

---

# Multimodal Deep Learning for Product Design Data

---

Yoon Sang Cho

DMQA Seminar

2019-10-04

# Contents

---

- Introduction
- Multimodal Deep Learning
- Application for Product Design Data
- Conclusions
- Future works

# Multimodal Deep Learning

---

?

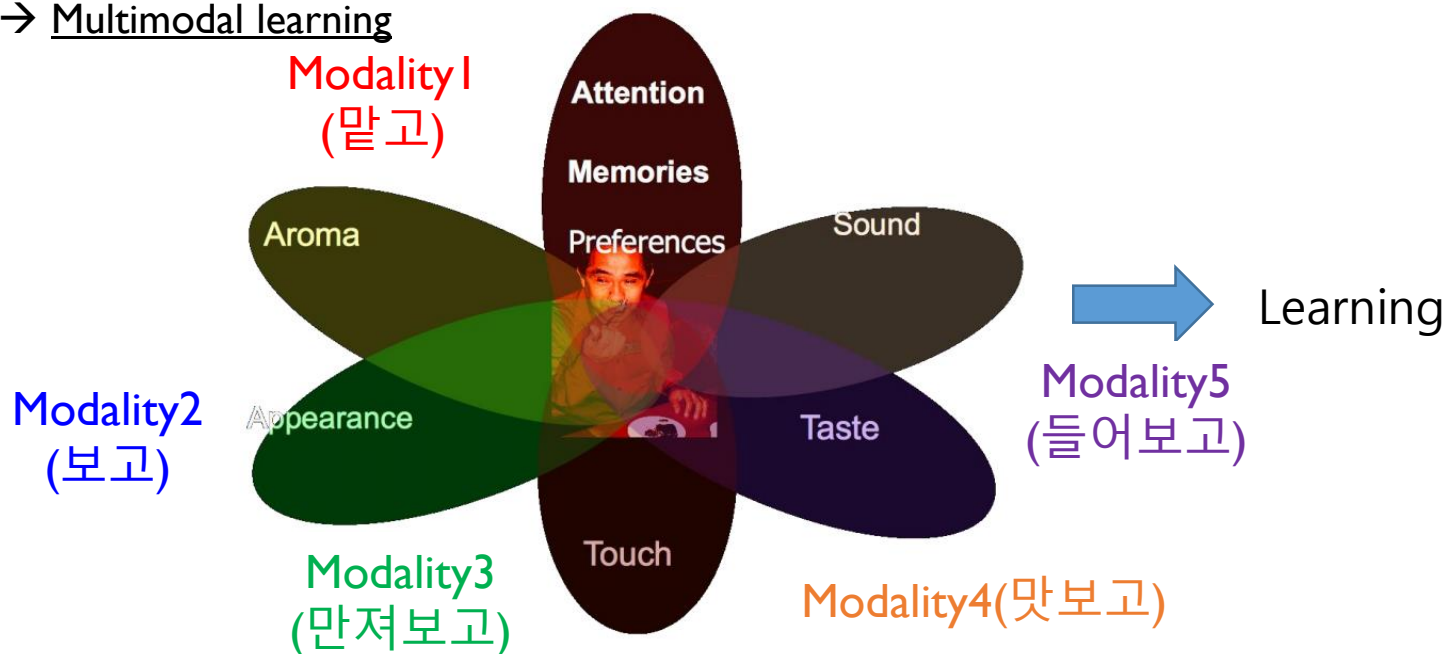
Multimodal 을 인공지능망 모델을 여러 층으로 쌓아 데이터 학습

# Introduction

## ❖ What is the 'Multimodal Learning' ?

- 인간은 살아가는 데 필요한 정보를 학습하기 위해 대표적으로 5개의 감각 기관으로부터 수집되는 데이터를 바탕으로 학습
- 인간의 인지적 학습법을 모방하여 다양한 형태(modality) 데이터로 학습하는 방법

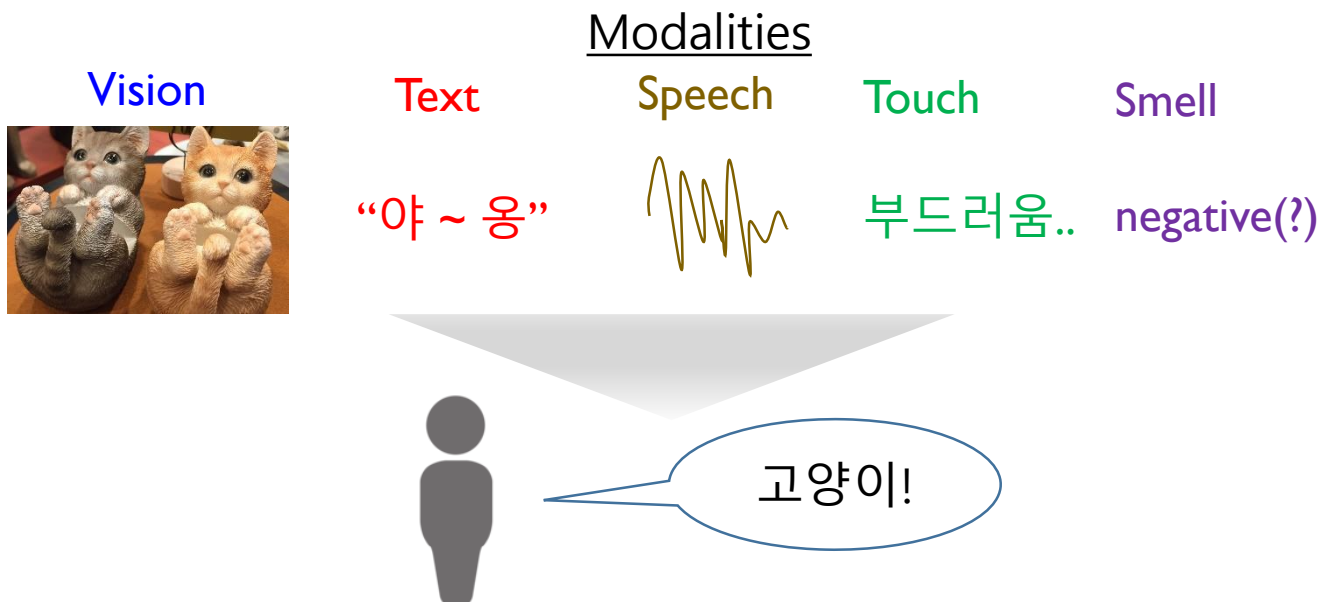
→ Multimodal learning



# Introduction

## ❖ What is the 'Multimodal Learning' ?

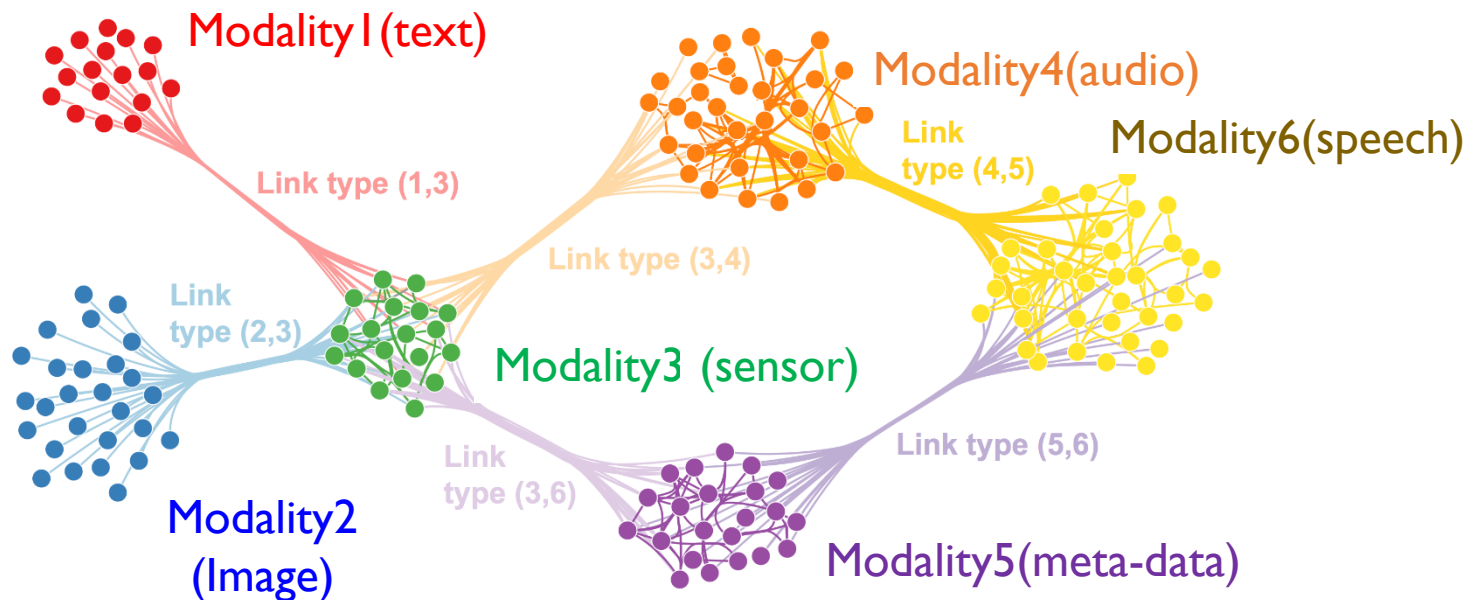
- 인간은 살아가는 데 필요한 정보를 학습하기 위해 대표적으로 5개의 감각 기관으로 부터 수집되는 데이터를 바탕으로 학습
- 인간의 인지적 학습법을 모방하여 다양한 형태(modality) 데이터로 학습하는 방법  
→ Multimodal learning



