Introduction to Zero-shot learning

DMQA Open Seminar

2021.08.06

유이경

Contents

1. **Introduction**
   - Background

2. **Zero-shot learning**
   - Definition
   - Side-information

3. **Learning method**
   - Zero-shot learning using attributes
   - Base model of embedding-based approach

4. **Conclusions**

발표자 소개

❖ 유이경

- 고려대학교 산업경영공학
- Data Mining & Quality Analytics Lab
- M.S. Student (2021.03 ~)

❖ Research Interest

- Machine Learning / Deep Learning
- Multi-task learning
- Meta-learning

❖ Contact

- E-mail: ylk0801@korea.ac.kr

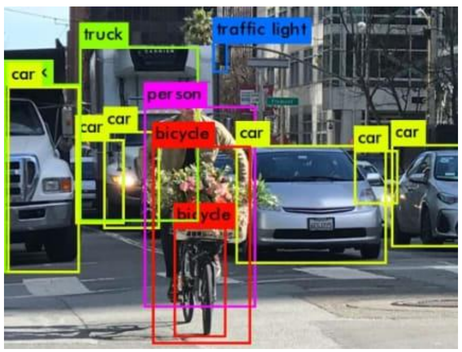
1. Introduction

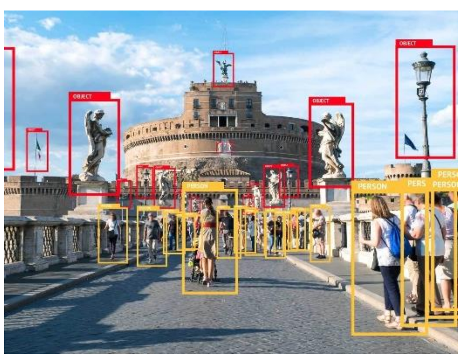
Introduction

Background

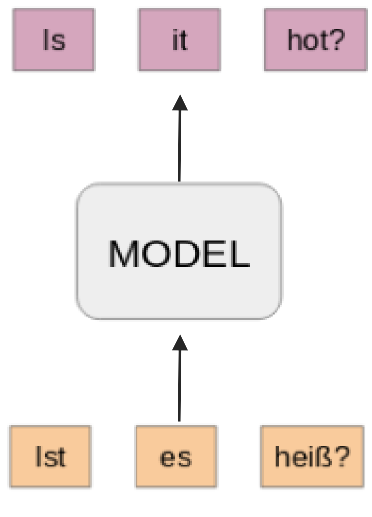
❖ 다양한 분야에서 우수한 성능을 보이는 딥러닝

Computer Vision

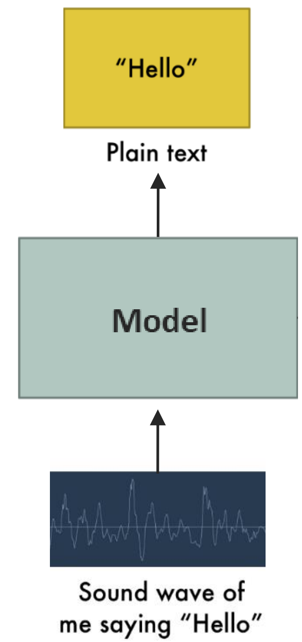




Natural Language Processing



Speech Recognition



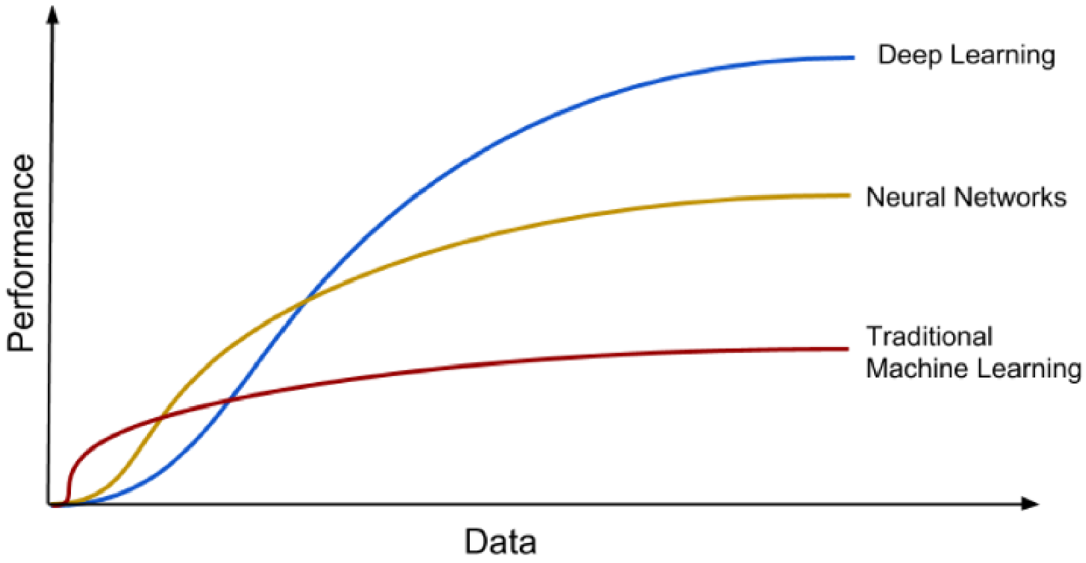
Data Mining Quality Analytics

https://www.analyticsvidhya.com/blog/2019/01/neural-machine-translation-keras/
https://medium.com/@ageitgey/machine-learning-is-fun-part-6-how-to-do-speech-recognition-with-deep-learning-28293c162f7a

Introduction

Background

❖ 이러한 우수한 성능에는 막대한 양의 데이터가 기반이 됨



Introduction

Background

❖ **그러나, 현실에서는…**

✓ 막대한 양의 데이터 중 **정답 레이블이 함께 존재하지 않는 데이터가 훨씬 많음**

✓ 레이블을 지정하는데 드는 **시간과 비용에 따른 제약**

✓ 경우에 따라 레이블을 지정하는 것은 **전문가만이 수행가능**

Labels

Rose

Tulip

?

