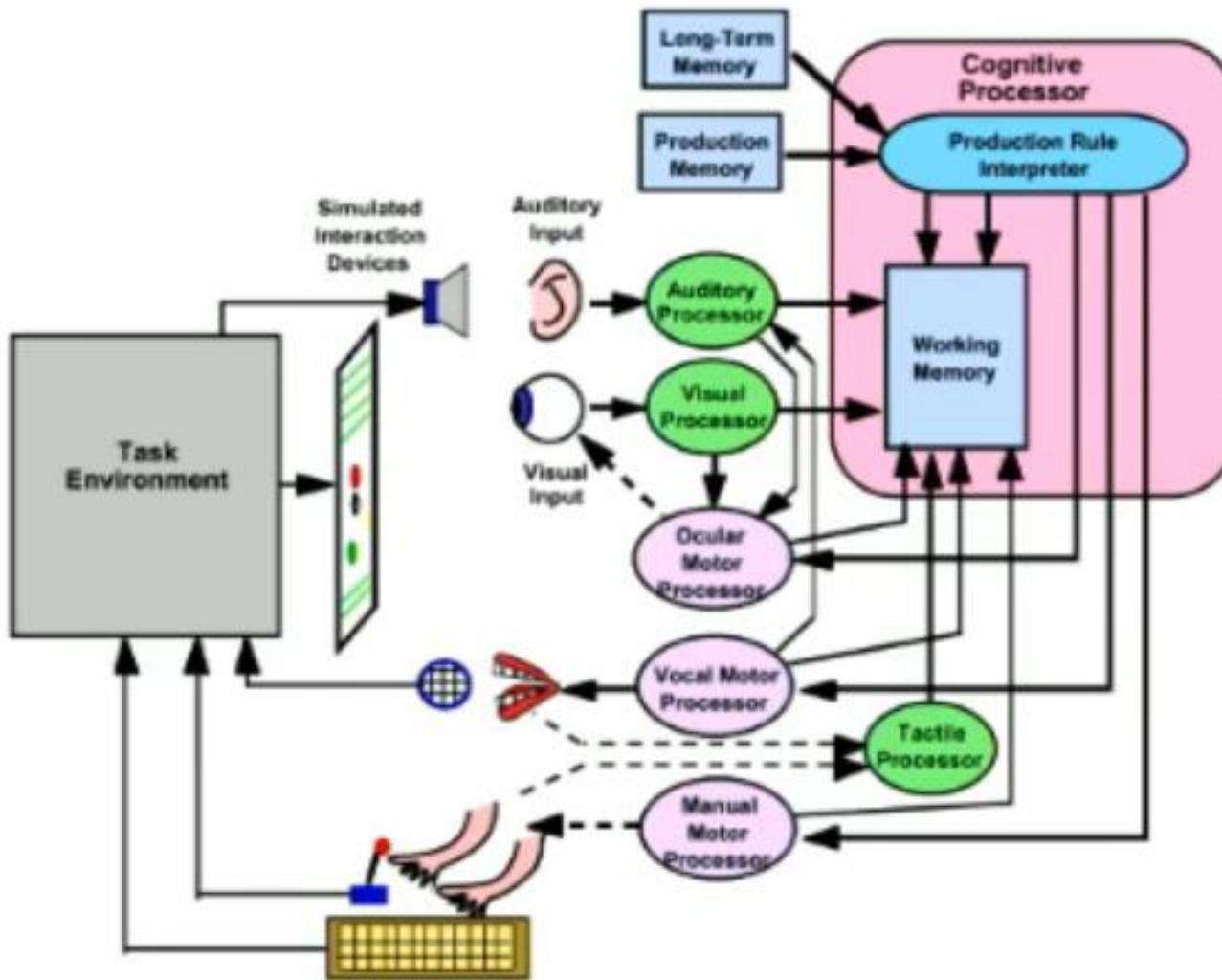
[Image by Intuitive Surgical, Inc.](#) is licensed under CC BY-SA 3.0; changes made

[Image by Oyundari Zorigtbaatar](#) is licensed under CC BY-SA 4.0

Artificial intelligence started as a field whose goal was to replicate human level intelligence in a machine.

