초격차 패키지 Online.

# 최신 논문

## High-level Computer Vision

Ch. 1 | Context Understanding

High-level computer vision을 위한 context 이해

Ch. 2 | Scene Understanding

Visual data의 또 다른 표현: Scene

Ch. 3 | Human Motion Understanding

High-level vision 관점에서의 human motion 연구

Ch. 4 | Video Understanding

Temporal 축이 추가된 visual data

# 강사 소개

## 이 준 현

**[Author]**
ICML, CVPR, NeurIPS
(1저자 및 공저자)

**[Reviewer]**
ICML, NeurIPS, ICLR 등

**[Program Committee]**
WWW Workshop

**[Community Organizer]**
GNN KR

**[Researcher]**
LAB145

https://www.facebook.com/groups/gnnkr
https://www.facebook.com/LAB145

GraphLearning @TheWebConf 2022

### PROGRAM COMMITTEE MEMBERS

Saloni Agarwal, University of Texas at Dallas
Ariful Azad, Indiana University Bloomington
Lei Bai, University of Sydney
Tanmoy Chakraborty, Indraprastha Institute of Information Technology Delhi
Michael Cochez, Vrije Universiteit Amsterdam
Tyler Derr, Vanderbilt University
Falih Febrinanto, Federation University Australia
Mingliang Hou, Dalian University of Technology
Zhao Kang, University of Electronic Science and Technology of China
Seyed Mehran Kazemi, Google Research
Zekarias Kefato, KTH Royal Institute of Technology
Junhyun Lee, Korea University
Radosław Michalski, Wrocław University of Science and Technology
Shirui Pan, Monash University
Chanyoung Park, Korea Advanced Institute of Science and Technology
Ciyuan Peng, Federation University Australia
Jonas Richiardi, University of Lausanne
Tara Safavi, University of Michigan
Vivek Sharma, MIT
Ke Sun, Dalian University of Technology
Pengyang Wang, University of Macau
Shan Xue, University of Wollongong
Leo Yu Zhang, Deakin University

Lab145
1월 30일 ·

[What does 2022 hold for Geometric & Graph ML?]
딥마인드의 Petar Veličković과 Michael Bronstein 교수의 글입니다.
레퍼런스 논문들도 함께 기재되어있으니 원글도 보시면 좋을 것 같습니다.

1. (Differential) Geometry는 ML에서 점점 더 중요해지고 있습니다.

2. Message passing은 여전히 GNNs에서 주요 패러다임입니다.
Graph 상에서 convolution을 구현하는 방법중 하나로 여겨지고 사용되어온 message passing 방법이 expressiveness등의 관점에서 근본적인 한계점이 있다고 약 2020년부터 지적되어왔습니다.
그로인해 다양한 오퍼레이션들이 제안되어왔지만, 여전히 주요 패러다임으로 인정받고 있습니다 (오히려 단점을 보완하여 더 나은 message passing을 만드는 방법론도 제안되고 있습니다).

3. NeuralODEs로 시작된 흐름이 그래프에도 확장되고 있습니다.
몇몇 연구들이 GNN 모델을 discretisations of continuous differential equation로 수식화 하는데, 이는 GNN의 고질적인 문제였던 oversmoothing과 oversquashing에 효과가 있습니다.

4. 신호처리, 뇌과학, 물리학 등에서 나온 예전 아이디어들이 재조명 받고 있습니다.
Graph signal processing은 spectral graph neural networks 등으로 Graph ML에서 주요하게 사용되어왔습니다.
이 외에도 고전적인 방법론들이 최신 모델들과 결합되어 새로운 결과를 보여주고 있습니다.

5. 복잡계를 모델링 ...
behaviours 때문에 ...
됩니다.

6. Reasoning, axio...
question입니다.
(특히 Petar는 근 ...
샵도 개최할 정도 ...

...ab145
... 22일 ·
...nsformer보다 좋은 게 나왔다고? : MSRA의 Swin Transformer
...transformer #ObjectDetection #Segmentation #videoaction
...omputervision
...요! 오늘 소개해 드릴 논문은 지난 포스팅(Swin Transformer)에
...rosoft Research Asia에서 공개한 Swin Transformer V2 모델입
...에서, Swin Transformer는 컴퓨터비전에서의 dense
...n task도 잘 한다고 이야기 했었었죠! 이번 논문에서는 Swin
...ner를 최대 30억 파라미터까지 확장하고 최대 1,536x1,536 해
...미지를 학습할 수 있는 Scaling up된 모델을 제시한다고 합니
...퓨터비전에서는 도메인 특성상 NLP와 다르게 transformer 기반
...odel에는 다음과 같은 한계가 있는데요!
...vision model에선 학습 시, instability issue가 있다.
...ownstream vision task들은 high resolution image나 large
...window를 요하는데 low resolution에서 pre-train된 model을
...움이 있다.
...ion model을 다음과 같은 방식으로 구현할
...que (layer norm의 위치를 바꿔줘요!)
...proach (단순히 dot product attention이
...lative position bias approach (coordinate
...cation, COCO object detection, ADE20K
...tics-400 video action classification에서
...l 아래 논문과 코드 링크를 달았으니 참고
...다음 포스팅에서 또 뵐게요!

Lab145
1월 20일 ·

[위아래도 없는 Graph에게 위치정보 가르치기 : Yoshua Bengio 공저 논문]
안녕하세요! 오늘 소개해 드릴 논문은 Graph Neural Networks with Learnable Structural and Positional Representations 입니다.
근 몇년간 가장 인기 있는 모델구조인 트랜스포머에서 이미지나 텍스트 데이터에 대해 위치 정보를 넣어주기 위해 사인함수 등으로 positional encoding vector를 만들어서 활용하는데요,
그러나 그래프에서는 위 아래 혹은 왼쪽 오른쪽의 개념이 없기 때문에 이를 고려하기 어렵습니다 (스펙트럴 영역으로 변환하여 고유벡터를 뽑으면 그래프 위에서의 사인함수를 구할 수는 있습니다).
이런 점 때문에 기존 GNN들은 isomorphic nodes나 graph symmetries에 대한 표현력(expressiveness)이 제한되었습니다.
본 논문에서는 이를 해결하기 위해 Random walk 기반의 structural and positional representation을 학습하는 방법을 제안합니다.
타 모델들과 결합하였을때, 일관된 성능향상을 보이며 기존의 linear complexity를 유지합니다.
아래에 1저자인 Vijay Prakash Dwivedi의 발표 영상과 슬라이드도 첨부드리니 더 관심 있으신 분들은 참고하시면 좋을 것 같습니다.

논문 : https://arxiv.org/abs/2110.07875
슬라이드 : https://hannes-stark.com/.../VPDwivedi_GNN_LSPE_LoGaG...
비디오 : https://www.youtube.com/watch?v=fft2Q0jEWi0