Introduction

Background

❖ 레이블이 존재하는 데이터가 없는 상황 속에서, 해당 카테고리의 데이터를 올바르게 예측하는 것은 현실에서 매우 중요 → "Zero-shot learning"을 통해!

Hyacinth

No labeled data

Unseen data

Model

Predicted!

Hyacinth

"Zero-shot learning"

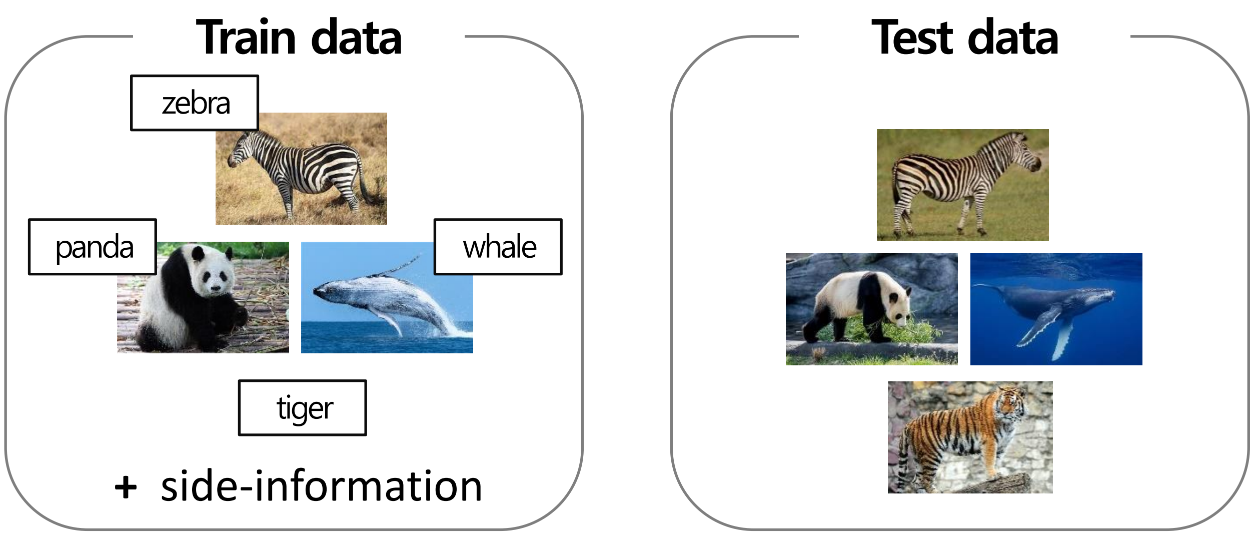
2. Zero-shot learning

Zero-shot learning

Definition

❖ **Zero-shot learning이란?**

- 레이블이 지정된 소수의 클래스 집합 데이터와 클래스에 대한 추가 정보만을 사용하여, 한 번도 본 적 없는 많은 클래스까지 잘 예측하도록 학습
- 학습 시, 레이블이 지정된 데이터와 추가 정보만을 사용해 학습
- 테스트 시, 학습 때 보았던 클래스의 데이터와 한 번도 본 적 없는 클래스의 데이터에 대해 레이블 예측 수행