인공 지능은 인간 수준의 지능을 기계에 복제하는 것을 목표로 시작되었습니다.





The EPIC architecture
for cognition
and performance
with application to
human-computer interaction

표상 (REPRESENTATION)

원래의 것과 같은 인상을 주는 이미지 또는 형상

- **정신적 표상은** 정신 안에서 비교적 일관되게 재생산되는 의미 있는 사물이나 대상에 대한 지각을 일컫는다.
- **관념적 표상은** 사고나 생각의 토대를 제공하는 정신적 표상으로서, 실질적으로는 정신적 표상과 동일하다.
- **본능적 표상은** 자기 표상 안에 존재하는 개인의 욕동(원본능) 측면들을 말한다.

표상은 자아의 하부 구조를 구성하며 자아 내용물의 일부로 간주된다.

표상을 쉽게 설명하면

우리가 대상에 대한 지식, 정보를 다룬다고 할 때, 실제 대상을 그대로 우리 머릿속으로 가져와서 다루는 것이 아니다. 인간은 실제 대상을 어떤 상징이나 다른 형태로 재표현하여, 즉 추상화하여 다룬다.

이러한 점에서 암, 정보를 '표상(表象; representations)' 이라 한다. 다시 말하여 실물 자체가 아니라, 다시 (re-)나타냄(presentation)의 결과가 우리 마음의 내용이기 때문이다.

예를 들어 우리가 사랑하는 사람을 생각한다고 할 때, 우리의 머리 속에 사랑하는 사람 실물이 들어 있는 것이 아니라, 그 사람에 대한 심상(image)이라던가 다듬어진 생각이나 언어화된 일화나 감정에 대한 기억이 들어있는 것이다. 자동차 한 대, 자동차 세 대라는 생각도 대상 자체가 아니라 심상(Image)이 표상되어서 우리 마음에 남는다고 본다. 즉 실제의 대상이 아니라 '다시-나타내어(표현되어)' 추상화되어진 어떤 내용이 상징으로, 표상으로 우리 마음속에 들어있는 것이다. 마음의 내용들이 곧 표상인 것이다.



사진을 보시고 무엇이 머리에 떠오르시나요?

스프링 – 소라, 다슬기, 용수철, 골뱅이, 짜배기, X 등 각각의 시선으로 바라봄 반면, 작가인 올덴버그는 인도 양 조개에서 모티브로 했고, 다슬기 모양, 한국의 도자기와 한복의 옷고름에서 영감을 받아 제작 했다고 함.⁹⁴



표상이란?

머릿속에 떠오른 무엇.