# High-level Computer Vision 이란?



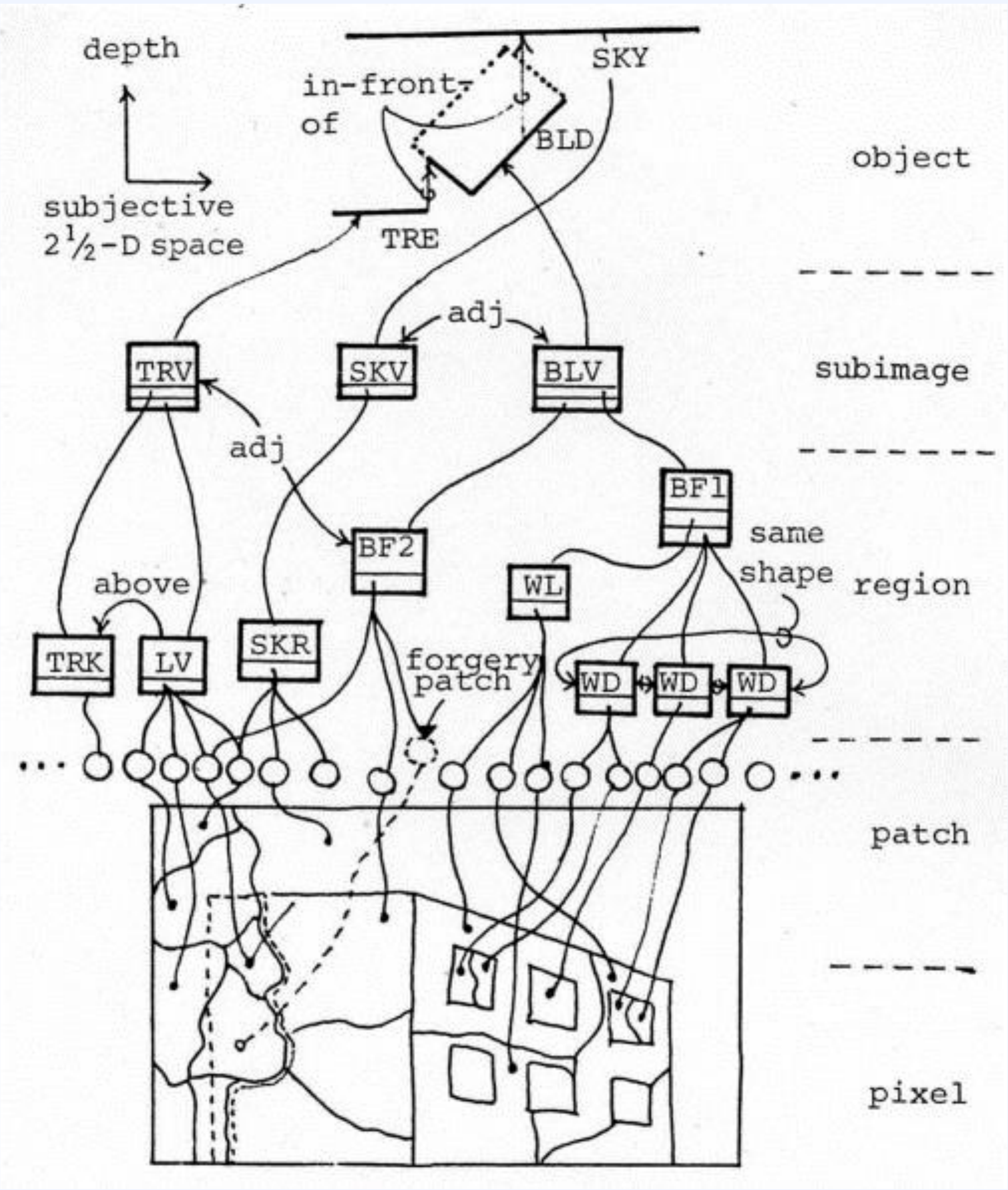
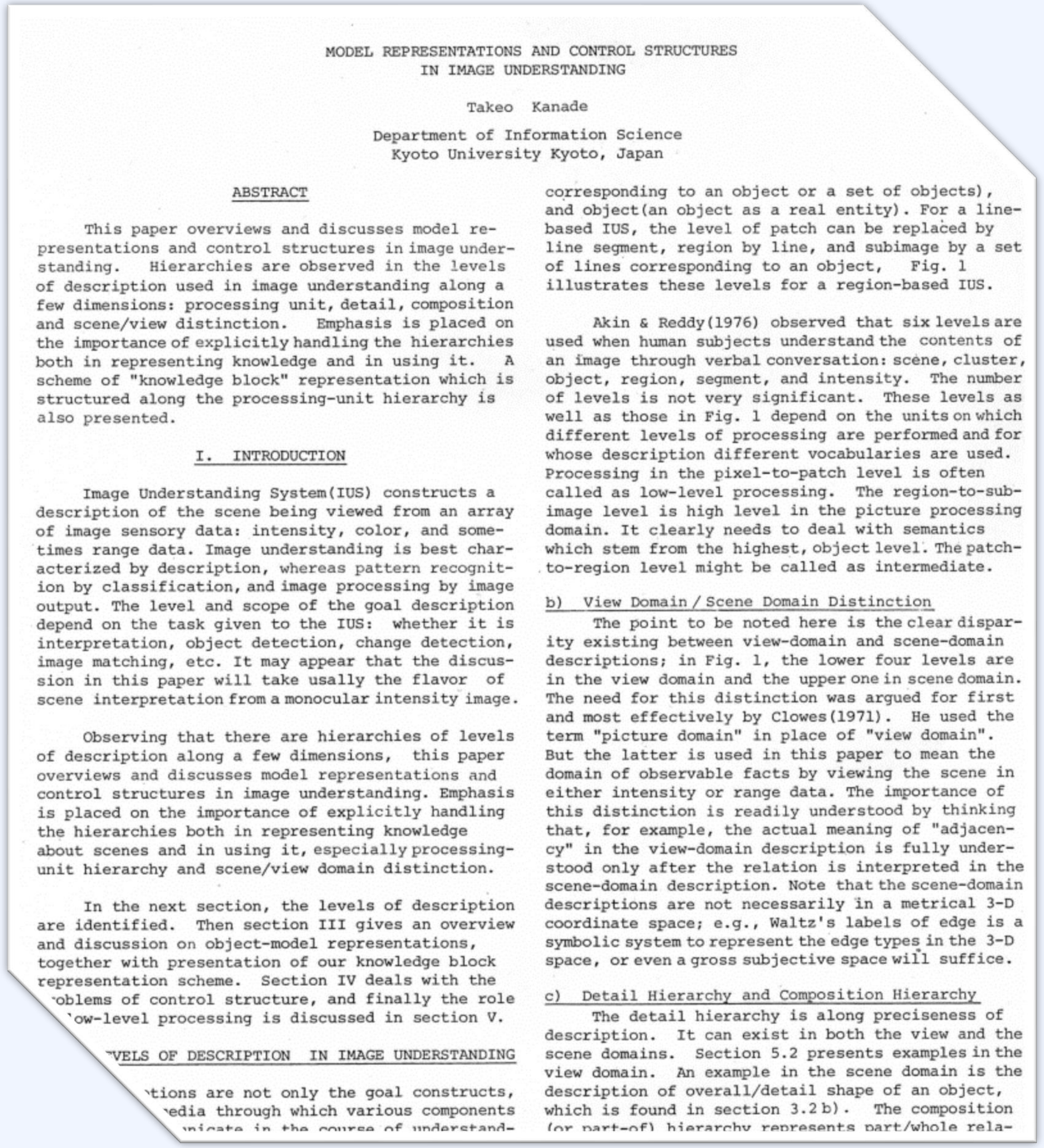



Illustration of Levels of Description in Processing-Unit Hierarchy



MODEL REPRESENTATIONS AND CONTROL STRUCTURES
IN IMAGE UNDERSTANDING

Takeo Kanade

Department of Information Science
Kyoto University Kyoto, Japan

ABSTRACT

This paper overviews and discusses model representations and control structures in image understanding. Hierarchies are observed in the levels of description used in image understanding along a few dimensions: processing unit, detail, composition and scene/view distinction. Emphasis is placed on the importance of explicitly handling the hierarchies both in representing knowledge and in using it. A scheme of "knowledge block" representation which is structured along the processing-unit hierarchy is also presented.

I. INTRODUCTION

Image Understanding System(IUS) constructs a description of the scene being viewed from an array of image sensory data: intensity, color, and sometimes range data. Image understanding is best characterized by description, whereas pattern recognition by classification, and image processing by image output. The level and scope of the goal description depend on the task given to the IUS: whether it is interpretation, object detection, change detection, image matching, etc. It may appear that the discussion in this paper will take usally the flavor of scene interpretation from a monocular intensity image.

Observing that there are hierarchies of levels of description along a few dimensions, this paper overviews and discusses model representations and control structures in image understanding. Emphasis is placed on the importance of explicitly handling the hierarchies both in representing knowledge about scenes and in using it, especially processing-unit hierarchy and scene/view domain distinction.

In the next section, the levels of description are identified. Then section III gives an overview and discussion on object-model representations, together with presentation of our knowledge block representation scheme. Section IV deals with the ·oblems of control structure, and finally the role ·ow-level processing is discussed in section V.

·VELS OF DESCRIPTION IN IMAGE UNDERSTANDING

·tions are not only the goal constructs, ·edia through which various components ·nicate in the course of understand-

corresponding to an object or a set of objects), and object(an object as a real entity). For a line-based IUS, the level of patch can be replaced by line segment, region by line, and subimage by a set of lines corresponding to an object, Fig. 1 illustrates these levels for a region-based IUS.

Akin & Reddy(1976) observed that six levels are used when human subjects understand the contents of an image through verbal conversation: scene, cluster, object, region, segment, and intensity. The number of levels is not very significant. These levels as well as those in Fig. 1 depend on the units on which different levels of processing are performed and for whose description different vocabularies are used. Processing in the pixel-to-patch level is often called as low-level processing. The region-to-subimage level is high level in the picture processing domain. It clearly needs to deal with semantics which stem from the highest, object level. The patch-to-region level might be called as intermediate.

b) View Domain / Scene Domain Distinction

The point to be noted here is the clear disparity existing between view-domain and scene-domain descriptions; in Fig. 1, the lower four levels are in the view domain and the upper one in scene domain. The need for this distinction was argued for first and most effectively by Clowes(1971). He used the term "picture domain" in place of "view domain". But the latter is used in this paper to mean the domain of observable facts by viewing the scene in either intensity or range data. The importance of this distinction is readily understood by thinking that, for example, the actual meaning of "adjacency" in the view-domain description is fully understood only after the relation is interpreted in the scene-domain description. Note that the scene-domain descriptions are not necessarily in a metrical 3-D coordinate space; e.g., Waltz's labels of edge is a symbolic system to represent the edge types in the 3-D space, or even a gross subjective space will suffice.

c) Detail Hierarchy and Composition Hierarchy

The detail hierarchy is along preciseness of description. It can exist in both the view and the scene domains. Section 5.2 presents examples in the view domain. An example in the scene domain is the description of overall/detail shape of an object, which is found in section 3.2 b). The composition (or part-of) hierarchy represents part/whole rela-

Model Representations and Control Structures in Image Understanding, Takeo Kanade, IJCAI 1977