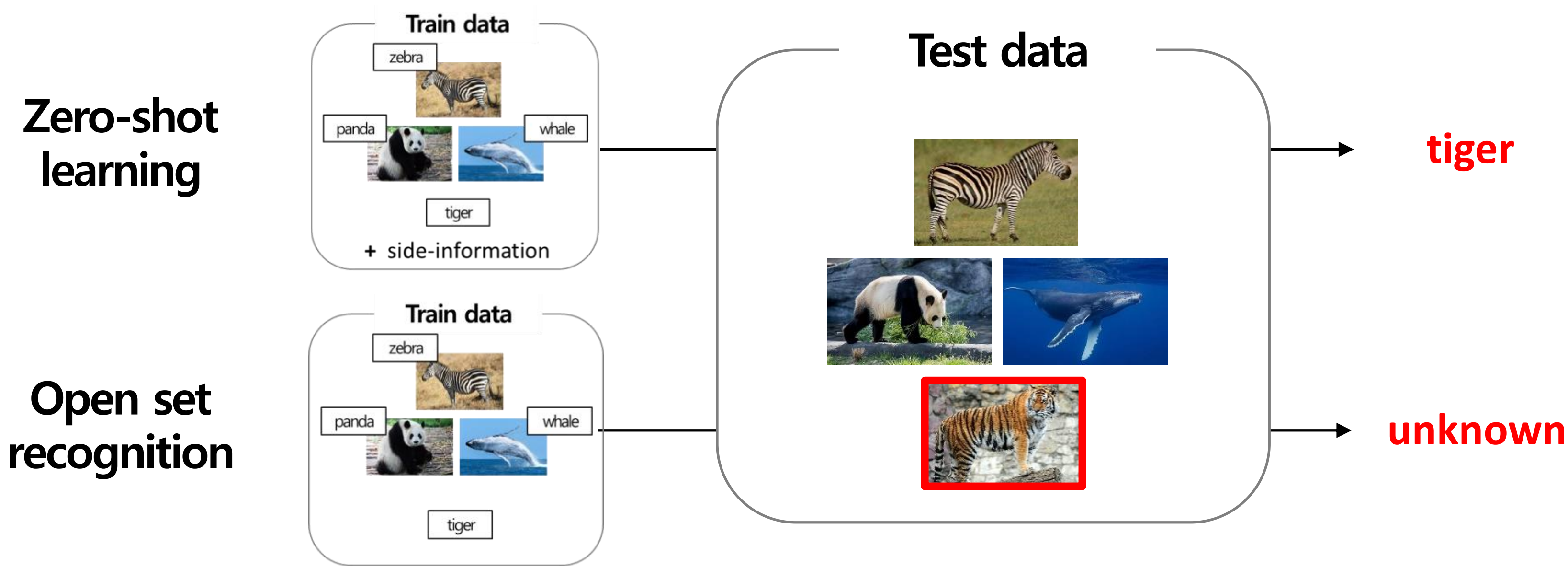


Zero-shot learning

Definition

❖ **Zero-shot learning vs. Open set recognition**

- Zero-shot learning의 목적은 알고 있는 class의 seen data와 unseen data 모두 올바른 class로 분류하는 것 → **기존 classification 성능 강화**
- Open set recognition의 목적은 알고 있는 class의 seen data는 올바른 class로 분류하고, unseen data는 특정 class가 아닌 unknown data 자체로 분류하는 것 → **unknown data detection**

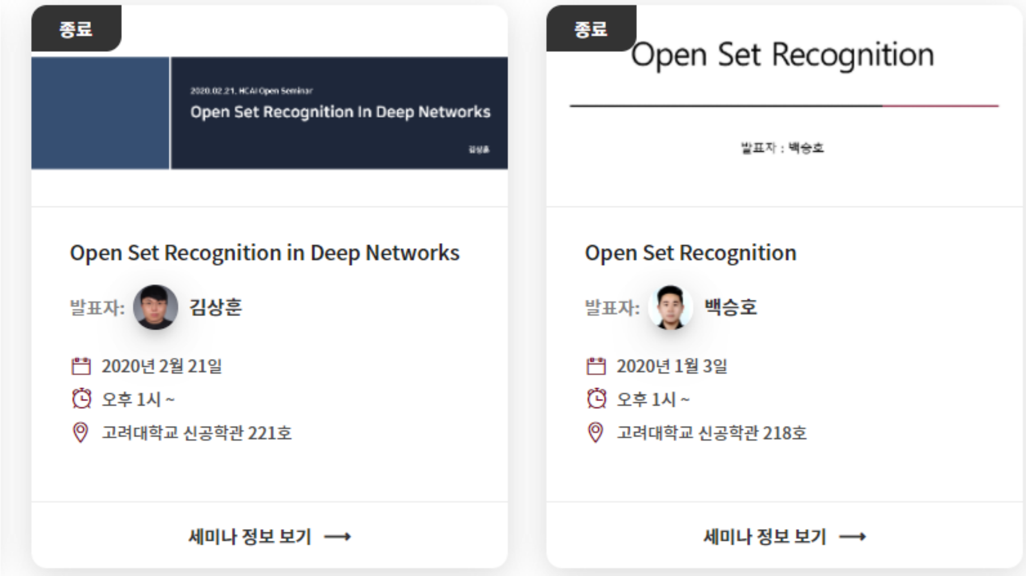


Zero-shot learning

Definition

❖ Zero-shot learning vs. Open set recognition

- Zero-shot learning의 목적은 알고 있는 class의 seen data와 unseen data 모두 올바른 class로 분류하는 것 → **기존 classification 성능 강화**
- Open set recognition의 목적은 알고 있는 class의 seen data는 올바른 class로 분류하고, unseen data는 특정 class가 아닌 unknown data 자체로 분류하는 것 → **unknown data detection**



Zero-shot learning

Definition

❖ **Zero-shot learning의 구성요소**

- Image: x
- Class label: y
- **Side-information: h_y**

x $\qquad\qquad$ y

$\longleftrightarrow$ h_y $\longleftrightarrow$ zebra

다양한 종류의 Side-information

✓ Attributes
✓ Hierarchy similarity measures
✓ Human gaze

⋮

Zero-shot learning

Side-information

❖ **Attributes as side-information**

x $\qquad$ y

$\longleftrightarrow$

black–stripe
has tail
lives on land
small

$\longleftrightarrow$

zebra

- ✓ **Attributes**
- ✓ Hierarchy similarity measures
- ✓ Human gaze

⋮

Zero-shot learning

Side-information

❖ **Objects descriptions as side-information**

x

$\longleftrightarrow$

WIKIPEDIA
The Free Encyclopedia

Article Talk Read View source View history

Zebra

From Wikipedia, the free encyclopedia

For other uses, see Zebra (disambiguation).

Zebras (UK: /ˈzɛbrə/, US: /ˈziː-/) (subgenus *Hippotigris*) are African equines with distinctive black-and-white striped coats. There are three extant species: the Grévy's zebra (*Equus grevyi*), plains zebra (*E. quagga*), and the mountain zebra (*E. zebra*). Zebras share the genus *Equus* with horses and asses, the three groups being the only living members of the family Equidae. Zebra stripes come in different patterns, unique to each individual. Several theories have been proposed for the function of these stripes, with most evidence supporting them as a form of protection from biting flies. Zebras inhabit eastern and southern Africa and can be found in a variety of habitats such as savannahs, grasslands, woodlands,

$\longleftrightarrow$

y

zebra

- ✓ **Attributes - Object descriptions from Wikipedia**
- ✓ Hierarchy similarity measures
- ✓ Human gaze

⋮

Zero-shot learning

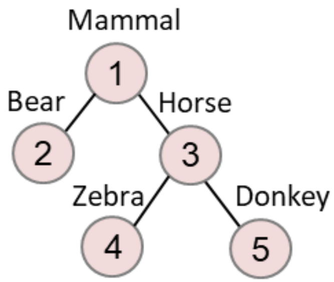
Side-information

❖ **Hierarchy similarity measures as side-information**

Hierarchy

x

$\longleftrightarrow$



Class 4 = [2 3 1 0 1]