# Introduction

## ❖ What is the 'Multimodal' ?

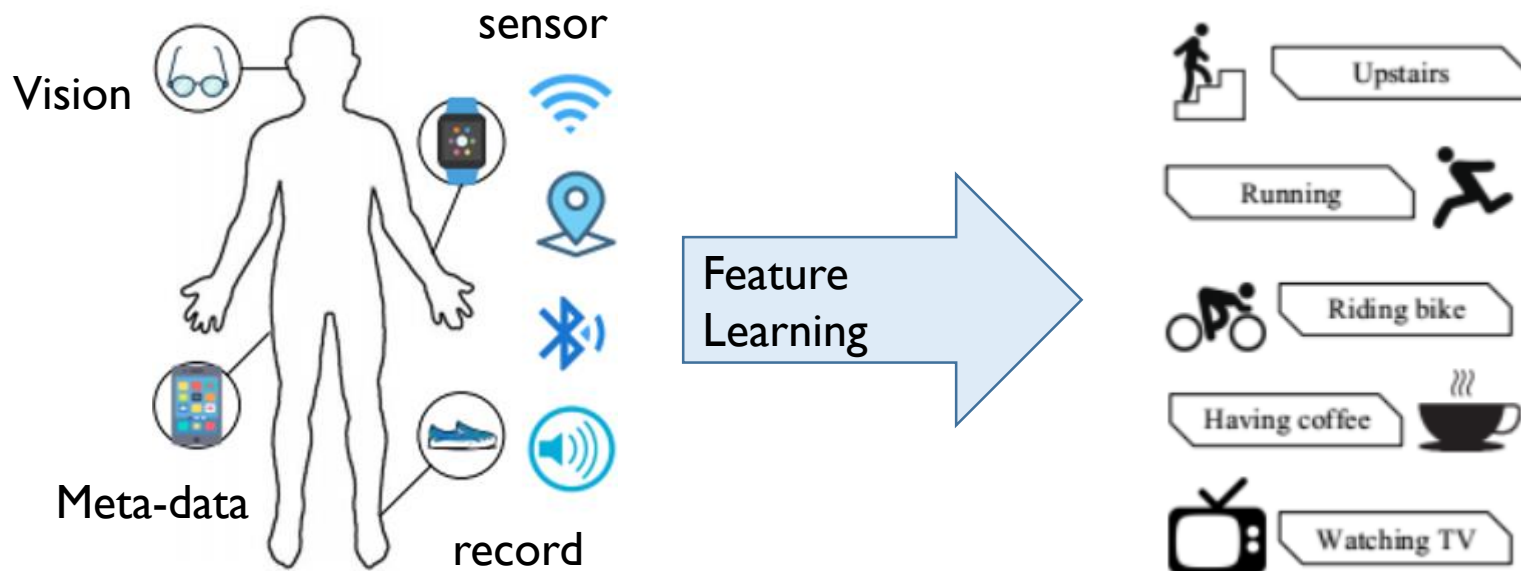
- Multimodal 이란? → Modality가 여러 개 존재
- Modality(양식): 특정 자원으로부터 수집된 데이터 표현 형식
- Multimodal data: 다양한 자원(source)로부터 수집된 데이터가 하나의 정보를 표현



# Introduction

## ❖ Multimodal learning 기법 사용 분야

- 인간 행동 인식 분야(Human Activity Recognition)




<https://becominghuman.ai/deep-learning-for-sensor-based-human-activity-recognition-970ff47c6b6b>

# Introduction

❖ What is the 'Multimodal Learning'?

$$f(X) = \text{Model}$$




	Text	Image	...	Audio	Sensor	Y
관측치 1	...	...	...	...	...	0
관측치 2	...	...	...	...	...	1
...	...	...	...	...	...	...
관측치 N	...	...	...	...	...	0



# Introduction

❖ What is the 'Multimodal Learning'?

$$f(X) = \text{Model}$$



	변수 1	변수 2	...	변수 P-1	변수 P	Y
관측치 1	...	...	...	...	...	0
관측치 2	...	...	...	...	...	1
...	...	...	...	...	...	...
관측치 N	...	...	...	...	...	0

변수가 많으면 Multimodal? NO! → 변수들의 데이터차원이 달라야 함!

# Introduction

## ❖ What is the 'Multimodal Learning'?

- Single-modal

$$\text{Model} = f(X_n^p)_{n=1, \dots, N, p=1, \dots, P}$$

	변수 1	변수 2	...	변수 P-1	변수 P	Y
관측치 1	...	...	...	...	...	0
관측치 2	...	...	...	...	...	1
...	...	...	...	...	...	...
관측치 N	...	...	...	...	...	0

# Introduction

## ❖ What is the 'Multimodal Learning'?

- Multi-modal

$$\text{Model} = f\left(X_{term}^{doc}, X_{x,y}^{color}, X_{time}^{voice}, X_{time}^{sensor}\right)$$

	Text	Image	...	Audio	Sensor	Y
관측치 1	...	...	...	...	...	0
관측치 2	...	...	...	...	...	1
...	...	...	...	...	...	...
관측치 N	...	...	...	...	...	0

# Introduction

---

## ❖ Single-modal 과 Multi-modal 차이점

- Single-modal 데이터: Image OR Text OR Audio OR ...
- Multi-modal 데이터: Image AND Text AND Audio AND ...



Multimodal Learning Model =  $f(X_{term}^{doc}, X_{x,y}^{color}, X_{time}^{voice}, X_{time}^{sensor})$



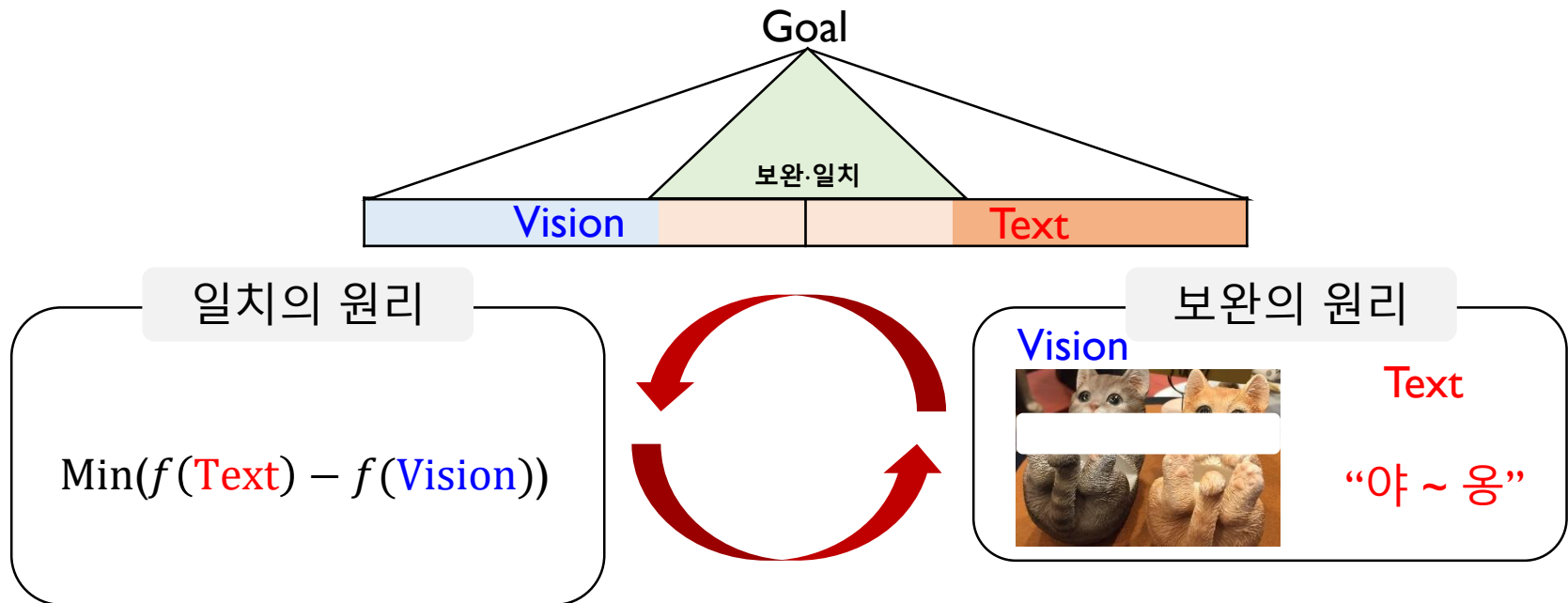
Multimodal learning은 특징 차원이 다른 데이터를 동시에 학습

어떻게 '잘' 학습할 수 있을까?