# High-level Computer Vision 이란?



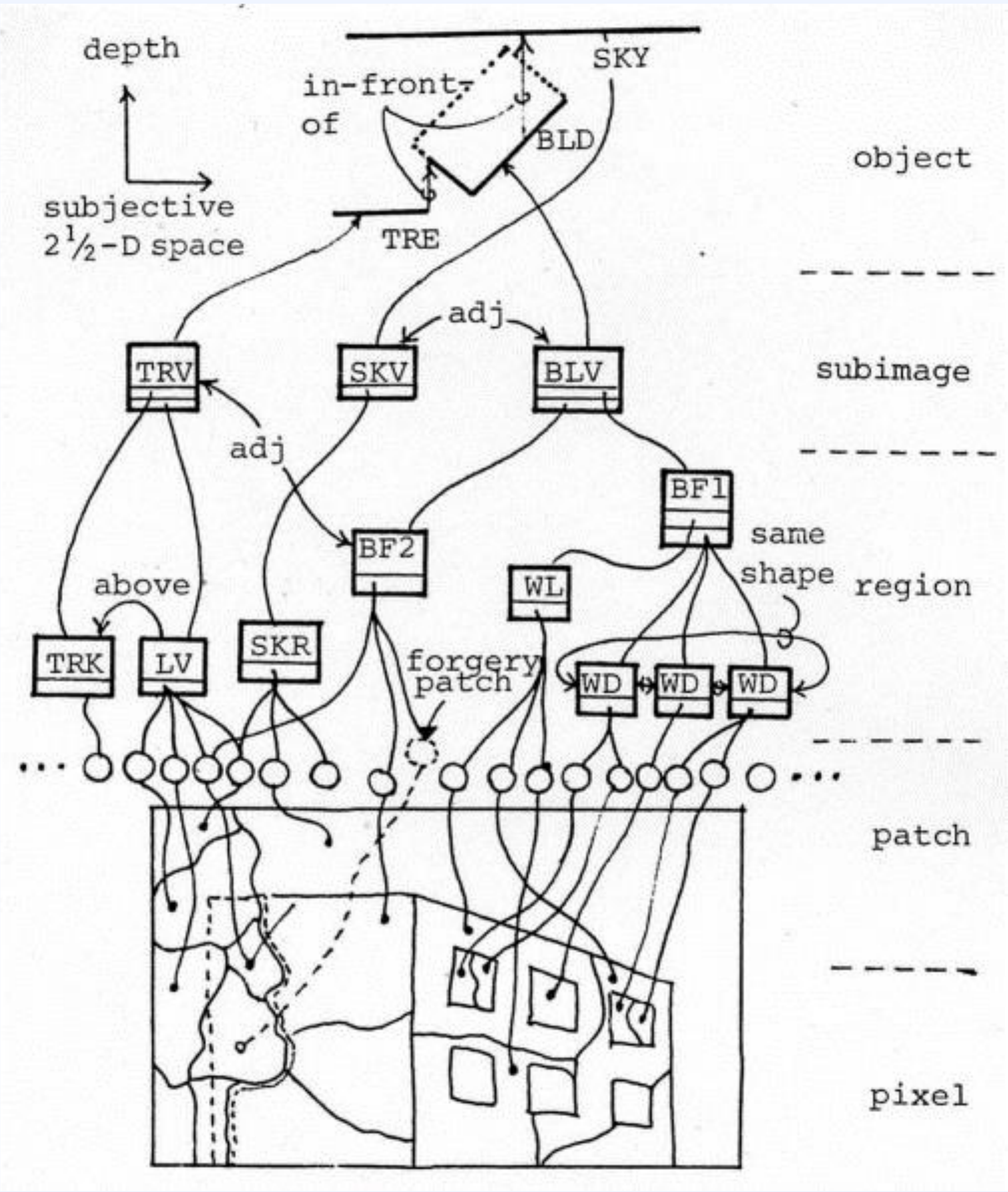


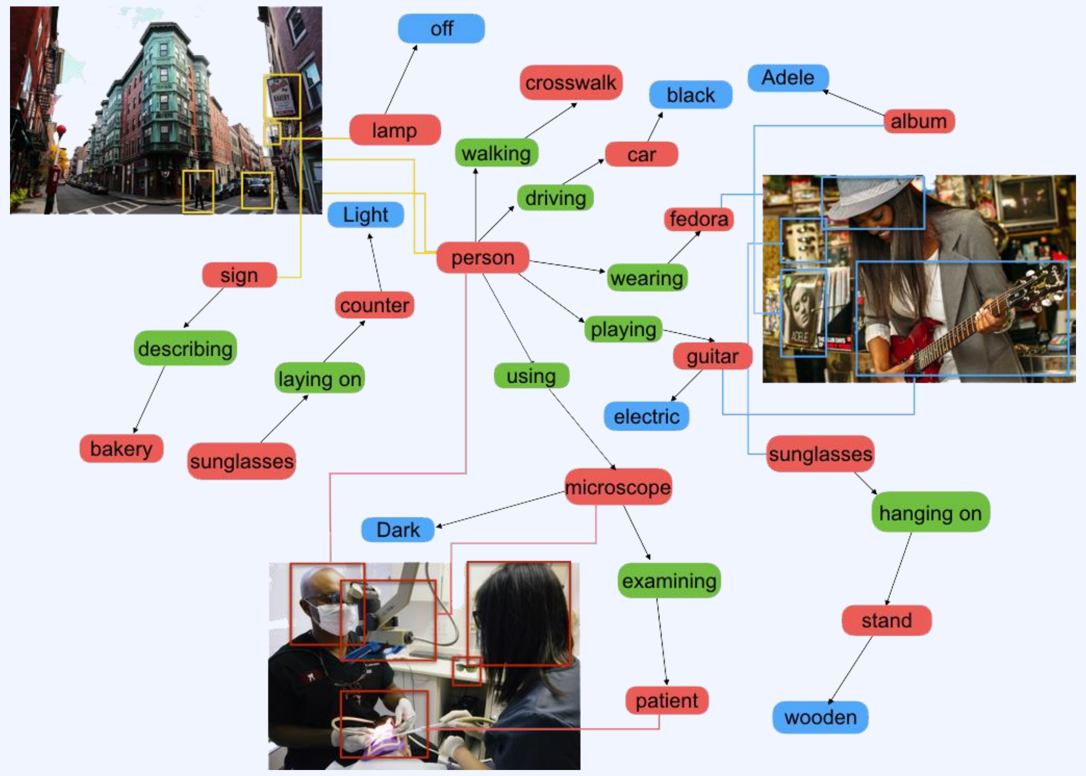

Illustration of Levels of Description in Processing-Unit Hierarchy

High

Mid

Low

Compositionality, Scene understanding…




Visual Genome

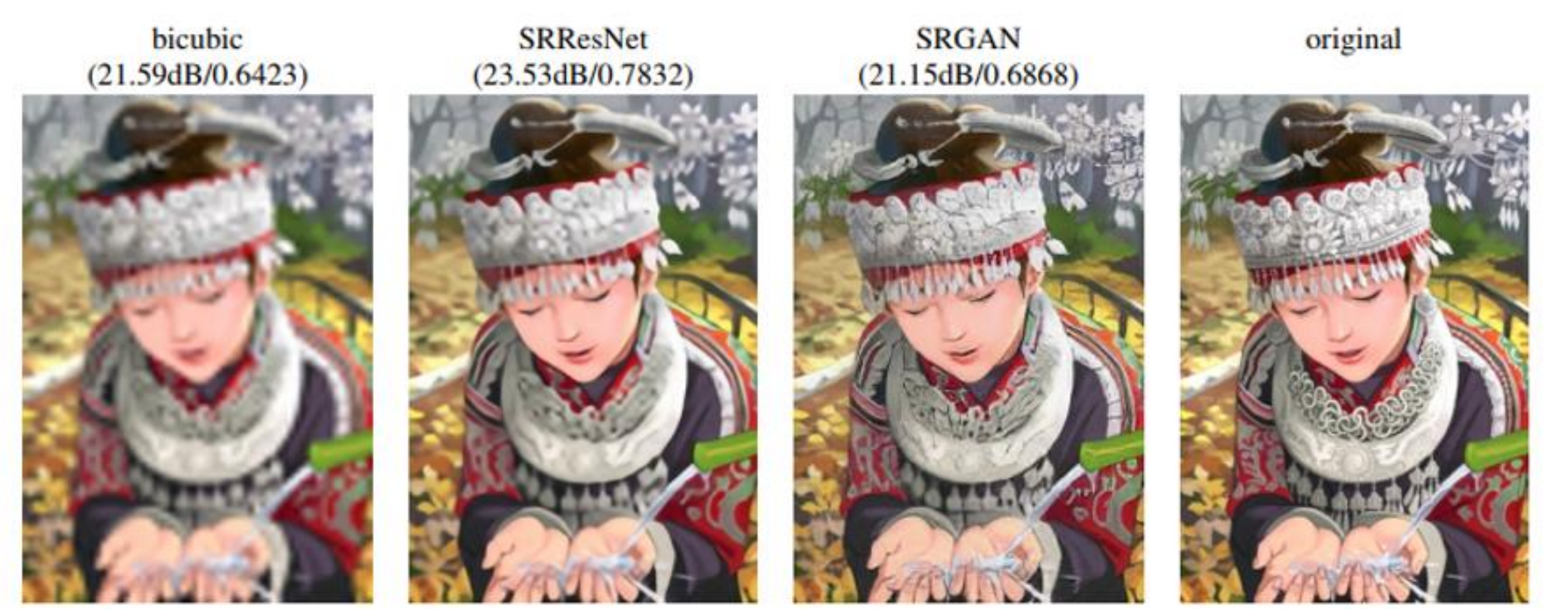



Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, Ledig et al., CVPR 2017 (oral)

Image processing, Feature detection, and Early segmentation…

# High-level Computer Vision 이란?



**이런 키워드들과 연관이 있어요!**

Compositionality

Semantic Interpretation

Cognitive Computer Vision

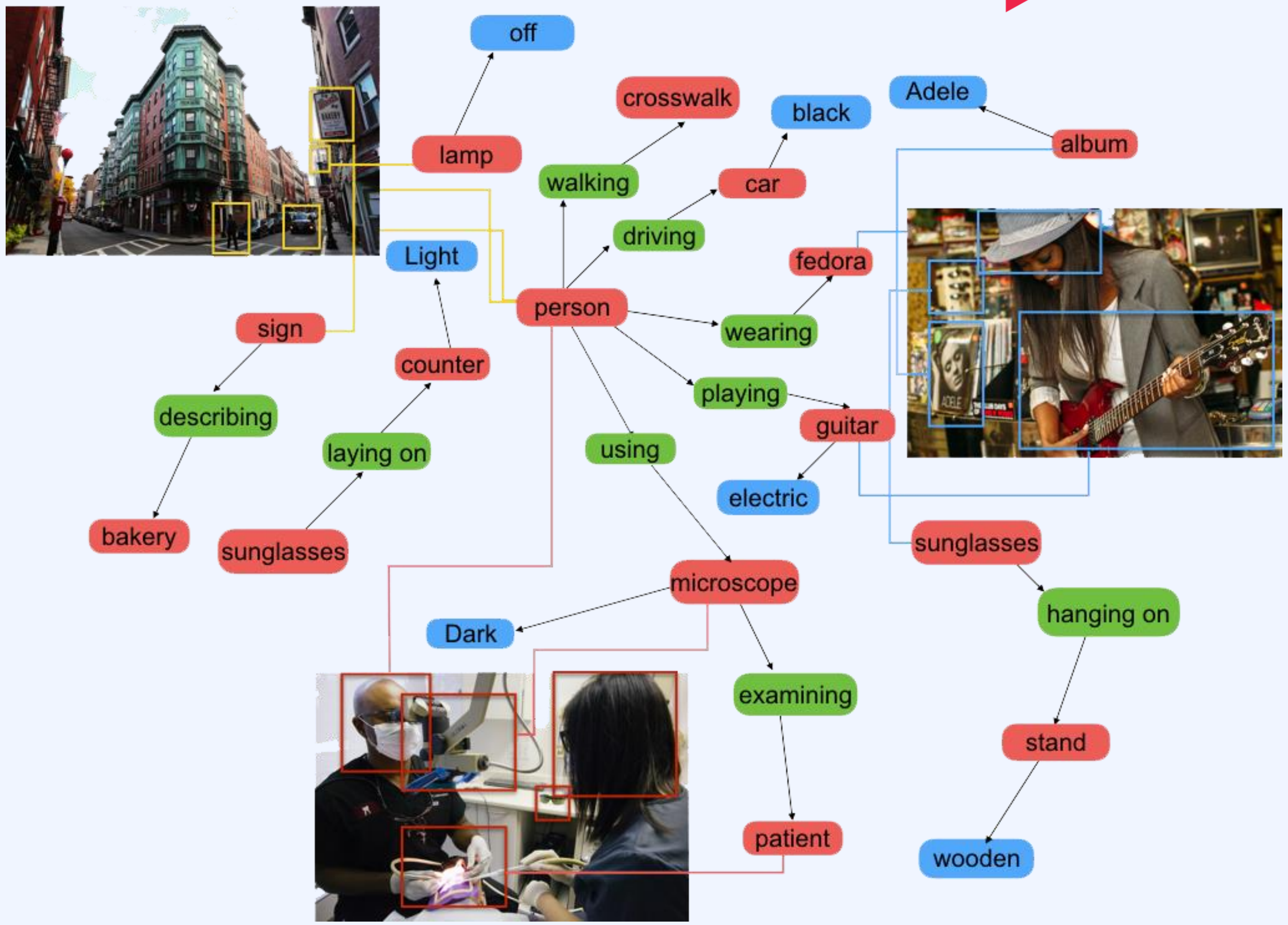
Scene Graph

Action recognition

Symbolic AI

Reasoning

⋮




Visual Genome
(https://visualgenome.org/)