$\longleftrightarrow$

y

zebra

- ✓ Attributes
- ✓ **Hierarchy similarity measures**
- ✓ Human gaze

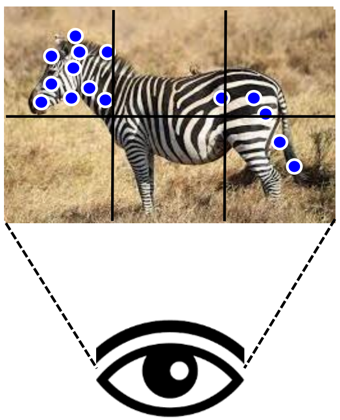
⋮

Zero-shot learning

Side-information

❖ **Human gaze as side-information**

x



Gaze points histogram

10	1	2
0	0	2

y

zebra

- ✓ Attributes
- ✓ Hierarchy similarity measures
- ✓ **Human gaze**

⋮

3. Learning method

Learning method

Zero-shot learning using attributes

❖ Attributes를 side-information으로 사용하여 학습하는 대표적 approach

Embedding-based approach

Generative model-based approach

Learning method

Zero-shot learning using attributes

❖ Attributes를 side-information으로 사용하여 학습하는 두 가지 대표적 approach

Embedding-based approach

Generative model-based approach

Learning method

Zero-shot learning using attributes

❖ **Embedding-based approach**

- Attributes를 사용하여 각 클래스에 해당하는 정보를 vector representation으로 변환
- 이미지에 해당하는 의미를 가진 semantic embedding 값

Image		**Attributes**		**Class**
	⟷	black–stripe has tail lives on land small	⟷	zebra [1 0 1 1 0 1]
	⟷	gray has tail lives in water big	⟷	[0 1 1 0 1 0] whale

Learning method

Zero-shot learning using attributes

❖ 두 이미지 간 동일한 attributes → 같은 벡터 값

Image	Attributes	Class
	black–stripe **has tail** lives on land small	zebra [1 0 **1** 1 0 1]
	gray **has tail** lives in water big	[0 1 **1** 0 1 0] whale

Learning method

Zero-shot learning using attributes

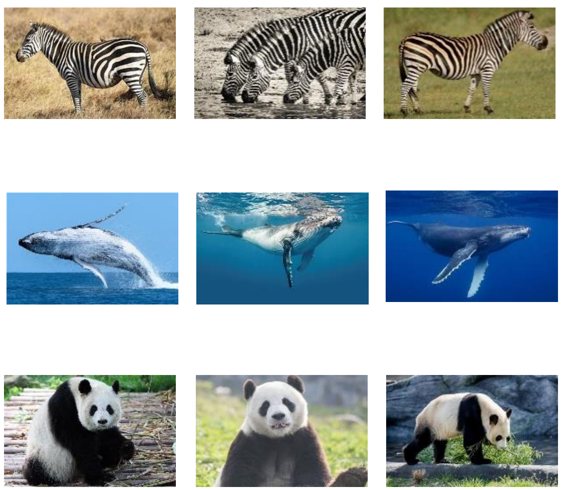
❖ 두 이미지 간 동일하지 않은 attributes → 다른 벡터 값

Image	Attributes	Class
	black–stripe has tail lives on land **small**	zebra [1 0 1 1 0 **1**]
	gray has tail lives in water **big**	[0 1 1 0 1 **0**] whale