# Multimodal Deep Learning

## ❖ Multimodal Learning의 지향점

- ① 일치의 원리(consistence principle)
- ② 보완의 원리(complementary principle)

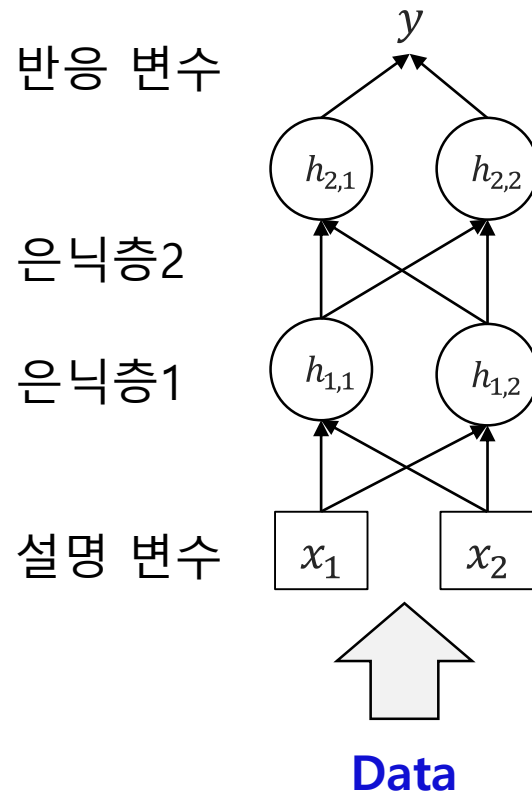


Deep learning 기반 Multimodal Learning 기법 소개

# Multimodal Deep Learning

## ❖ Deep Learning이란?

- 인공 신경망을 여러 층으로 깊게 쌓아 예측력을 높이는 모델
- 인공신경망(Artificial Neural Network, ANN)



→ Y를 잘 예측하는 방향으로.

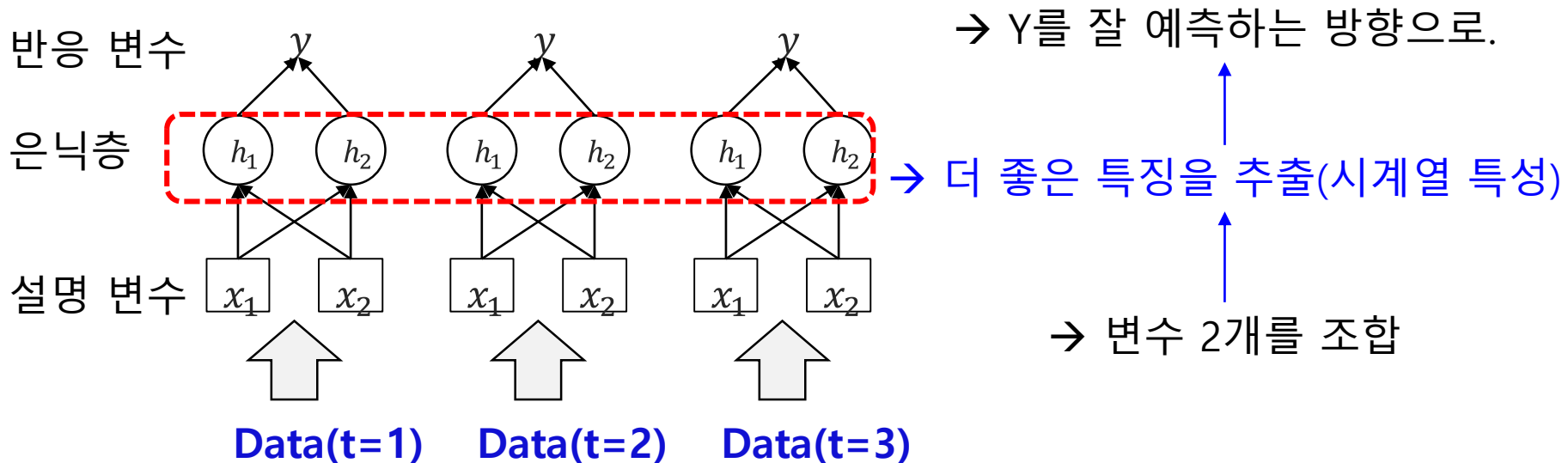
→ 더 좋은 특징을 추출

→ 변수 2개를 조합

# Multimodal Deep Learning

## ❖ Deep Learning이란?

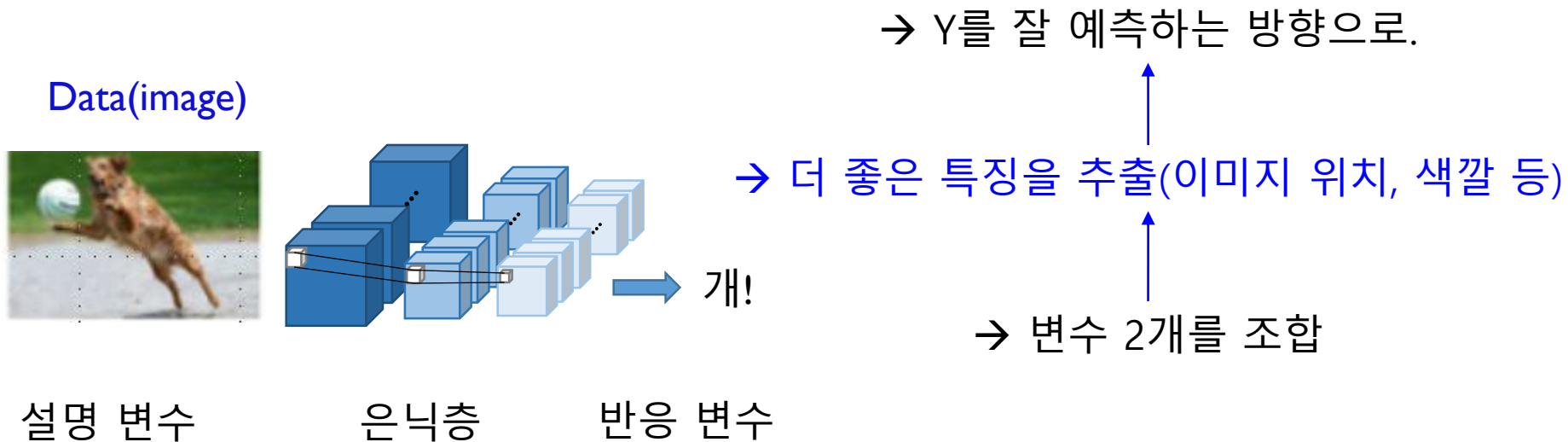
- 인공 신경망을 여러 층으로 깊게 쌓아 예측력을 높이는 모델
- 순환신경망(Recurrent Neural Network, RNN)



# Multimodal Deep Learning

## ❖ Deep Learning이란?

- 인공 신경망을 여러 층으로 깊게 쌓아 예측력을 높이는 모델
- 합성곱신경망(Convolutional Neural Network, CNN)