쇼펜하우어의 위치

플라톤

이데아



현상계

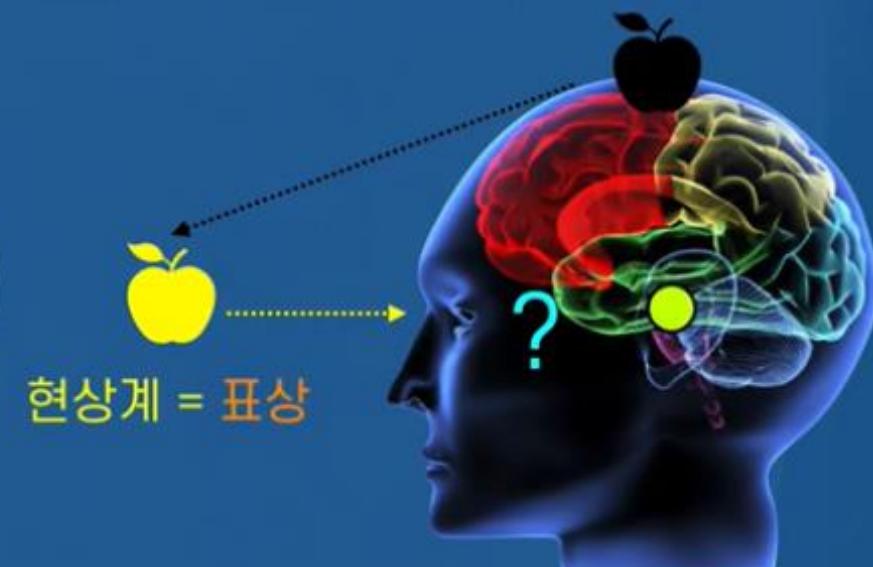
이데아 = 물자체



칸트

쇼펜하우어

이데아 = 물자체 = 의지



현상계 넘어 이데아를 추구하라

인간은 물자체를 인식할 수 없으므로,
현상계에서 보이는 것만
이성으로 올바르게 추구하라

물자체는 욕망이라는 의지이며,
현상계는 의지[욕망]의 표상이므로,
의지를 줄여라



물 자체는 욕망이라는 의지이며, 현상계는 의지[욕망]의 표상이므로, 의지를 줄여라

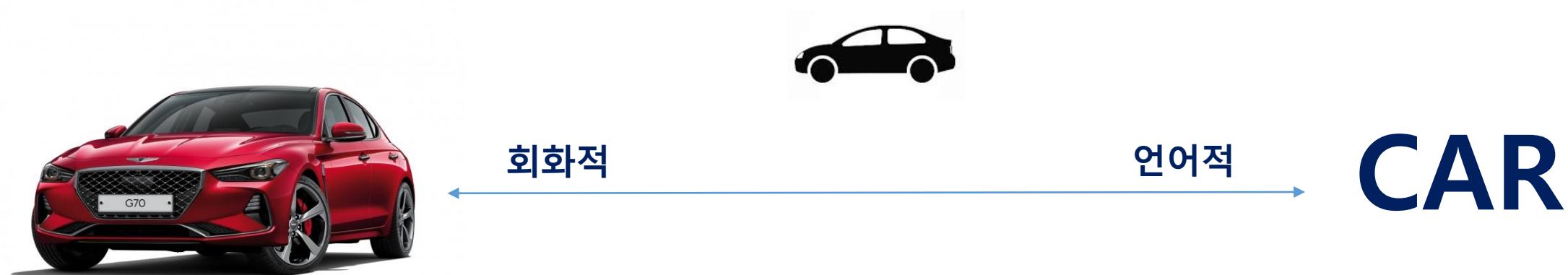
Representation

Optical Character Recognition

심상 표상과 심상 모형

인간은 경험을 통해 사물, 사건, 일, 사람들을 하나씩 표상 함

- 명제적 표상 (propositional representation) : 논리와 기호, 언어 세계와 연관된 표상
- 심상(Mental Imagery) : 대상에 대한 지각적인(perceptual) 특징을 보전하는 특수한 표상 (구체적이고 실제 대상과 근접한 표상, 회화적(picture-like) 표상)
- 심상 모형(mental model) : 위 둘 사이에 존재 (각 개인이 그들의 경험을 이해하고 설명하기 위해 세우는 지식 구조)



표상(표현: representation)의 강조 (Fodor, 1975)

인간과 컴퓨터가 자극 정보를 어떠한 상징으로 기억에 저장한다는 것은 자극 자체를 저장하는 것이 아니라 자극에 대한 표상(표현)을 저장하는 것이며 이는 마음과 컴퓨터 모두가 자극의 정보를 내적 기호(상징)로 변화시켜 기억에 보유한다는 것이다. 따라서 무엇을 안다는 것은 이들 **표상간의 연관을 찾거나 새로운 관계성을 만들어 낸다는 것을 의미한다.** 따라서 암의 과정에 대한 연구는 자극들이 어떻게 상징(기호) 표상들로 전환되고 또 활용 되는가를 연구하는 것이라 하겠다. 즉 **인지과학의 핵심 연구주제는 마음이나 컴퓨터에서의 표상의 처리과정(계산)과, 표상의 본질 및 그 구조적 특성의 연구라고 할 수 있다.**