Learning method

Zero-shot learning using attributes

❖ 학습 시 관측된 데이터가 없는 클래스 값이더라도, embedding 값을 학습해 라벨 예측 가능

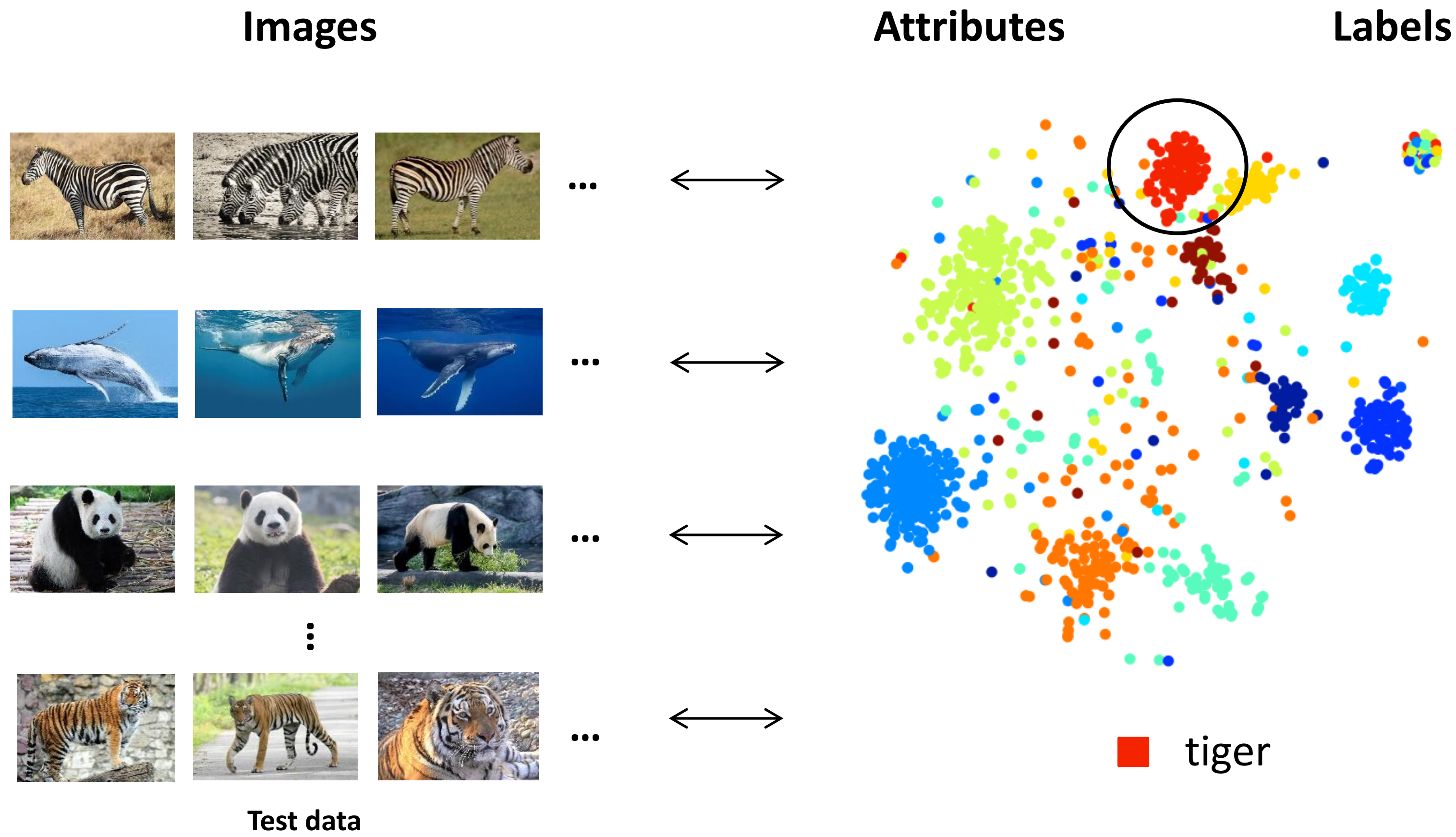
Images		Attributes		Labels
 …	⟷	black–stripe has tail lives on land small	⟷	zebra [1 0 1 1 0 1]
…	⟷	gray has tail lives in water big	⟷	whale [0 1 1 0 1 0]
…	⟷	black–stripe no tail lives on land medium	⟷	panda [1 0 1 1 0 1]
⋮				⋮
No images	⟷	black–stripe has tail lives on land medium	⟷	tiger [1 1 1 0 1 0]

Train data

Learning method

Zero-shot learning using attributes

❖ 학습 시 관측된 데이터가 없는 클래스 값이더라도, embedding 값을 학습해 라벨 예측 가능



Learning method

Base model of embedding-based approach

❖ DeViSE: A Deep Visual-Semantic Embedding Model

- NIPS 2013 / Google이 발표한 논문
- 2021년 8월 5일 기준 2045회 인용
- Embedding-based approach의 근간이 되는 기본 모델 제안

DeViSE: A Deep Visual-Semantic Embedding Model

Andrea Frome*, Greg S. Corrado*, Jonathon Shlens*, Samy Bengio
Jeffrey Dean, Marc'Aurelio Ranzato, Tomas Mikolov
* These authors contributed equally.

{afrome, gcorrado, shlens, bengio, jeff, ranzato[†], tmikolov}@google.com
Google, Inc.
Mountain View, CA, USA

Abstract

Modern visual recognition systems are often limited in their ability to scale to large numbers of object categories. This limitation is in part due to the increasing difficulty of acquiring sufficient training data in the form of labeled images as the number of object categories grows. One remedy is to leverage data from other sources – such as text data – both to train visual models and to constrain their predictions. In this paper we present a new *deep visual-semantic embedding model* trained to identify visual objects using both labeled image data as well as semantic information gleaned from unannotated text. We demonstrate that this model matches state-of-the-art performance on the 1000-class ImageNet object recognition challenge while making more semantically reasonable errors, and also show that the semantic information can be exploited to make predictions about tens of thousands of image labels not observed during training. Semantic knowledge improves such *zero-shot* predictions achieving hit rates of up to 18% across thousands of novel labels never seen by the visual model.

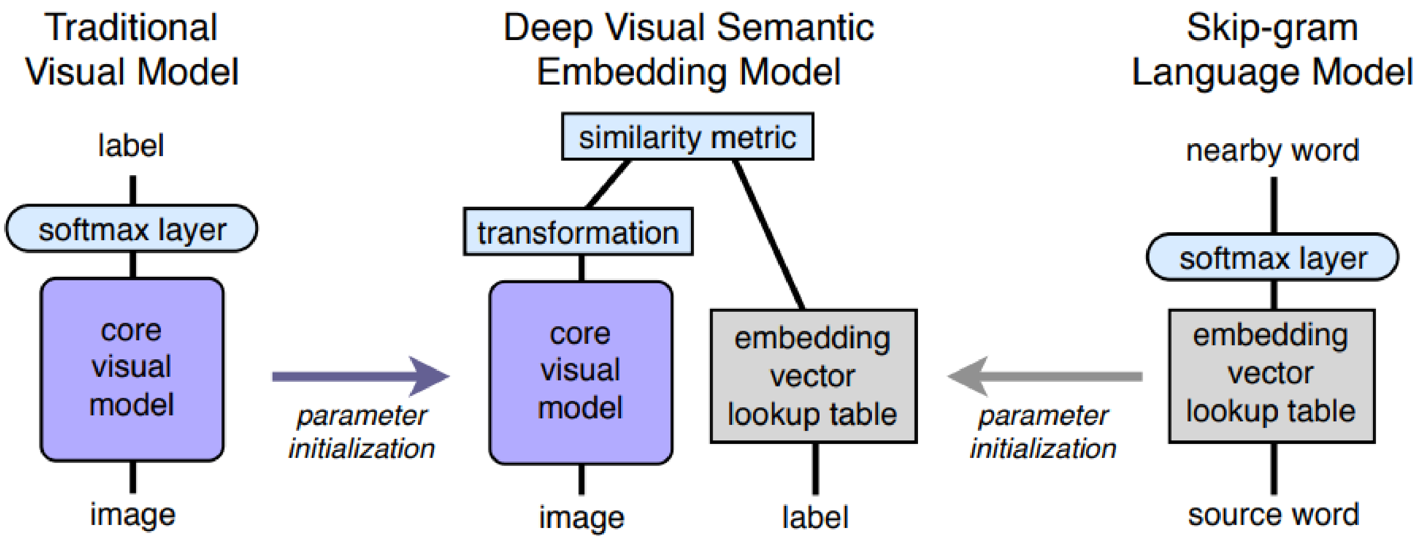
Frome, A., Corrado, G., Shlens, J., Bengio, S., Dean, J., Ranzato, M. A., & Mikolov, T. (2013). Devise: A deep visual-semantic embedding model.