# Multimodal Deep Learning

---

Multimodal learning은 특징 차원이 다른 데이터를 동시에 학습

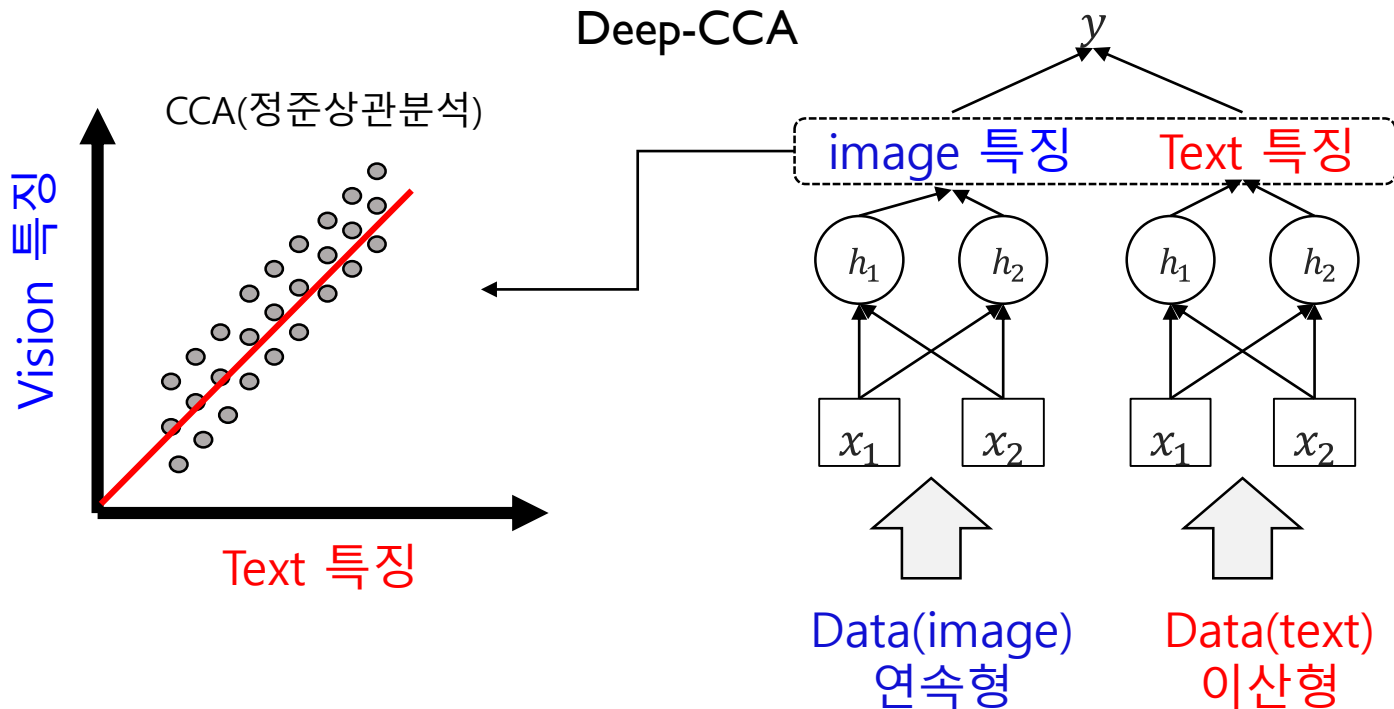
어떻게 '잘' 학습할 수 있을까?

→ 각 데이터 특성을 '잘' 통합해야함.

# Multimodal Deep Learning

❖ Multimodal Learning은 데이터 통합 방식에 따라 구분

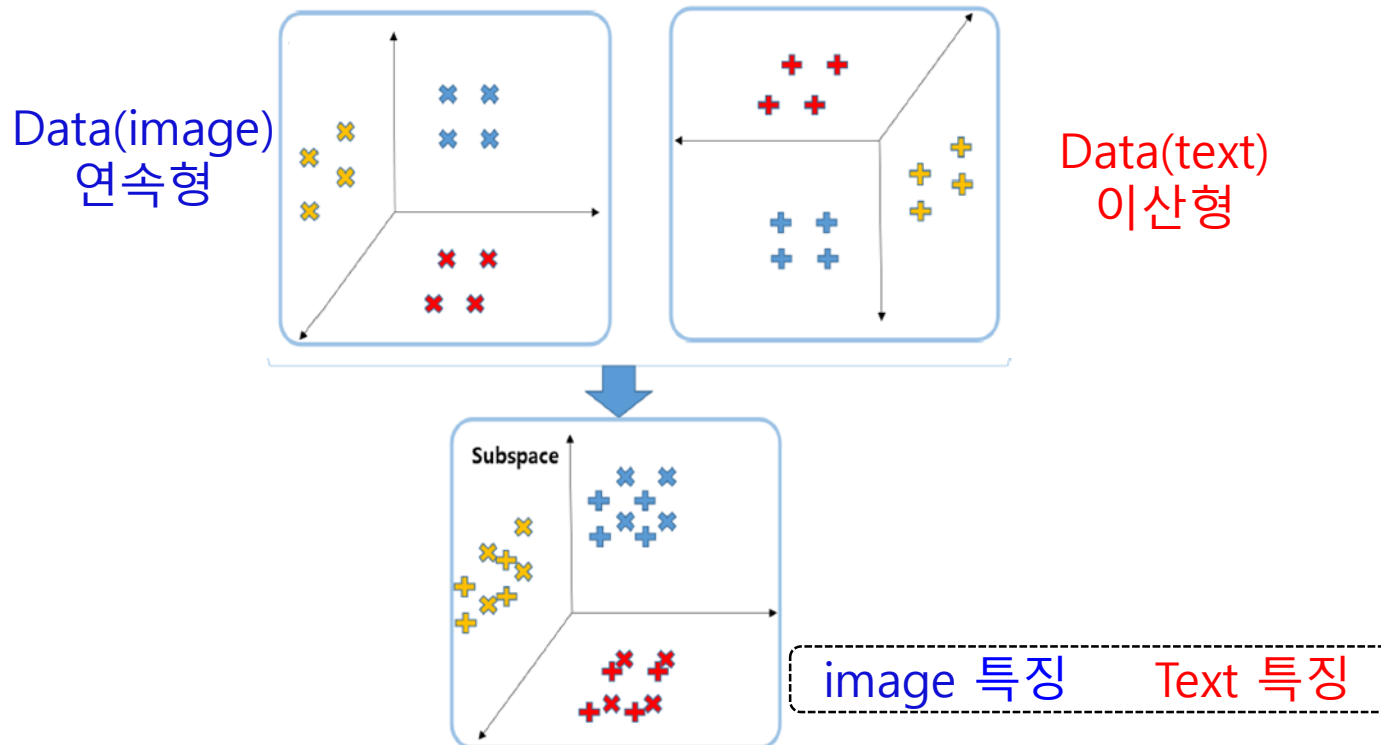
(I) 데이터 ‘차원’의 통합: 다른 특성 데이터가 공유하는 하나의 특징공간으로 투영  
e.g. Deep Canonical Correlation Analysis(Deep CCA, 심층-정준상관분석)



# Multimodal Deep Learning

❖ Multimodal Learning은 데이터 통합 방식에 따라 구분

(I) 데이터 차원의 통합: 다른 특성 데이터를 Embedding 하여 특성이 같은 데이터로 추출  
e.g. Deep Canonical Correlation Analysis(Deep CCA, 심층-정준상관분석)