핵심어 -> 연관성과 추상화

인간들이 보여주는 다양한 종류의 의사 결정 행동과 문제 풀이 행동 또 학습 등은 내적 표상의 도움이 없이는 불가능한 듯 보이기 때문이다. 인간에게 내적 표상 체계가 있음을 주장하며 포더가 드는 예는 체스 경기이다. 포더에 따르자면, 체스 경기를 할 때 복잡한 수를 마음 속으로 이리 저리 두어볼 수 있다는 사실은 우리에게 체스 경기를 표현할 수 있는 표상 체계가 존재함을 증명하는 것이다. 더 나아가 포더는 학습을 위해서도 내적 표상이 요구됨을 지적한다. 학습에는 마음 속으로 가설을 형성하고 수정하는 작업이 요구되는데, 내적 표상이 없다면 그런 가설을 표현할 수 없기 때문이다.

표상의 존재론을 수용하고 나면, 이제 풀어야 문제는 그 같은 표상이 과연 어떤 구조를 가지고 있는가 하는 것이다. 이 문제에 대한 표상적 심리론의 표준적 견해는 표상의 구조가 마치 자연 언어 같다는 것이다. 이른바 **사고 언어 가설**(language of thought hypothesis)로 불려지는 이같은 생각은 포더에 의해 최초로 제안되었으며, 오늘날에는 대부분의 표상적 심리론자들에 의해 어떤 형태로든 수용되고 있다.

포더는 특히 사고 언어 가설

인간의 언어 습득과 의사소통의 핵심적 국면을 설명하기 위한 일종의
‘최선의 설명’으로서 요구된다는 점을 강조함.

사고언어가설(language of thought hypothesis)과 계산주의 마음이론(computational theory of mind)을 주창한 철학자다. 사고언어가설이란 인간의 마음이 언어 구조와 동일한 구조를 가졌다는 이론이다. 인간의 자연 언어는 의미론적 성질과 통사적 성질을 갖는 언어들로 이루어져 있으며 이 두 성질을 동시에 갖는 것들은 기호라고 불린다. 이 가설에 따르면 인간의 인지는 자연언어가 의미론적 성질과 통사적 성질을 갖는 것과 마찬가지로 동일한 구조를 가졌다. 계산주의 마음이론은 이런 사고언어가설을 토대로 하고 있다. 즉 마음(인지)는 앞에 설명한 바와 같은 언어적 구조를 가졌는데 이 구조는 기호 논리학에서의 추론 과정의 형식을 지녔으며 단일한 명제들의 결합을 통해 부분에서 전체로 인지적 구조가 형성된다. 계산주의 마음이론은 컴퓨터나 인간이나 이런 식의 정보처리시스템을 가졌다고 보는 철학적 입장이다.

흥미로운 점은 포더는 초창기 본인이 주장하던 계산주의 마음이론을 비판했다는 것이다. 포더에 따르면 마음은 계산주의 마음이론에서 설명하는 것처럼 단순하지 않으며 (대중에서의 인공지능에 대한 의견과 다르게) 현대 인지과학은 마음의 본질을 밝히려면 한참 멀었다고 한다.