Learning method

Base model of embedding-based approach

❖ **Visual model과 Language model을 결합**

- 레이블이 지정된 이미지 데이터와 해당 이미지와 관련된 텍스트에서 수집한 의미 정보를 모두 사용하여 개체를 식별하도록 훈련

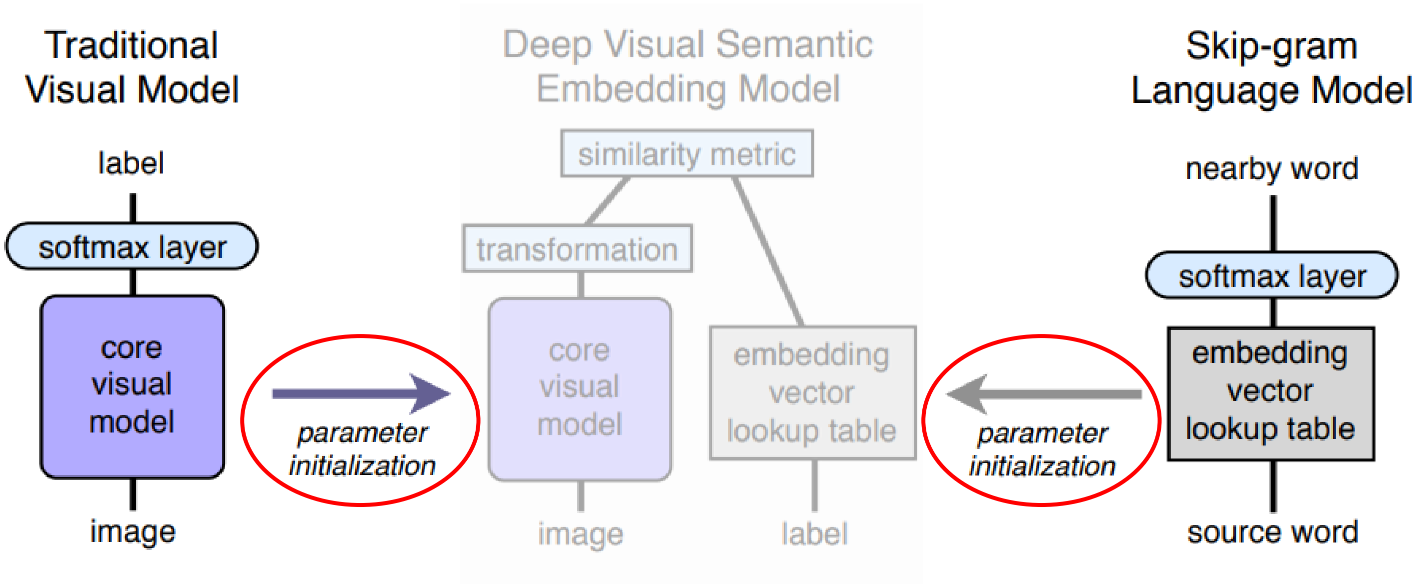


Learning method

Base model of embedding-based approach

❖ **Pre-training**

- Visual model은 **AlexNet**, Language model은 **Skip-gram LM**으로 각각 사전 학습
- 사전 학습된 파라미터로 모델 파라미터 초기화