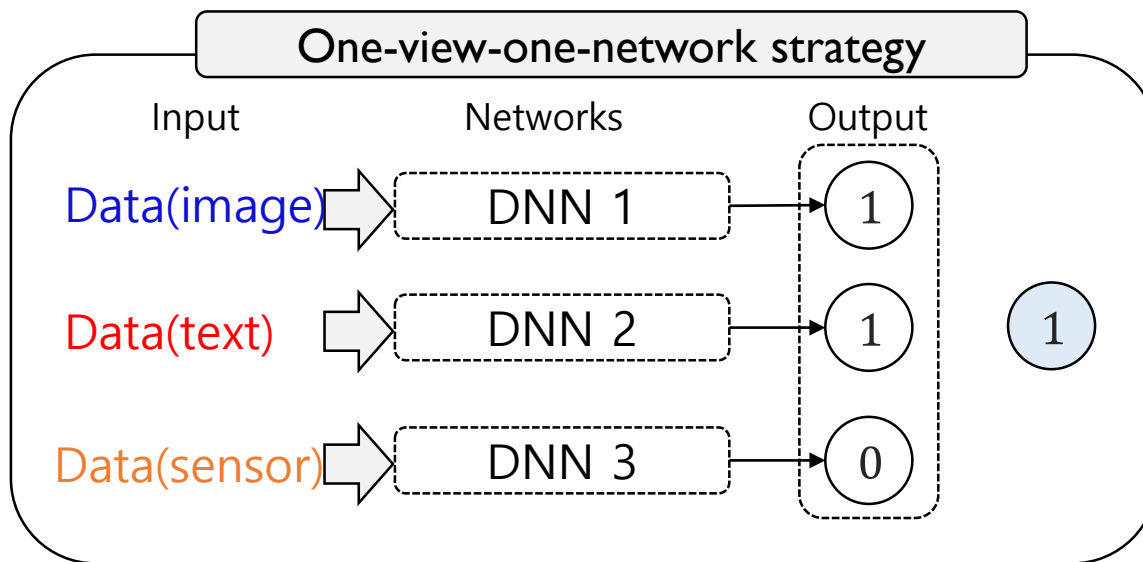


# Multimodal Deep Learning

❖ Multimodal Learning은 데이터 통합 방식에 따라 구분

(2) 분류기 통합: 여러 예측 모델의 결과를 결합하여 예측

→ Co-training, Ensemble



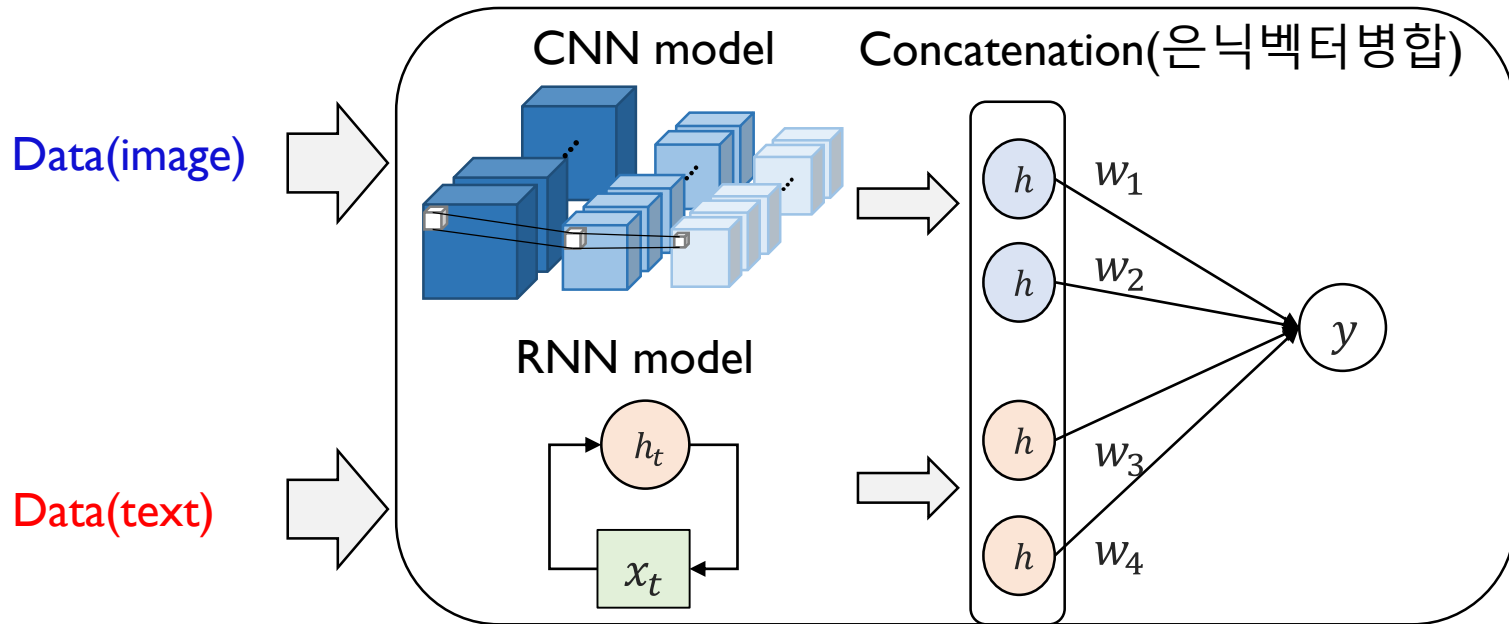
$$\text{목적 함수} \quad P(\hat{Y}_j = c | I_j) = \sum_{\phi \in \{1,2,3\}} \gamma_{\phi} P_{\phi}(\hat{Y}_j = c | I_j; \theta_{\phi})$$

$\gamma$ : 분류 결과에 대한 가중치

# Multimodal Deep Learning

❖ Multimodal Learning은 데이터 통합 방식에 따라 구분

(3) 학습된 표현 간의 통합: 다른 신경망으로 학습하여 추출된 특징을 선형 결합

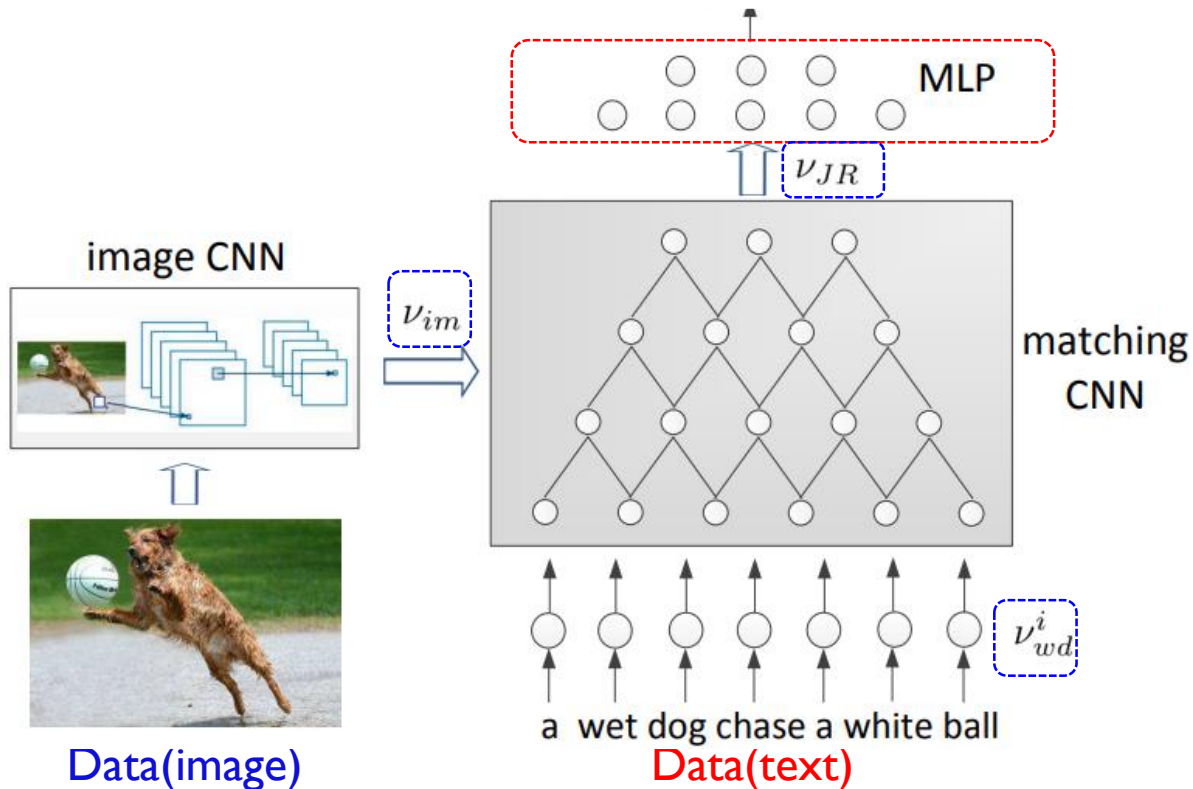


# Multimodal Deep Learning

## ❖ Multimodal CNN(m-CNN)

- Multimodal CNN: 이미지와 연관된 텍스트의 관계를 학습(Ma et al, 2015)

$Y = \text{Data}(\text{image}) \text{와 } \text{Data}(\text{text}) \text{ 유사도}$



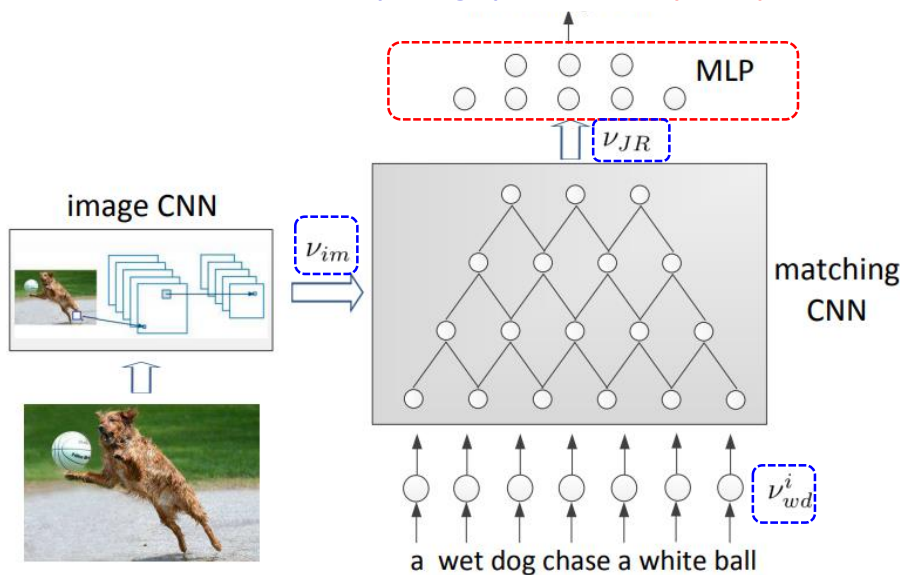
Ma, L., Lu, Z., Shang, L., & Li, H. (2015). Multimodal convolutional neural networks for matching image and sentence. In *Proceedings of the IEEE international conference on computer vision* (pp. 2623-2631)