# Context Understanding

## 1 Overview

# Context Understanding
## Overview



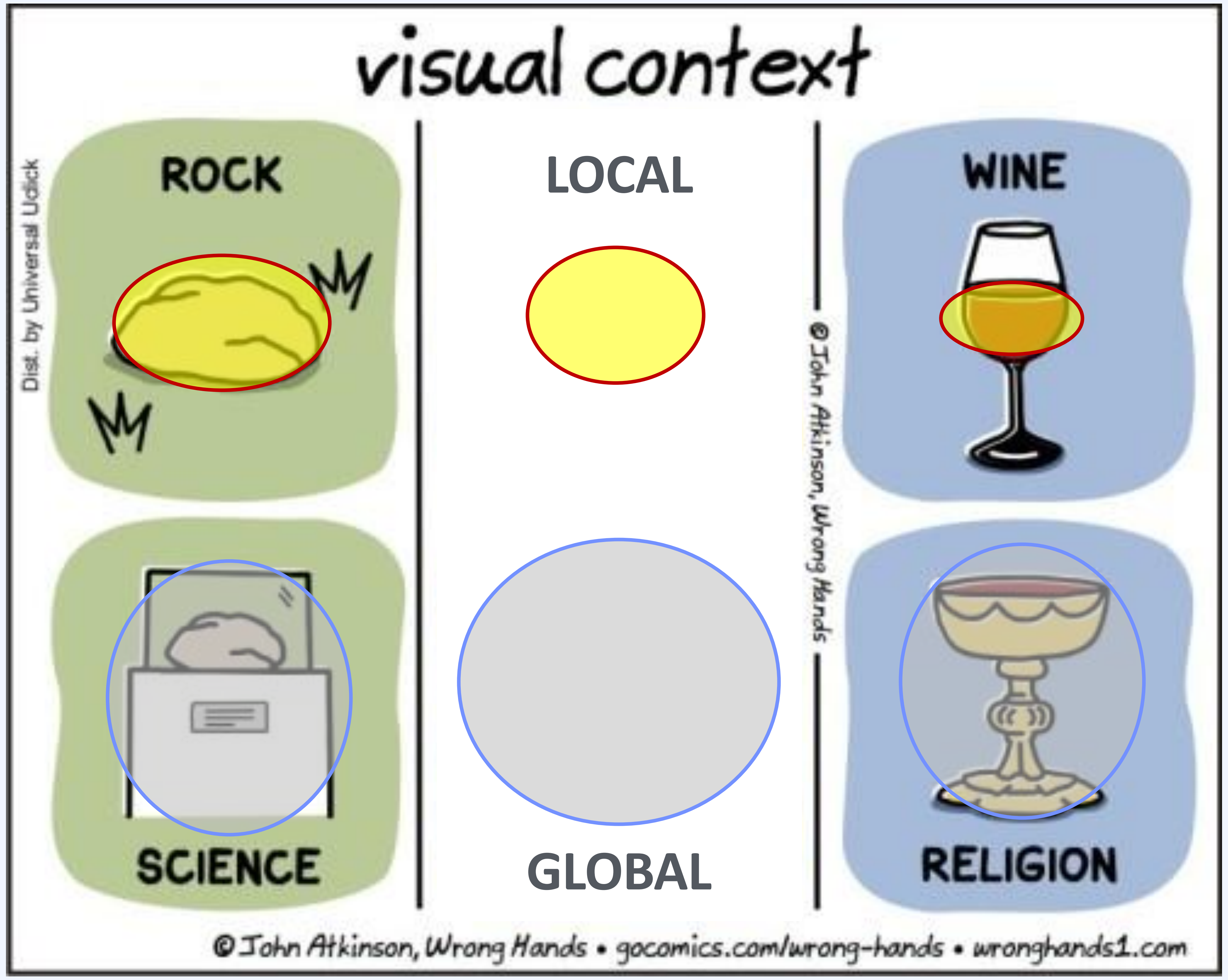

# Context Understanding
## Overview

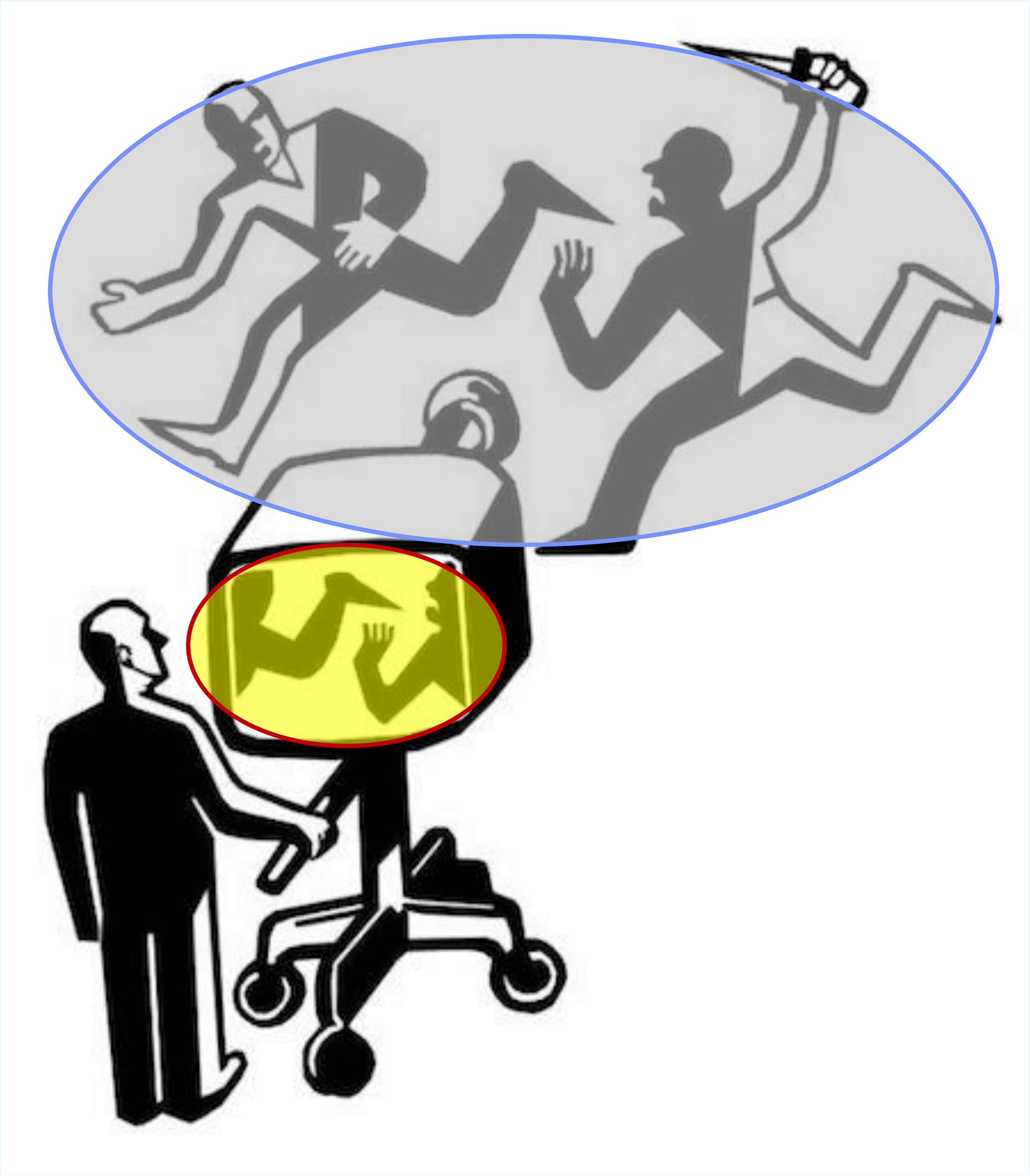

마르셀 뒤샹, <샘>, 1917 /
Wikimedia Commons

Context Understanding
Overview



**Local feature가 같더라도 global context 에 따라 scene의 의미가 달라질 수 있습니다.**



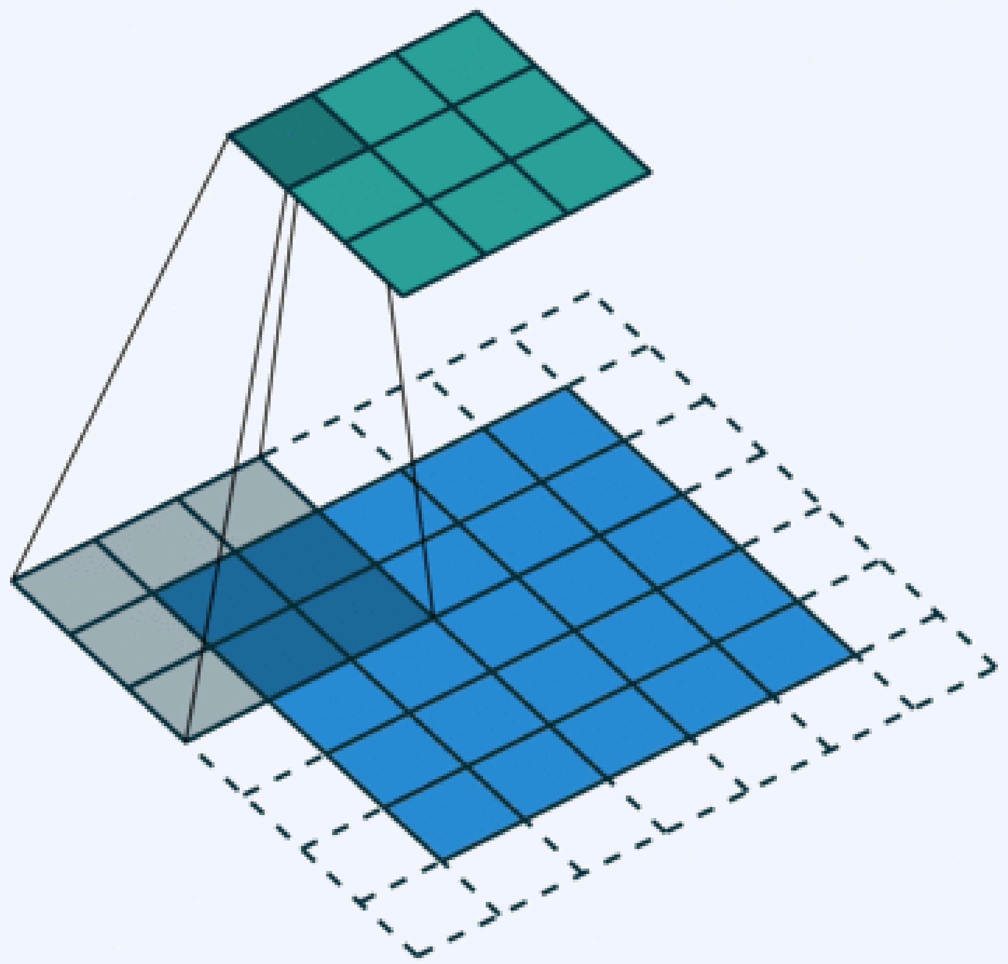