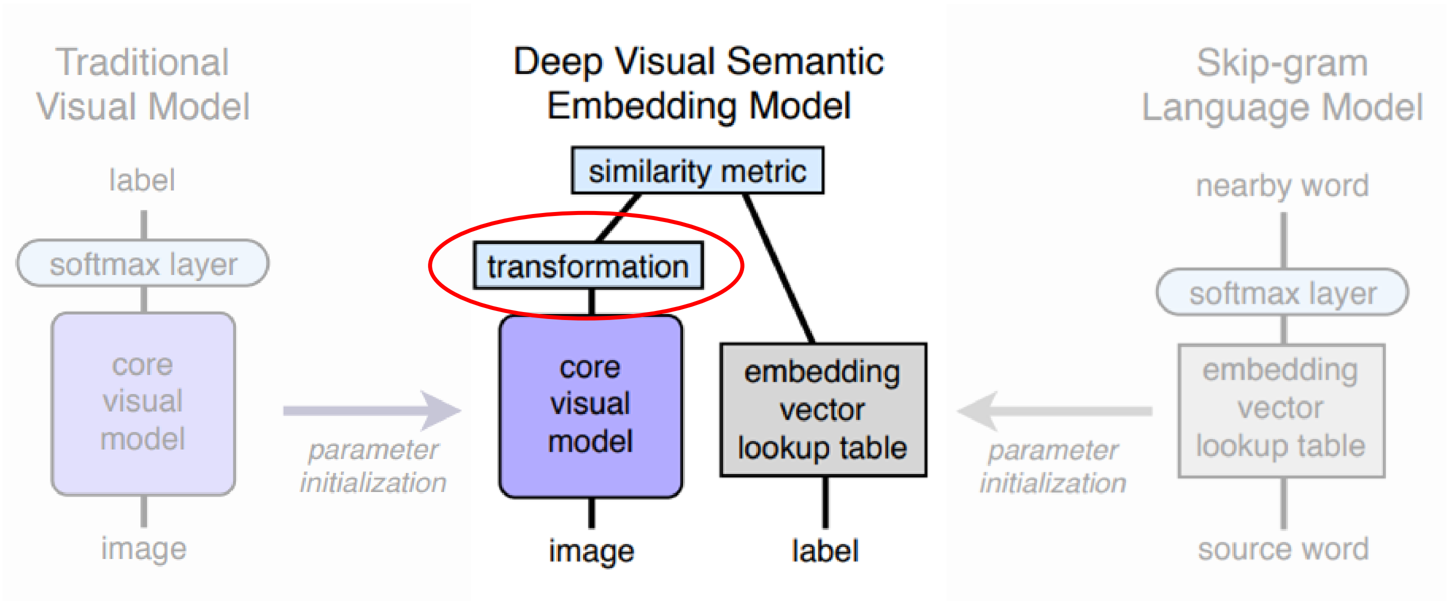


Learning method

Base model of embedding-based approach

❖ Transformation

- Visual model의 상단에 있는 n차원 표현을 Language model 고유의 m차원 표현으로 매핑하는 선형 변환 (본 논문에서는 4,096 → 500/1000차원으로 변환)

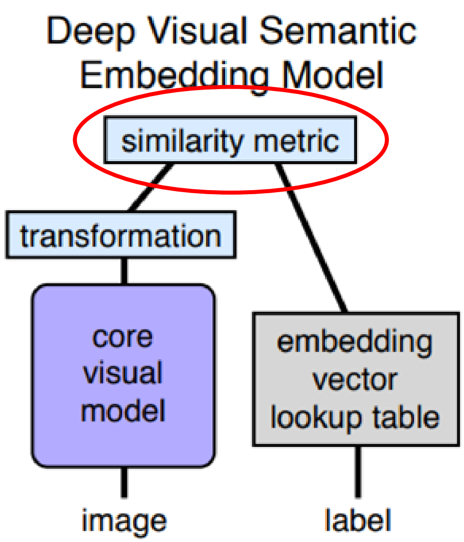


Learning method

Base model of embedding-based approach

❖ Similarity metric

- (1)**이미지와 정답 레이블의 벡터**로 계산된 코사인 유사도가 (2)**이미지와 무작위로 선택된 다른 레이블의 벡터**로 계산된 코사인 유사도보다 크도록 학습이 이루어짐



(1)  ↔ zebra [1 0 1 1 0 1]

(2)  ↔ whale [0 1 1 0 1 0]

Similarity: (1) > (2)

Learning method

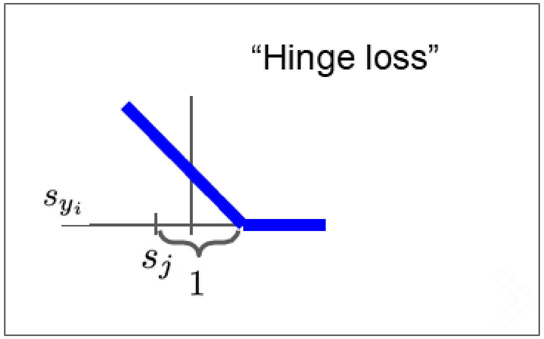
Base model of embedding-based approach

❖ Similarity metric

- 코사인 유사도와 Hinge rank loss의 조합으로 구성된 손실함수

$$loss(image, label) = \boxed{\sum_{j \neq label} \max[0, margin - \vec{t}_{label} M \vec{v}(image) + \vec{t}_j M \vec{v}(image)]}$$

cf) Hinge loss



$$L_i = \sum_{j \neq y_i} \begin{cases} 0 & \text{if } s_{y_i} \geq s_j + 1 \\ s_j - s_{y_i} + 1 & \text{otherwise} \end{cases}$$
$$= \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

Given an example (x_i, y_i)
where x_i is the image and
where y_i is the (integer) label,

and using the shorthand for the scores vector: $s = f(x_i, W)$

the SVM loss has the form:

$$L_i = \boxed{\sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)}$$

Learning method

Base model of embedding-based approach

❖ **Experiments**

	SUN		CUB		AWA		aPY	
Method	**SS**	**PS**	**SS**	**PS**	**SS**	**PS**	**SS**	**PS**
DAP [22]	38.9	39.9	37.5	40.0	57.1	44.1	35.2	33.8
CONSE [26]	44.2	38.8	36.7	34.3	63.6	45.6	25.9	26.9
CMT [34]	41.9	39.9	37.3	34.6	58.9	39.5	26.9	28.0
SSE [42]	54.5	51.5	43.7	43.9	68.8	60.1	31.1	34.0
LATEM [39]	56.9	55.3	49.4	49.3	74.8	55.1	34.5	35.2
ALE [3]	59.1	**58.1**	53.2	54.9	**78.6**	59.9	30.9	39.7
DEVISE [11]	57.5	56.5	53.2	52.0	72.9	54.2	35.4	**39.8**
SJE [4]	57.1	53.7	**55.3**	53.9	76.7	**65.6**	32.0	32.9
ESZSL [32]	57.3	54.5	55.1	53.9	74.7	58.2	34.4	38.3
SYNC [7]	**59.1**	56.3	54.1	**55.6**	72.2	54.0	**39.7**	23.9

Table 3: Zero-shot on SS = Standard Split, PS = Proposed Split using ResNet features (top-1 accuracy in %).

도메인 변경에
취약

Xian, Yongqin, Bernt Schiele, and Zeynep Akata. "Zero-shot learning-the good, the bad and the ugly." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.