# Multimodal Deep Learning

## ❖ Multimodal CNN(m-CNN)

- Multimodal CNN: 이미지와 연관된 텍스트의 관계를 학습(Ma et al, 2015)
- $S_{match}$  : 텍스트와 이미지 결합표현( $V_{JR}$ )을 기준으로 인공신경망 형성

Y = Data(image)와 Data(text) 유사도



목적함수

$$s_{match} = w_s (\sigma(w_h)(v_{JR}) + b_h) + b_s$$

예측에 집중하는 신경망

$$\mathbf{v} = v_{wd}^i \parallel v_{wd}^{i+1} \parallel \dots \parallel v_{wd}^{i+k_{rp}-1} \parallel v_{im}$$

특징을 결합하는 신경망





# Multimodal Deep Learning

---

## ❖ Multimodal Deep Learning 요약

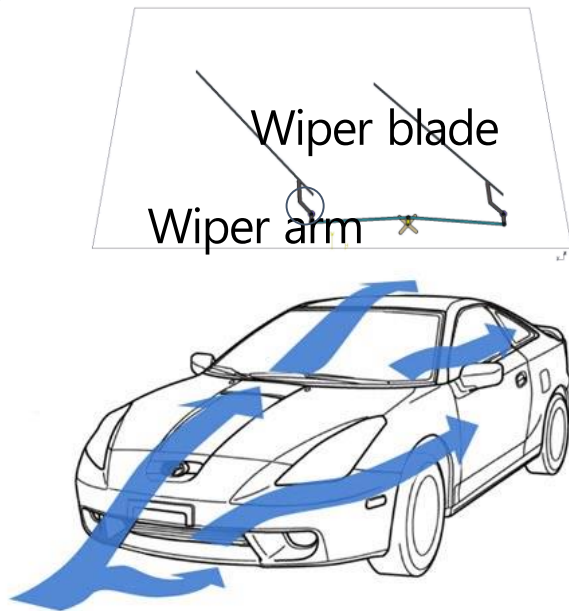
- Deep learning 기법의 도입으로 새로운 변화를 맞이함.
- Multimodal learning 기법은 인간행동인식, 의학, 정보검색, 표정인식 등 다양한 분야에서 문제를 해결
- Singlemodal Learning 기법과 달리 여러가지 정보를 보완적으로 이용
- 목표: 다양한 정보를 이용하여 단일 정보만 사용했을 때보다 학습 성능 ↑
- 현재 진행하고 있는 프로젝트에 적용

# Applications

## ❖ Multimodal Deep Learning for Product Design Data

- 후드 형상 이미지 및 와이퍼 설계 데이터 기반 고속 부상 속도 예측

차량 주행 시험



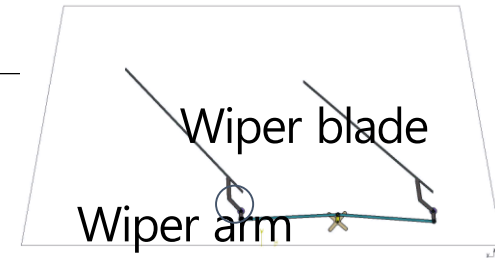
와이퍼 부상



부상 속도

낮으면 불합격!  
⋮  
140 km  
150 km  
-----  
160 km  
170 km  
⋮  
높으면 Good!

# Applications



## ❖ Multimodal Deep Learning for Product Design Data

- 와이퍼 설계 데이터 및 후드 형상 이미지 기반 고속 부상 속도 예측

	Wiper blade 설계치(각도) 1	Wiper blade 설계치(각도) 2	...	Wiper arm 설계치(각도) 1	Wiper arm 설계치(각도) 2	속도
관측치 1			...			140
관측치 2			...			150
...	...	...	...	...	...	...
관측치 N			...			170

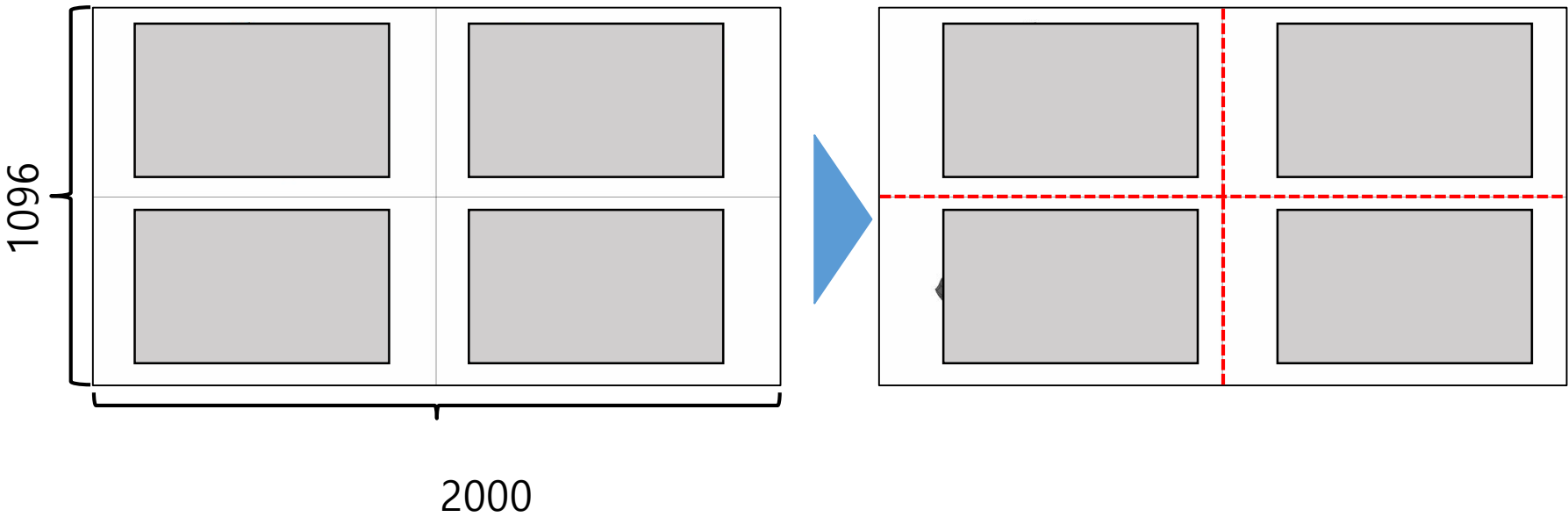
# Applications

## ❖ Multimodal Deep Learning for Product Design Data

- 와이퍼 설계 데이터 및 **후드 형상 이미지** 기반 고속 부상 속도 예측

관측치 1(color-RGB)

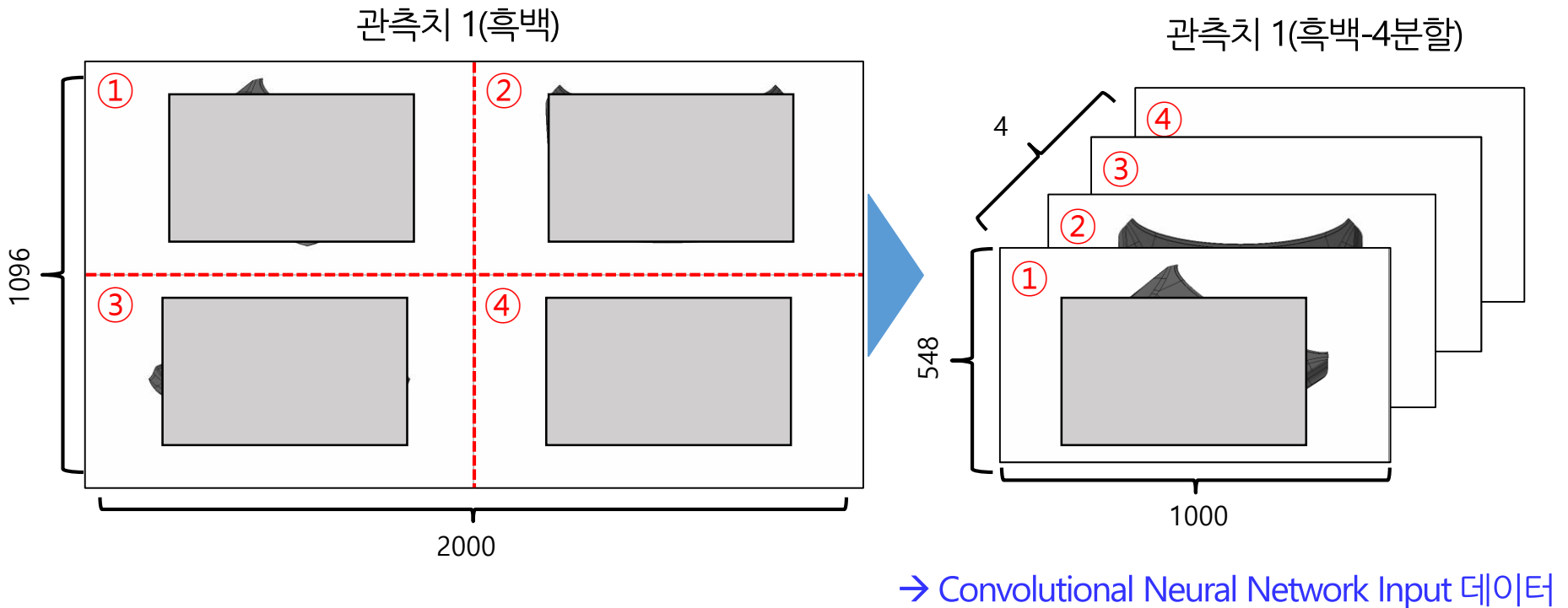
관측치 1(흑백)