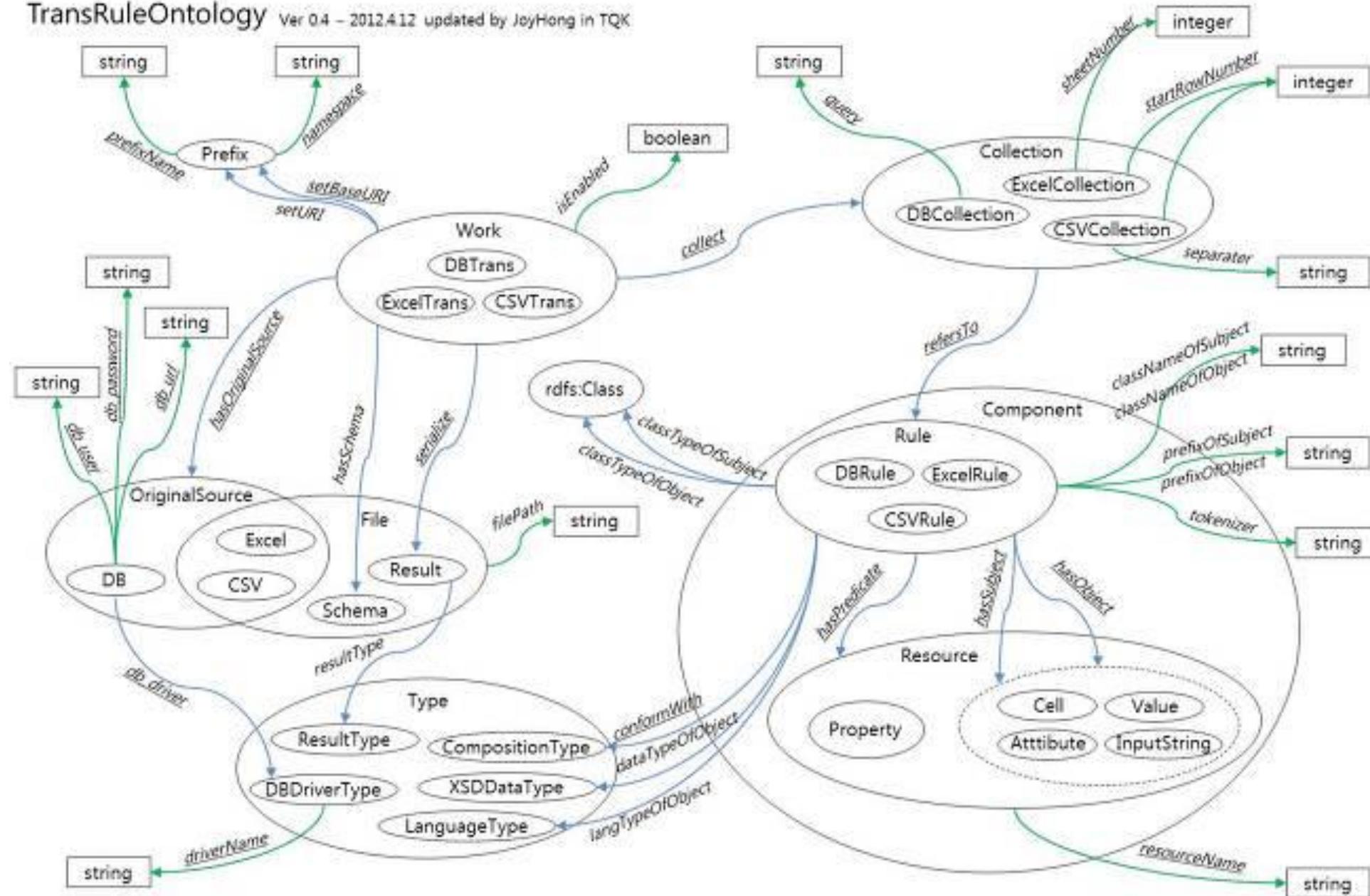
그렇다고 계산주의 마음 이론을 완전히 부정한 것은 아니다. 포더의 요지는 말단 인지는 계산주의로 설명할 수 있지만, 중앙 인지는 그렇지 못하다는 것이다. 중앙 인지가 계산주의로 설명될 수 있는지의 문제는 진화심리학과도 관련이 있다. 계산주의는 인지가 모듈적이라는 주장과 연관이 있는데, 포더는 말단 인지는 모듈적이지만 중앙 인지는 모듈적이지 않다고 주장하고 진화심리학자들은 말단 인지뿐만 아니라 중앙 인지까지 모듈적이라고 주장한다.^[1] 이와 관련해 포더가 내놓은 책이 '마음은 그렇게 작동하지 않는다'인데, 제목은 진화심리학자인 스티븐 핑커의 책 '마음은 어떻게 작동하는가'를 겨냥한 것이다. 유의할 점은 포더 본인의 자연선택설에 대한 비판적 입장과는 별개로, 진화심리학 비판이 곧 진화론 비판은 아니라는 것이다. 진화심리학 항목을 보면 알 수 있듯이 진화심리학은 기존 진화생물학계에서도 아직 여러 가지 비판을 받고 있는, (적어도 아직까지는) 충분히 발달되지 않은 학문이다.