4. Conclusions

Conclusions

1. **Why Zero-shot learning?**

2. **Side-information for Zero-shot learning**

3. **Embedding-based approach for Zero-shot learning**

Conclusions

1. Why Zero-shot learning?

2. Side-information for Zero-shot learning

3. Embedding-based approach for Zero-shot learning

Conclusions

1. Why Zero-shot learning?

?

- ✓ 정답 레이블이 함께 존재하지 않는 데이터가 훨씬 많음
- ✓ 전문가만이 수행가능한 레이블 지정 有
- ✓ 시간과 비용에 따른 제약

Conclusions

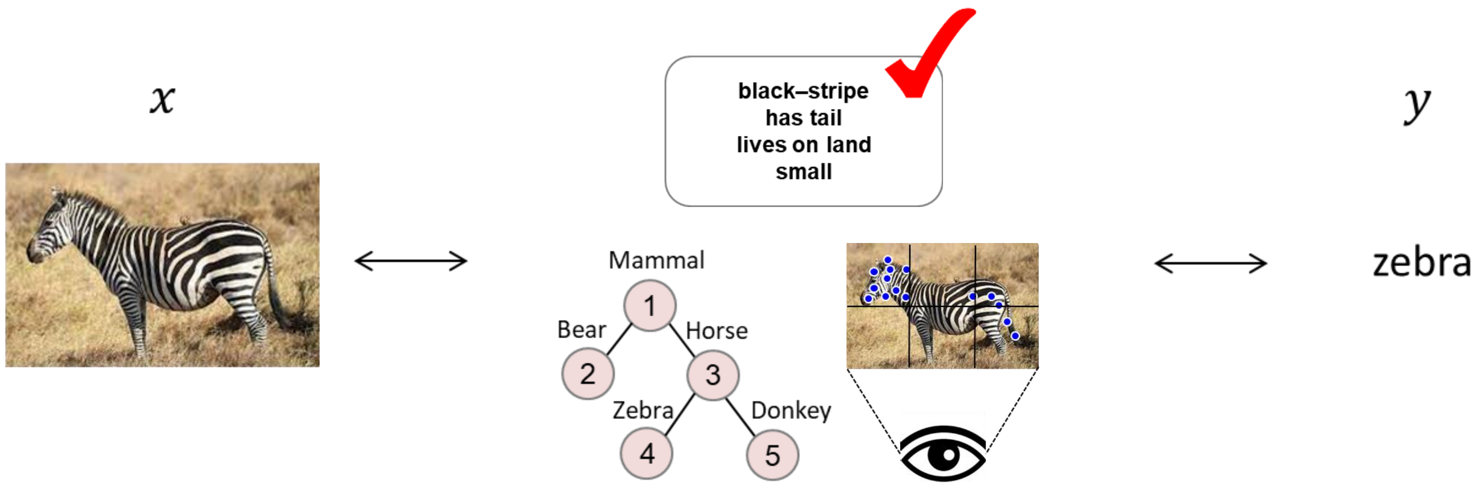
1. Why Zero-shot learning?

2. Side-information for Zero-shot learning

3. Embedding-based approach for Zero-shot learning

Conclusions

2. Side-information for Zero-shot learning



Conclusions

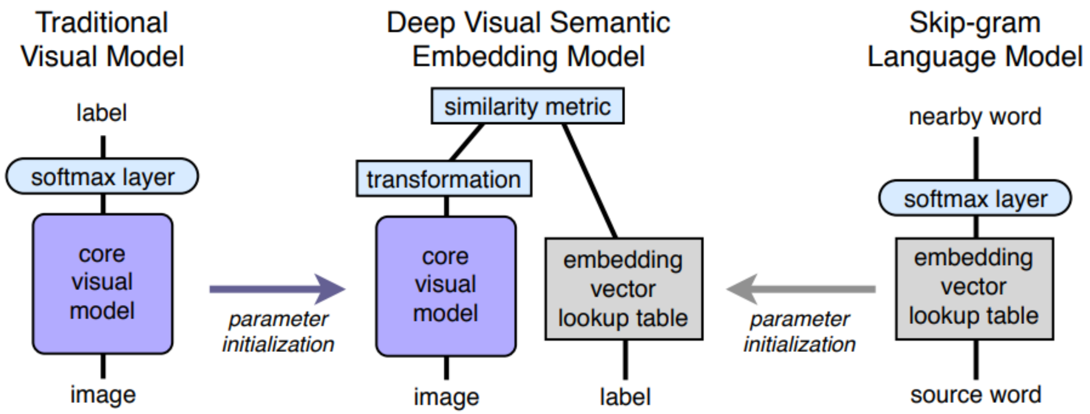
1. **Why Zero-shot learning?**

2. **Side-information for Zero-shot learning**

3. **Embedding-based approach for Zero-shot learning**

Conclusions

3. Embedding-based approach for Zero-shot learning



Thank you

References

1. Frome, A., Corrado, G., Shlens, J., Bengio, S., Dean, J., Ranzato, M. A., & Mikolov, T. (2013). Devise: A deep visual-semantic embedding model.
2. Akata, Z., Perronnin, F., Harchaoui, Z., & Schmid, C. (2013). Label-embedding for attribute-based classification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 819-826).
3. Wang, W., Zheng, V. W., Yu, H., & Miao, C. (2019). A survey of zero-shot learning: Settings, methods, and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 10(2), 1-37.
4. Xian, Y., Schiele, B., & Akata, Z. (2017). Zero-shot learning-the good, the bad and the ugly. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4582-4591).
5. https://www.youtube.com/watch?v=dE4nU5OaQqA
6. https://aihub.or.kr/node/24015
7. http://dmqm.korea.ac.kr/activity/seminar/301