# Applications

## ❖ Multimodal Deep Learning for Product Design Data

- 와이퍼 설계 데이터 및 **후드 형상 이미지** 기반 고속 부상 속도 예측

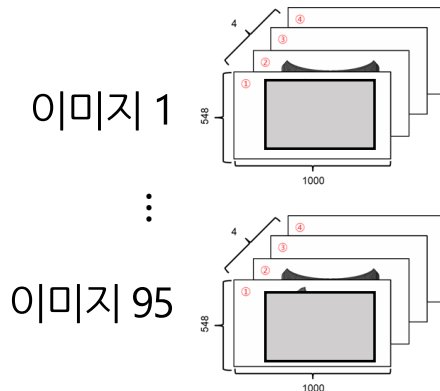


# Applications

## ❖ Multimodal Deep Learning for Product Design Data

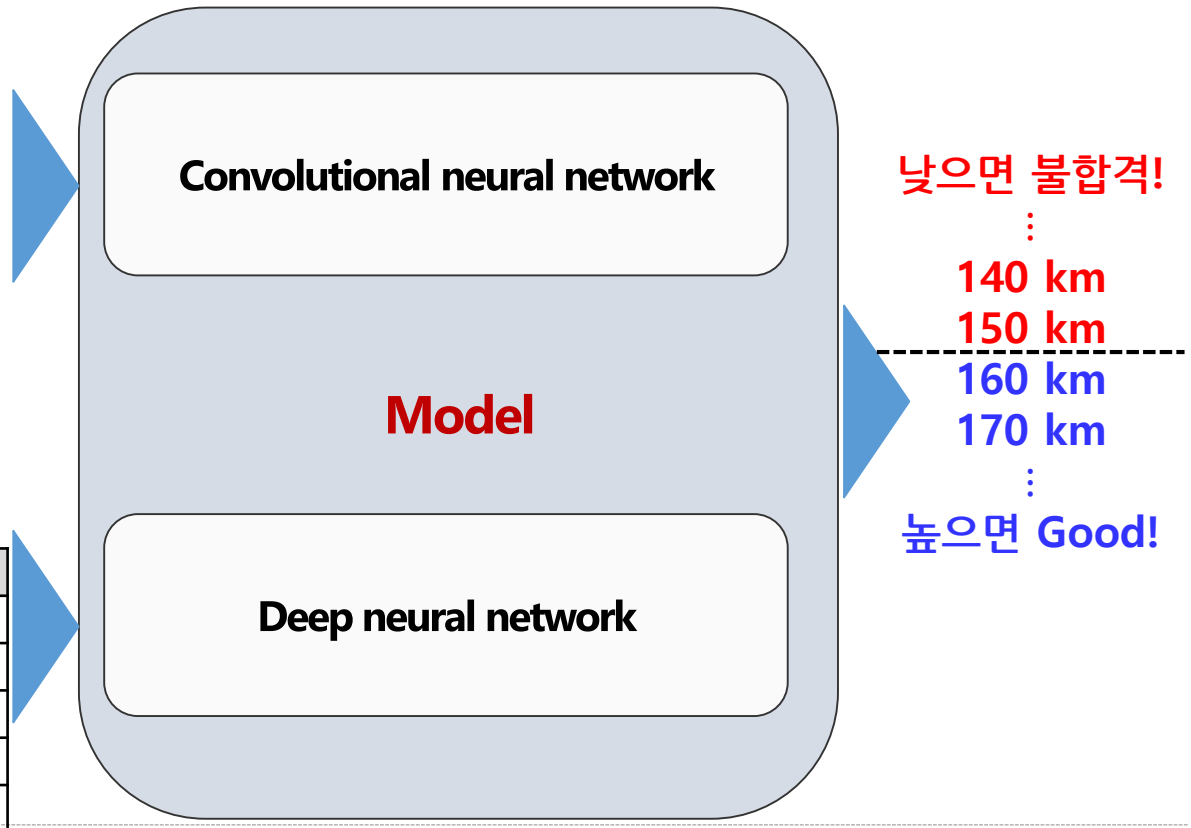
- 와이퍼 설계 데이터 및 후드 형상 이미지 기반 고속 부상 속도 예측

후드형상이미지



와이퍼 설계 데이터

관측치 ID	와이퍼 설계데이터
1	...
2	...
...	...
94	...
95	...

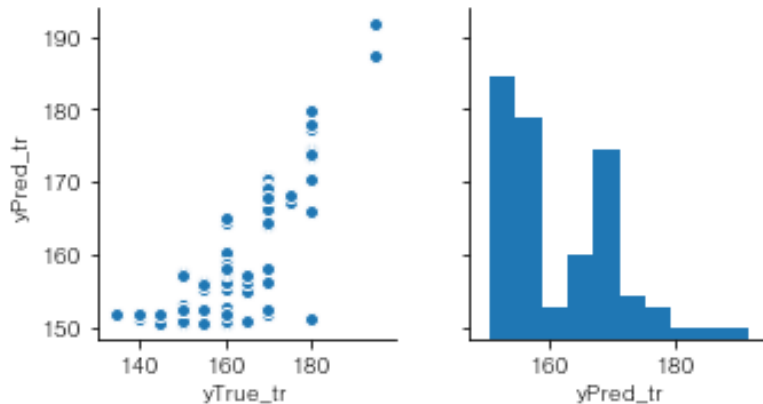


# Applications

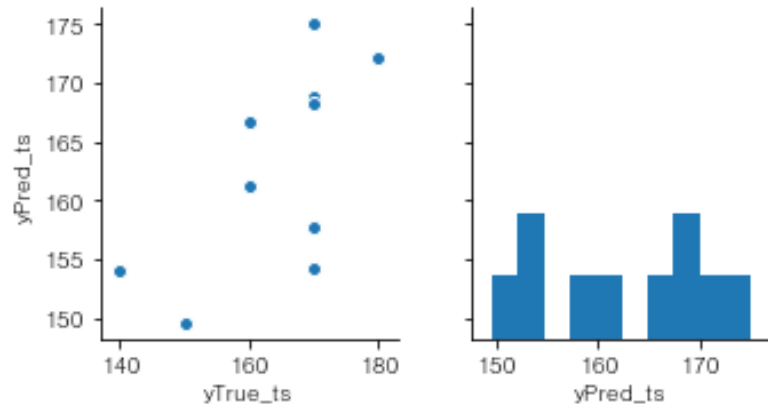
## ❖ Multimodal Deep Learning for Product Design Data

- 후드 형상 이미지 및 와이퍼 설계 데이터 기반 고속 부상 속도 예측

학습데이터 예측성능



테스트데이터 예측성능



# Applications

## ❖ Multimodal Deep Learning for Product Design Data

- 후드 형상 이미지 및 와이퍼 설계 데이터 기반 고속 부상 속도 예측
- Single modal과 Multimodal learning의 성능 비교

예측 모델	데이터	MAE	MSE
LASSO	Single modal-와이퍼 설계 데이터	6.76	72.02
CNN	Single modal-후드 형상 이미지 데이터	9.14	132.58
DNN+CNN	Multimodal - 와이퍼 설계 및 후드 형상 이미지	<b>6.62</b>	<b>73.7</b>



# Conclusions

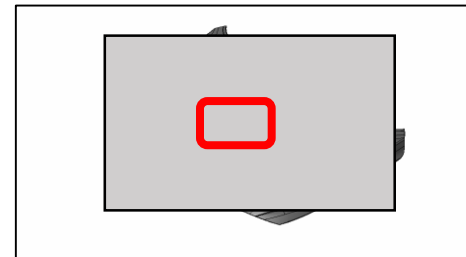
## ❖ Summary

- 후드 이미지 및 와이퍼 설계 데이터 기반 Multimodal Learning 기법을 적용
- 두 데이터를 모두 적용했을 때 더 좋은 성능을 보임

## ❖ Future work: Explainable Multimodal Deep learning

- 인자분석을 위한 Multimodal 모델 구조 디자인

관측치 ID	설계 인자1		설계 인자1
1	...	...	...
2	...	...	...
...	...	...	...
94	...	...	...
95	...	...	...