[출처] <https://namu.wiki/w/%EC%A0%9C%EB%A6%AC%20%ED%8F%AC%EB%8D%94>

정보처리적 패러다임의 인지과학은 마음에 대한 보는 틀을 이와 같이 상정하고 나서, 정보처리체계로서의 마음의 작용을 감각, 지각, 학습, 기억, 언어, 사고, 정서 등의 여러 과정으로 나눈 다음, 각 과정에서 어떠한 정보처리가 일어나는가, 각 과정들은 어떻게 상호 작용 하는가를 묻고, 각 과정에서 어떠한 정보(지식)구조, 즉 표상(표현)구조가 관련 되는가를 규명하려 한다. 따라서 마음의 현상, 즉 심리적 사건은 정보의 내용 및 정보를 처리하는 사건으로 개념화 되어지는 것이다.

TransRuleOntology

Ver 0.4 – 2012A12 updated by JoyHong in TQK



노만(D. Norman), HCI에 심성 모형의 개념을 처음 도입

컴퓨터와 관련되어 사람들이 갖고 있는 심성 모형은 대개 부정확(1983) 좋은 인터페이스 제공을 통해 사용자의 심성 모형이 정확히 형성될 수 있게 도와준다면 (인지적) 사용성은 올라갈 것이라 함.

Norman이 제시한 세 개의 다른 모형

심성 모형 (mental models)

- 사용자가 시스템을 사용하는데 필요한 표상

개념 모형 (conceptual models)

- 시스템 디자이너가 가지고 있는 시스템의 정확한 모형
- 심성 모형과 개념 모형은 다를 수 밖에 없고, 사용자들 사이의 심성 모형도 다름

시스템 이미지 (system image)

- 사용자와 디자이너의 대화는 시스템을 통해서 발생(시스템 = 매개자)
- 시스템의 인터페이스/동작/반응을 통해 둘 사이의 소통이 이루어짐
- 시스템이 사용자에게 줄 수 있는 중요한 표상은 외형 표상



표상 학습(Representation learning)

표상 학습(representation learning) 혹은 특성 학습(feature learning)은 직접적인 데이터 대신, 유용한 정보를 더 쉽게 추출할 수 있게 만들어진 데이터의 표상(representation)을 통해 분류기나 다른 예측 기계를 학습시키는 것을 뜻한다. -> **Neural Networks**

확률적으로 좋은 표상은 입력을 설명할 수 있는 내재적인 설명 요인들의 posterior 분포를 포함할 때가 많다. 대화 인식, 신호 처리, 물체 인식, 자연어 처리와 같은 분야들에서 표상 학습 방법론은 실증적인 성공을 이루어 냈으며, 다양한 확률 모델과 인공 신경망, 그리고 딥러닝이 자동으로 표상 학습을 위한 특성 추출에 이용된다.

DOCUMENTS ①

(Doc #1) John likes to eat apples and oranges. Mary likes oranges.

(Doc #2) Mary likes to eat mellon and watch football.