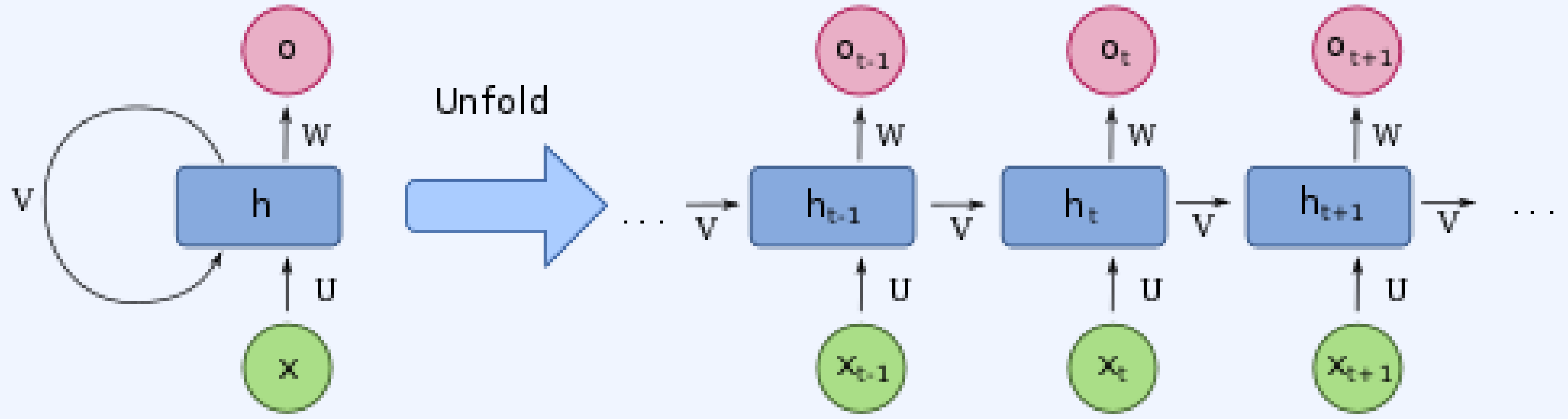




## 과거의 Deep learning 주류 모델들 : inductive bias



Convolution 연산 /
Wikimedia Commons



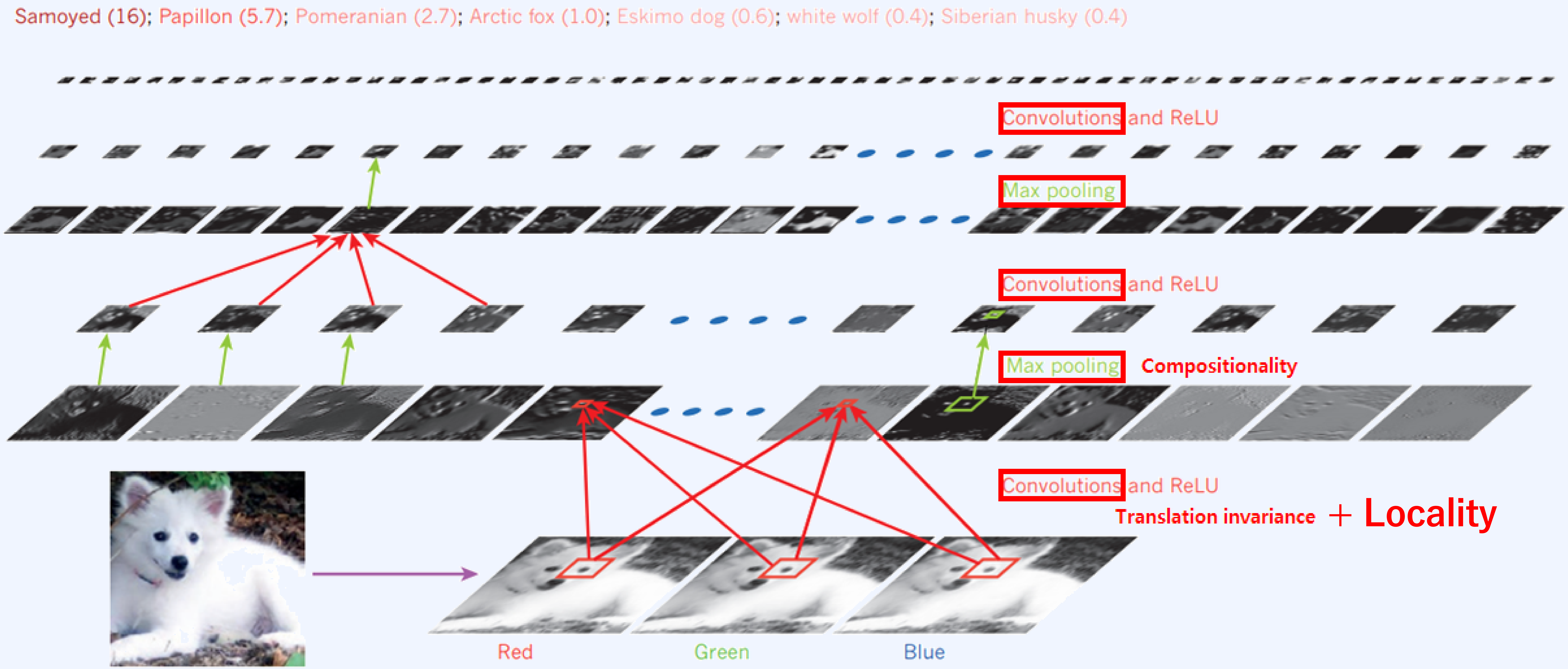

RNN /
Wikimedia Commons





# 과거의 Deep learning 주류 모델들 : inductive bias




[Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, 2015.]

# Context Understanding
## Overview

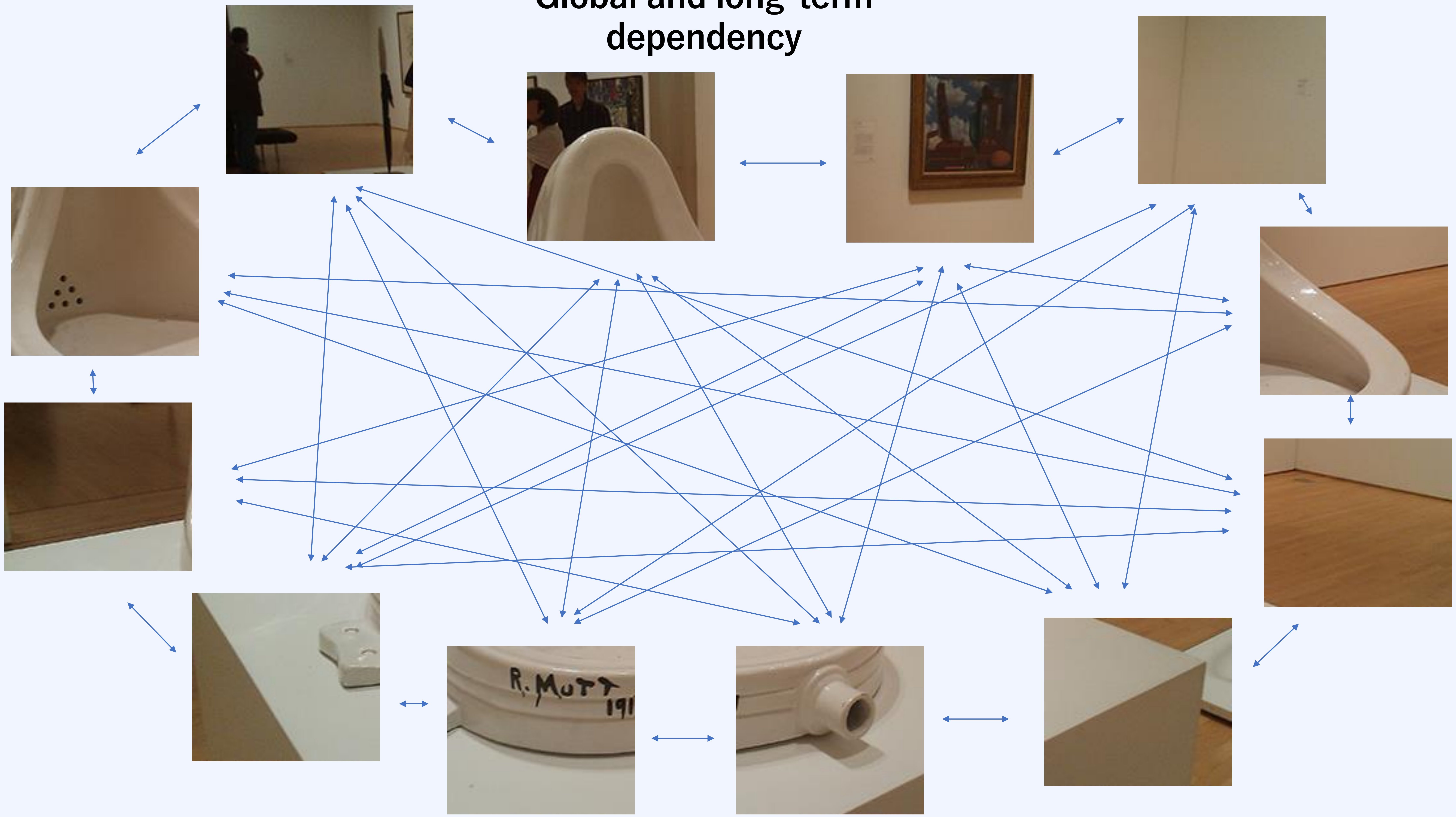


**Global and long-term dependency**

# Context Understanding
Overview



**Global and long-term dependency**



# Context Understanding
## Overview

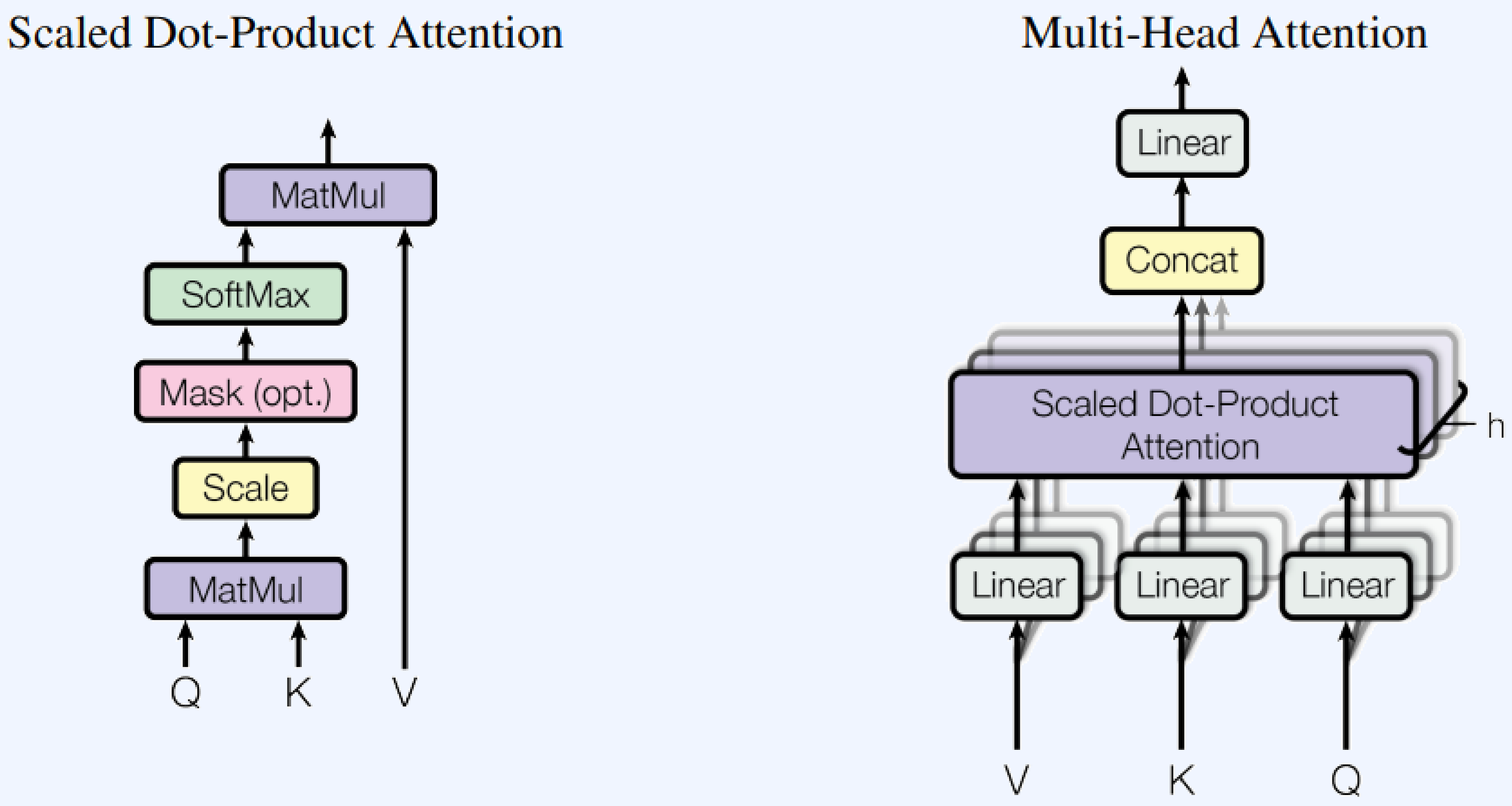


Attention Is All You Need,
Vaswani et al., NIPS 2017

**Transformer**

## Context Understanding 챕터에서는…

**Attention block**



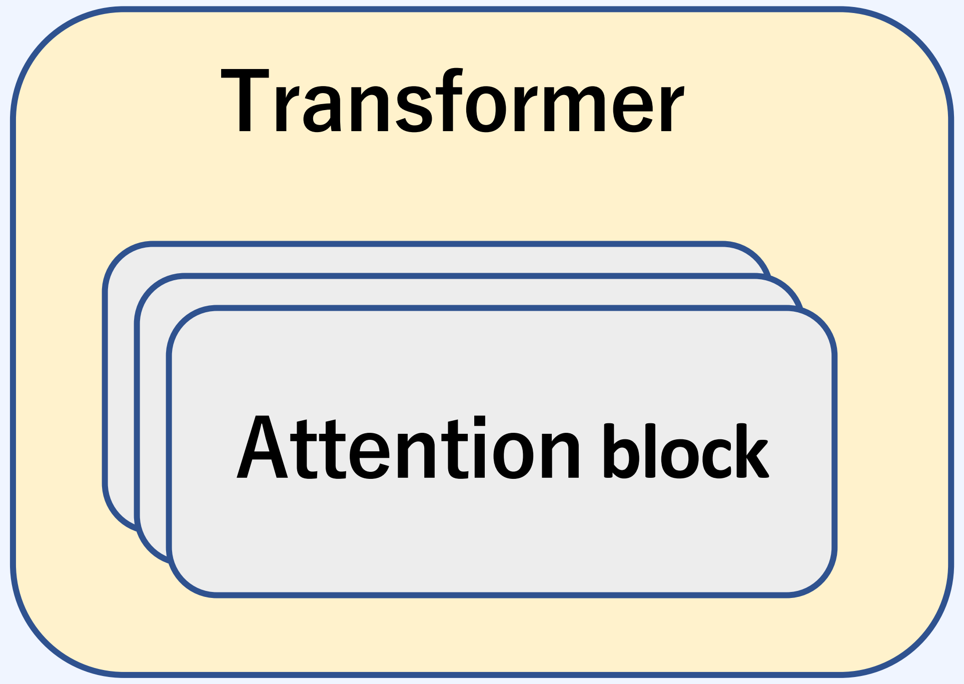