DICTIONARY ②

{ and
apples
eat
football
John
likes
Mary
mellan
oranges
to
watch }

REPRESENTATIONS ③

Doc #1

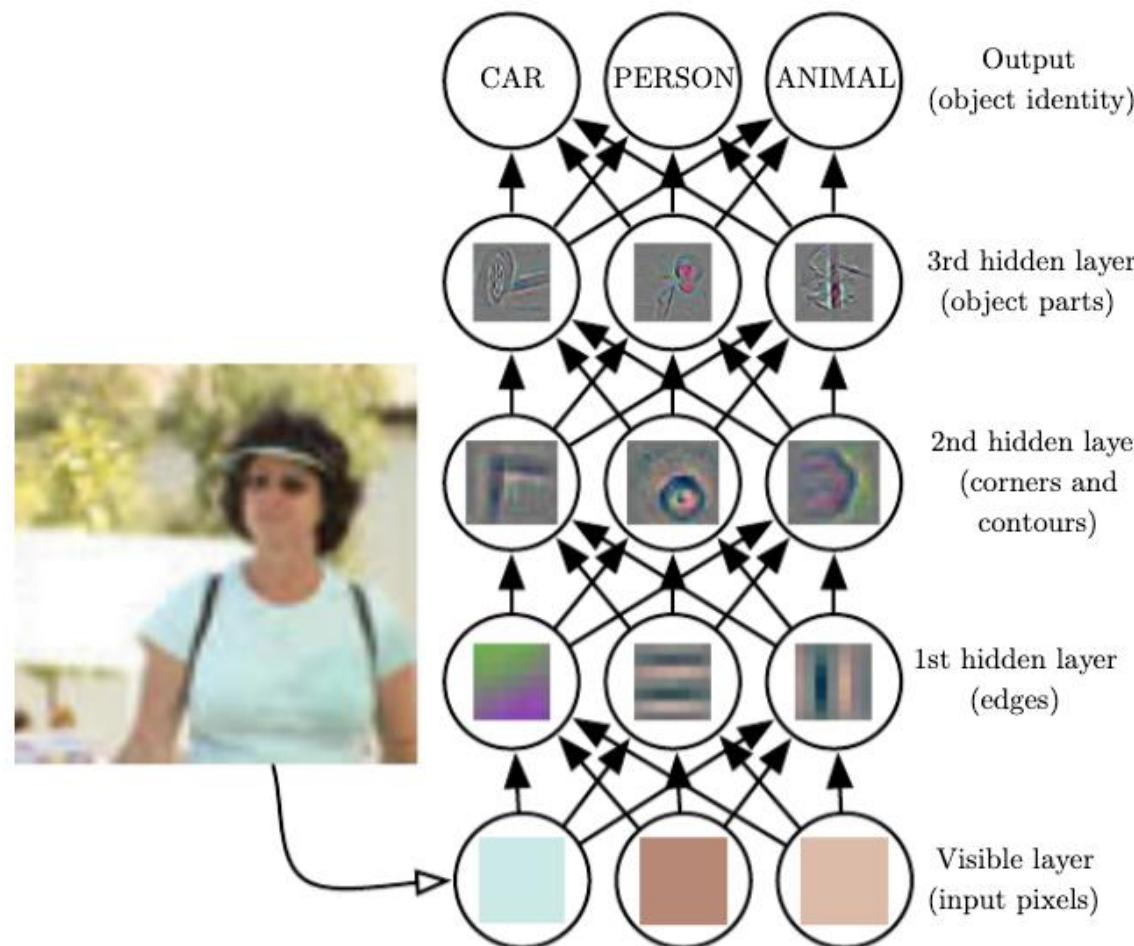
1
1
1
0
1
2
1
0
2
1
0

Doc #2

1
0
1
1
0
1
1
1
0
1
1

From a set of documents, build a dictionary containing the set of unique words, then each document is represented as a feature vector containing the count (the number of times) of each word in that document.

Deep Learning Model



Deep neural networks learn representations by combining simple concepts to derive complex structures in a hierarchical pipeline. Each layer iteratively refines the information from the layers before. In the end, a classifier takes the transformed representations and draws linear boundaries among the classes.



정 준 수 / Ph.D (jsjeong@hansung.ac.kr)

- 前) 삼성전자 연구원
- 前) 삼성의료원 (삼성생명과학연구소)
- 前) 삼성SDS (정보기술연구소)
- 現) (사)한국인공지능협회, AI, 머신러닝 강의
- 現) 한국소프트웨어산업협회, AI, 머신러닝 강의
- 現) 서울디지털재단, AI 자문위원
- 現) 한성대학교 교수(겸)
- 전문분야: Computer Vision, 머신러닝(ML), RPA
- <https://github.com/JSeong-me/>