Attention Is All You Need, Vaswani et al., NIPS 2017

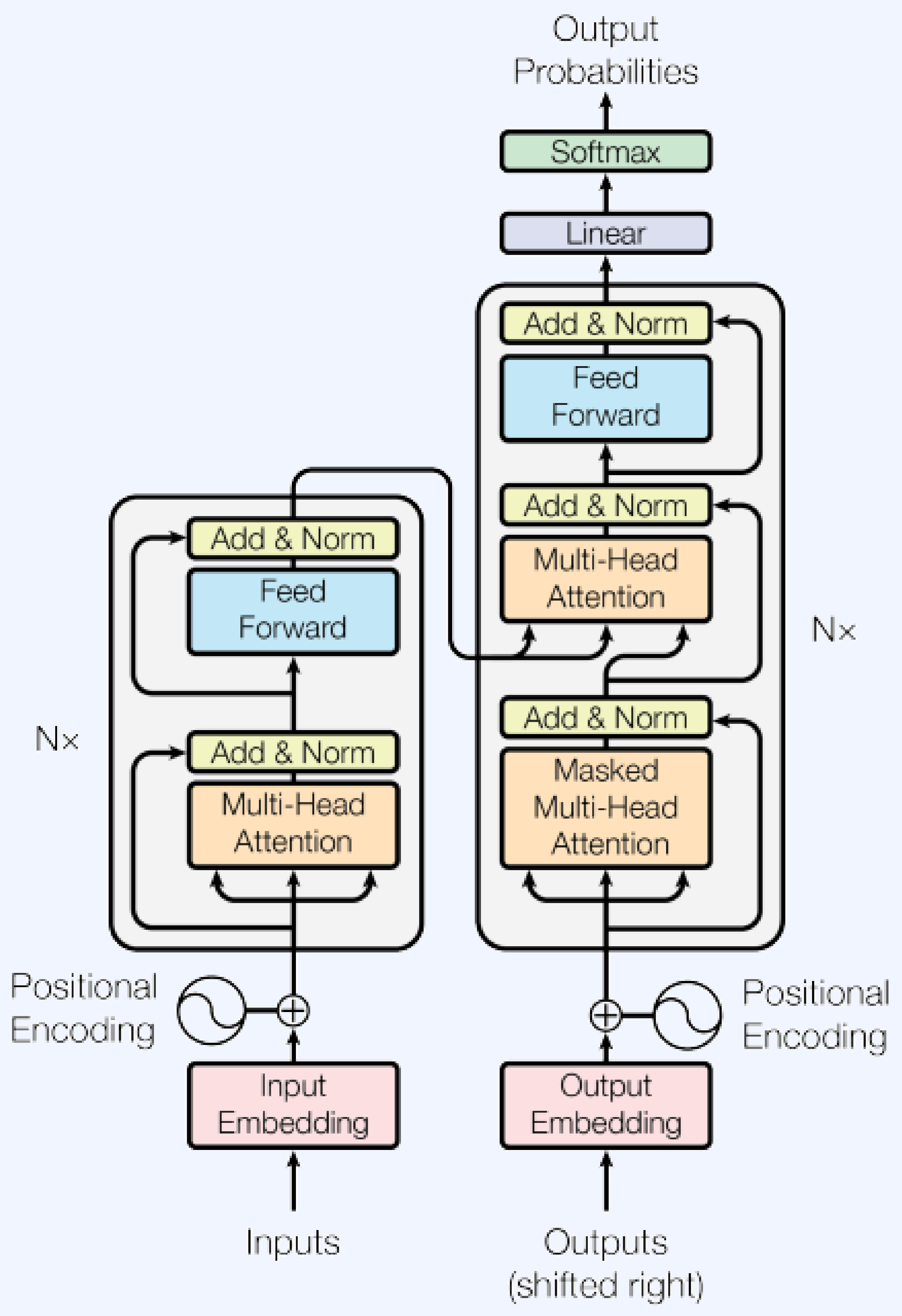
 





## Context Understanding 챕터에서는…




Attention Is All You Need, Vaswani et al., NIPS 2017

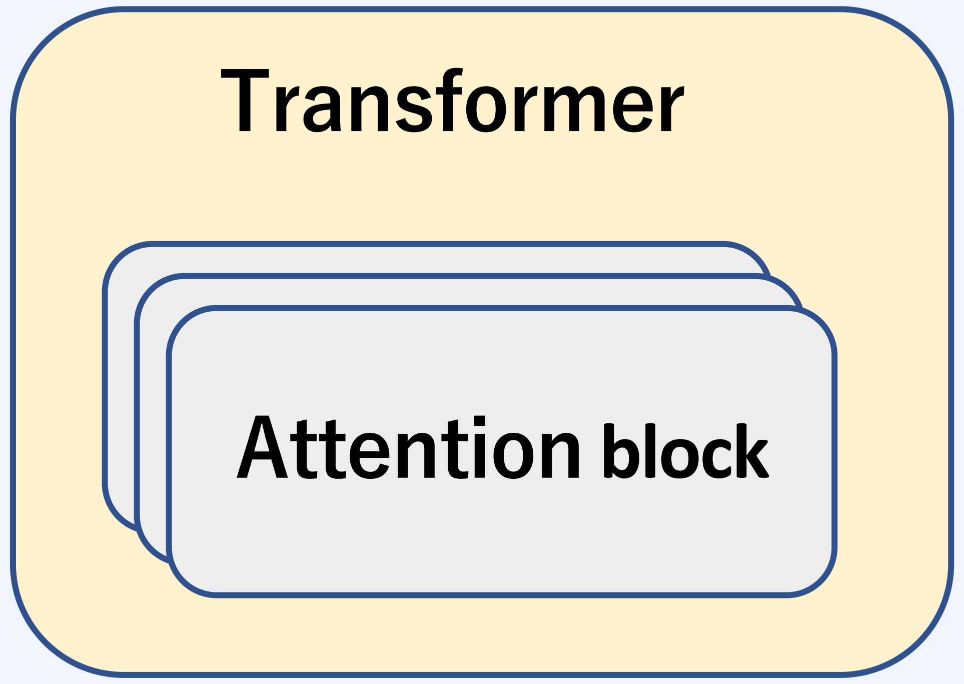
 

Context Understanding

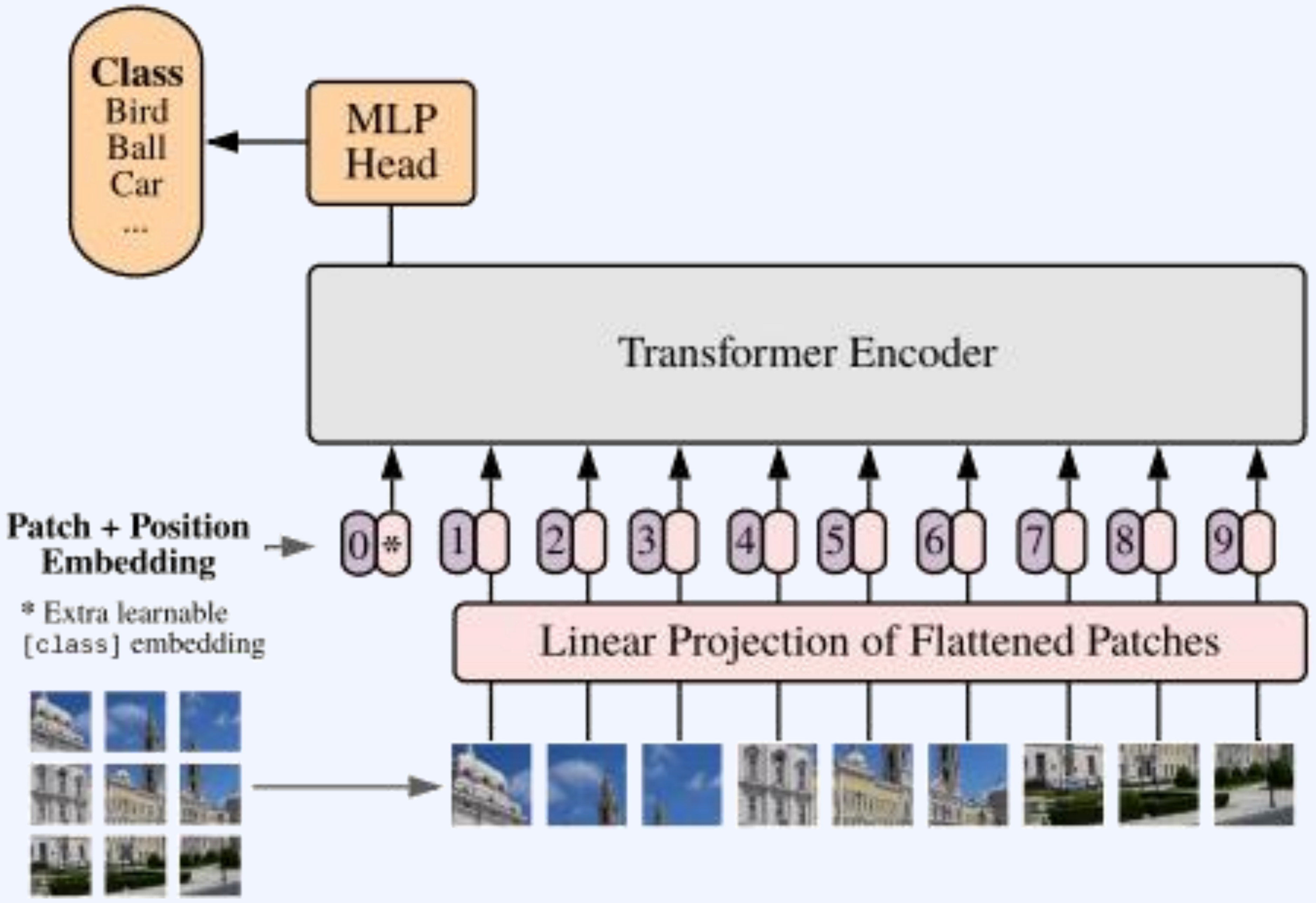
Overview



## Context Understanding 챕터에서는…



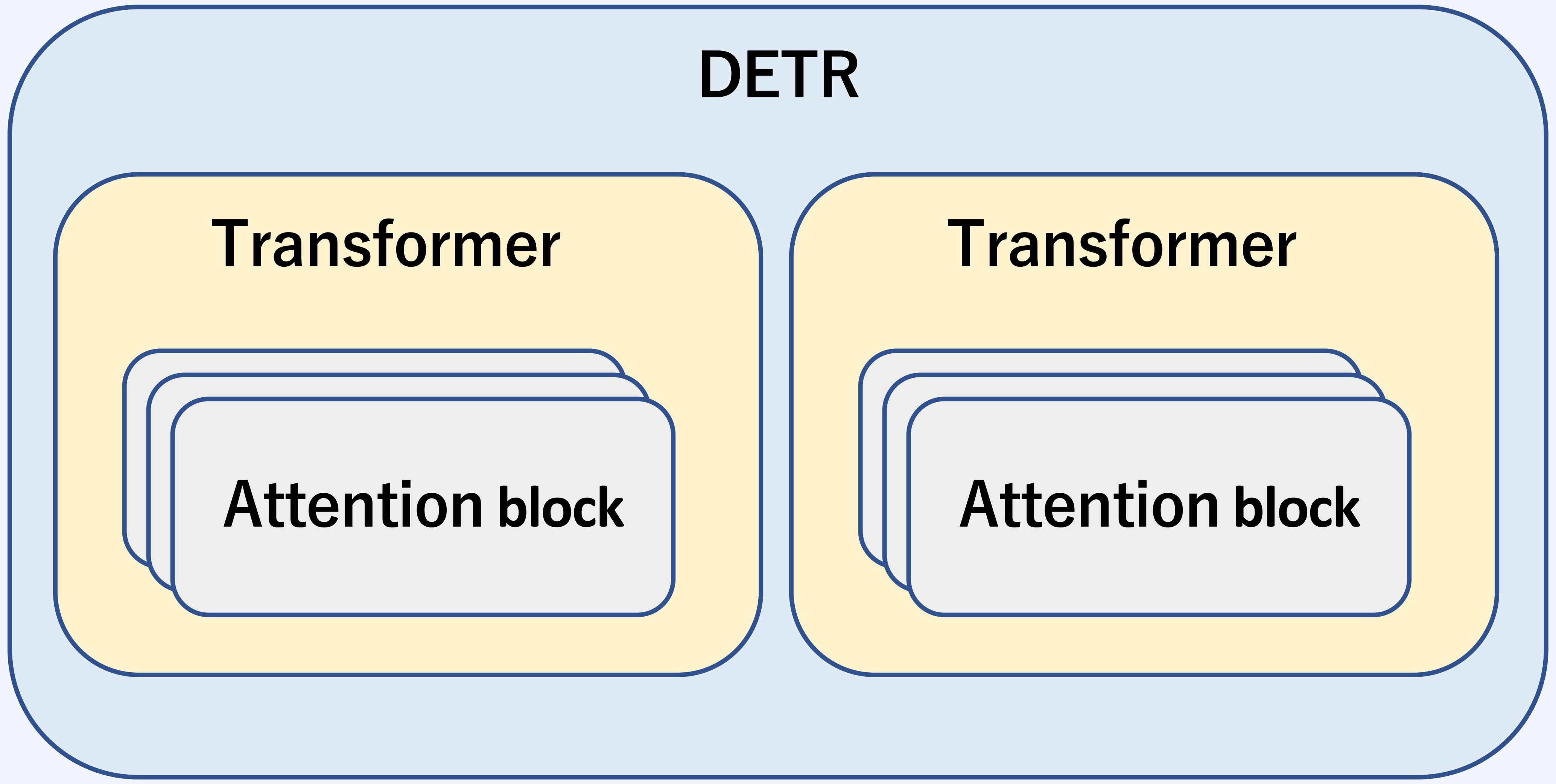

Visual Transformers

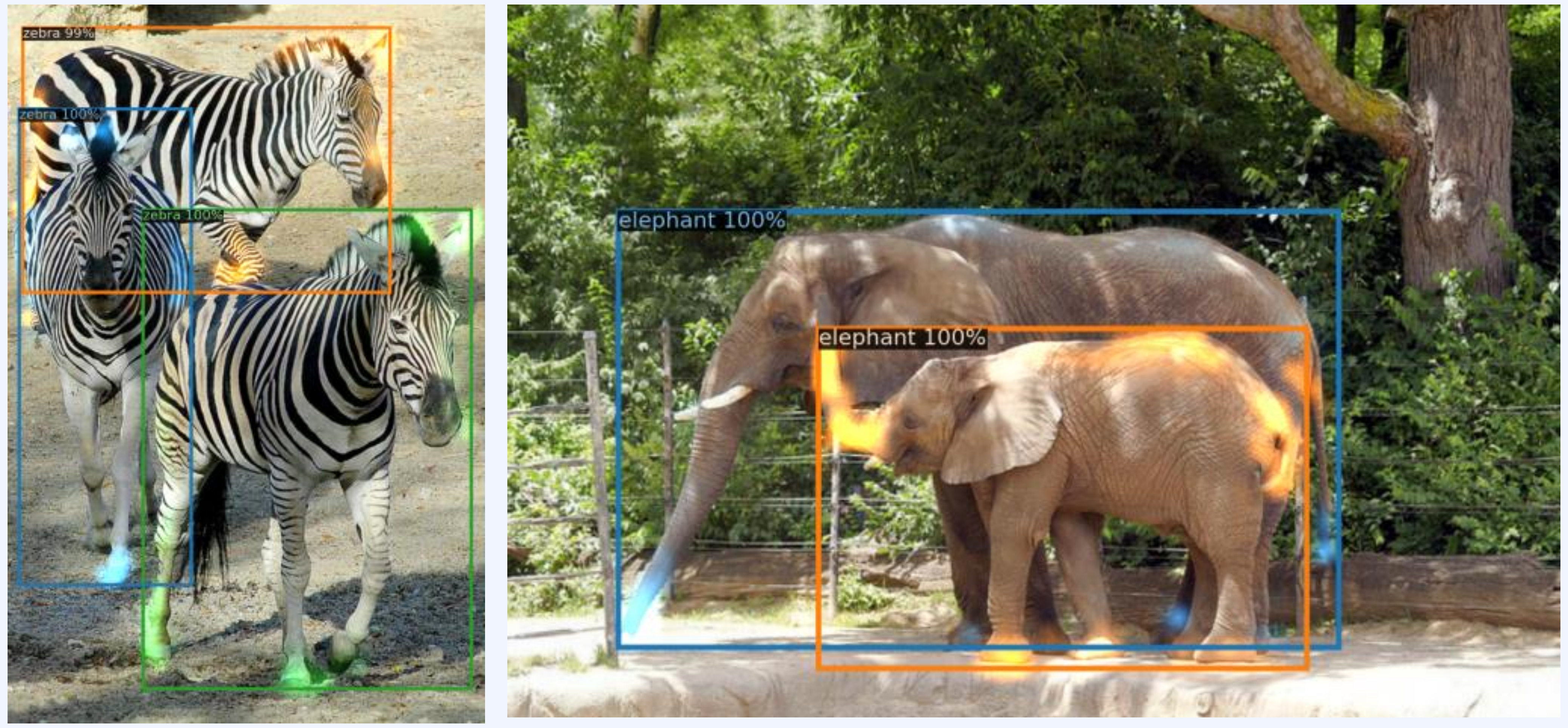


Context Understanding
Overview



## Context Understanding 챕터에서는…

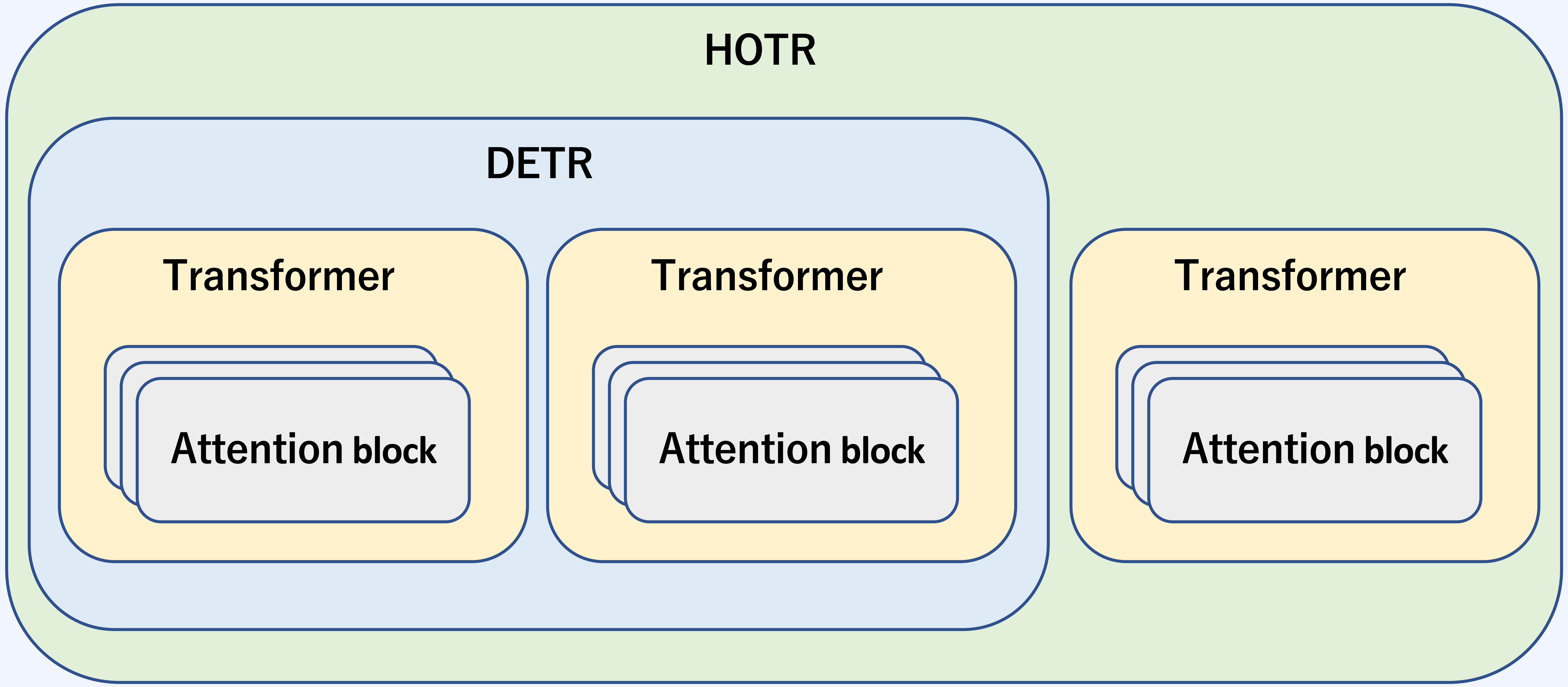


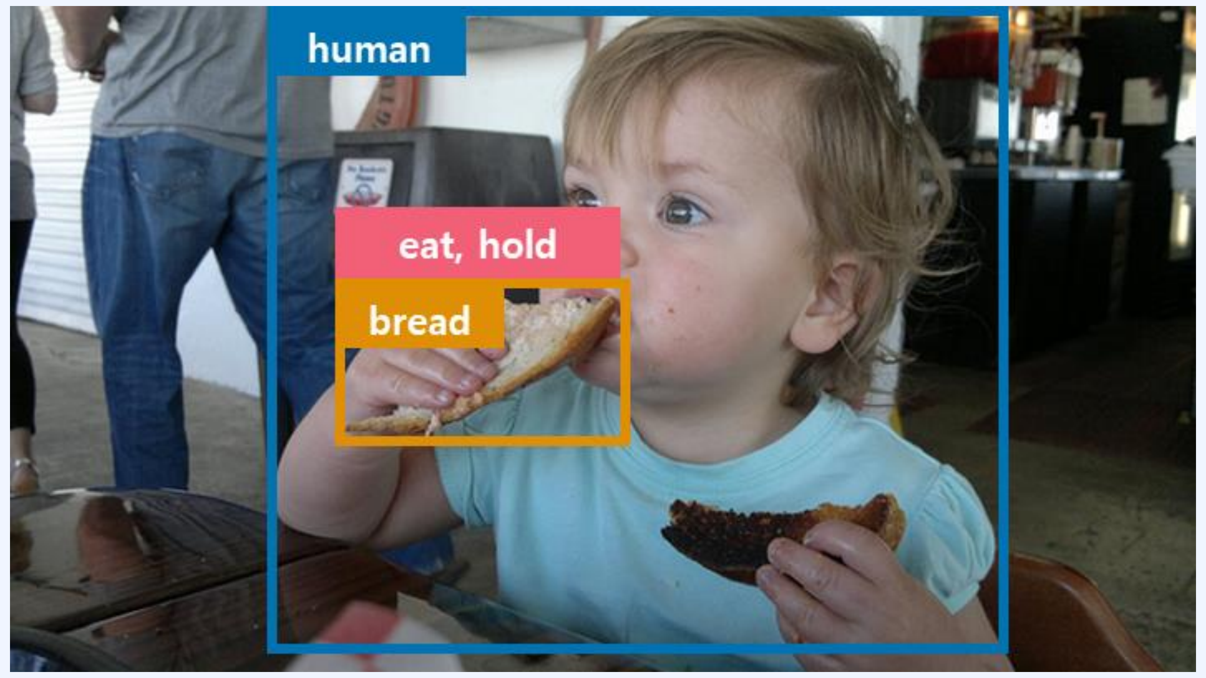




End-to-End Object Detection with Transformers, Carion et al., ECCV 2020 (oral)

# Context Understanding
## Overview

**Context Understanding 챕터에서는…**




HOTR: End-to-End Human-Object Interaction Detection with Transformers, Kim et al., CVPR 2021 (oral)