# Multimodal Deep Learning

멀티 뷰 기법	모형	언어	패키지	비고
데이터 차원의 통합	CCA	R	CCA	
		R	stats	
		Python	sklearn	
	KCCA	R	kernlab	
		Python	pyrcca	<a href="https://github.com/gallantlab/pyrcca">https://github.com/gallantlab/pyrcca</a>
	GCCA	R	drCCA	
		R	RGCCA	
		Python	numpy 기반	<a href="https://github.com/rupy/GCCA">https://github.com/rupy/GCCA</a>
분류기 차원의 통합	Co-training	R	SSL	
		Python	sklearn 기반	<a href="https://github.com/jjrob13/sklearn_cotraining/blob/master/sklearn_cotraining/classifiers.py">https://github.com/jjrob13/sklearn_cotraining/blob/master/sklearn_cotraining/classifiers.py</a>
	Seq2seq	Python	Tensorflow 기반	
학습된 표현 간의 통합	m-RNN	Python	Tensorflow 기반	<a href="http://www.stat.ucla.edu/~junhua.mao/m-RNN.html">http://www.stat.ucla.edu/~junhua.mao/m-RNN.html</a>
	Deep CCA	Python	Tensorflow 기반	<a href="https://github.com/VahidooX/DeepCCA">https://github.com/VahidooX/DeepCCA</a>
		Python	Theano 기반	<a href="https://github.com/msamribeiro/deep-ccaA">https://github.com/msamribeiro/deep-ccaA</a>
	multimodal DBM	Python	Tensorflow 기반	<a href="https://github.com/abyoussef/DRBM_Project">https://github.com/abyoussef/DRBM_Project</a>
	Corr-AE	Python	numpy 기반	<a href="https://github.com/huyt16/Twitter100k">https://github.com/huyt16/Twitter100k</a>
	JMVAE	Python	Theano 기반	<a href="https://github.com/masa-su/jmvae">https://github.com/masa-su/jmvae</a>
	Text-Image GAN	Python	Tensorflow 기반	<a href="https://github.com/paarthneekhara/text-to-x-image">https://github.com/paarthneekhara/text-to-x-image</a>

Multi-view learning review understanding methods(2019)

# Thank You

# Multimodal Deep Learning

---

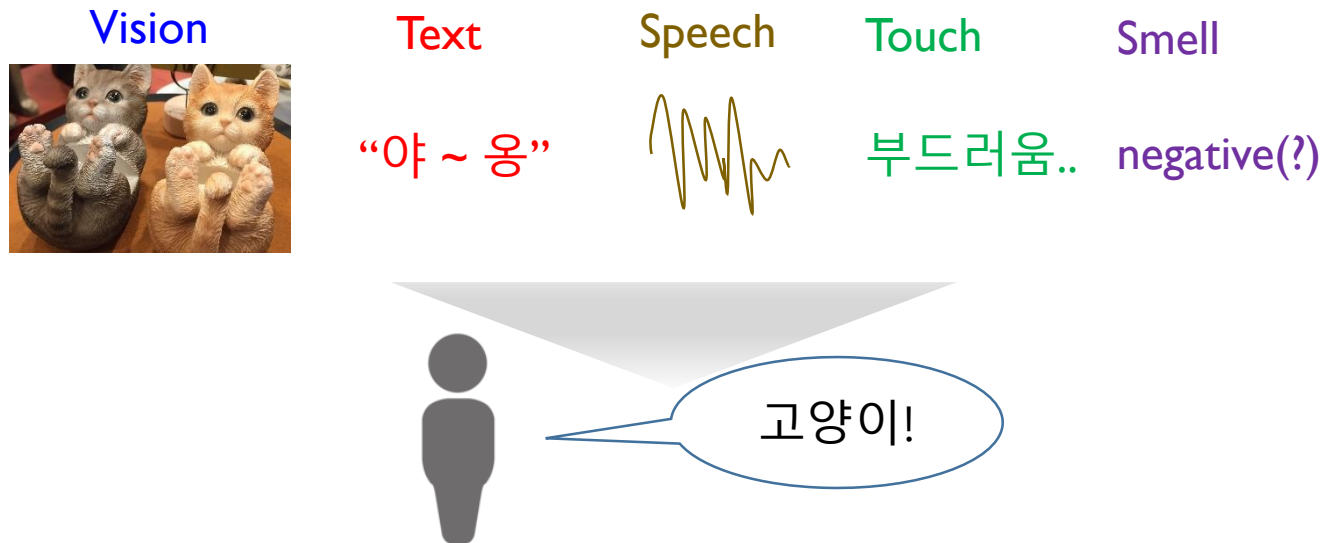
?

Multimodal 을 인공지능망 모델을 여러 층으로 쌓아 데이터 학습

# Introduction

## ❖ What is the ‘Multimodal Learning’ ?

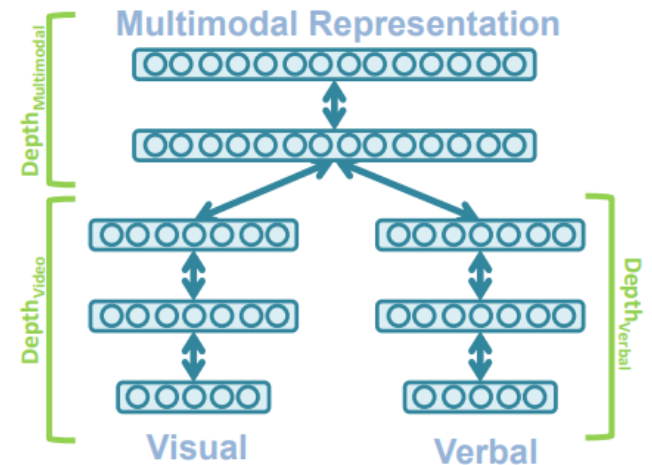
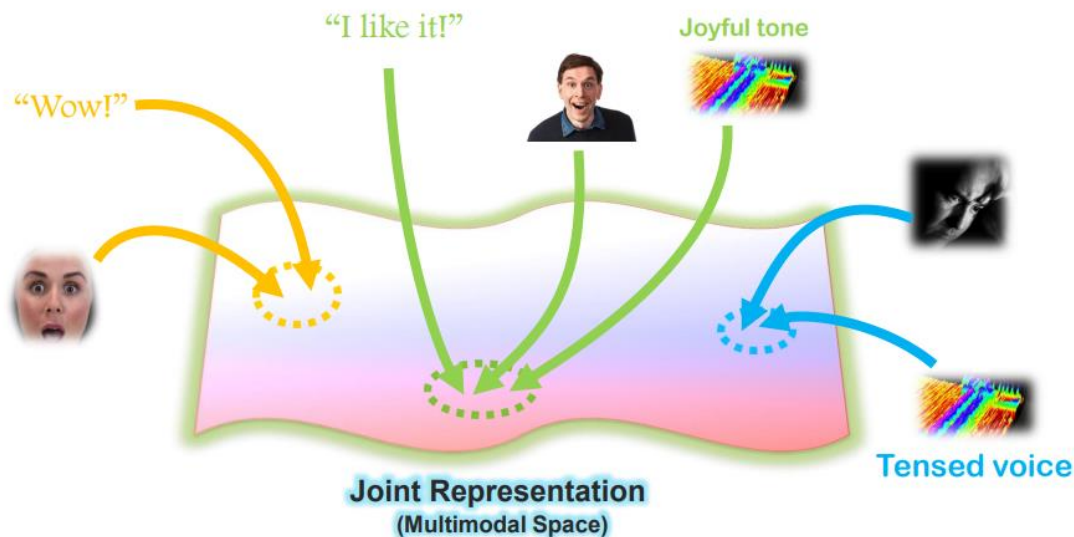
- 인간은 어떤 사물을 인식하거나 예측할 때 다양한 정보를 활용
- 인간의 인지적 학습방법을 모방하여 다양한 형태(modality) 데이터로부터 학습하는 방법



# Multimodal Deep Learning

## ❖ Core Challenge I: Representation

- Definition: Learning how to represent and summarize multimodal data in away that exploits the complementarity and redundancy.
- Learn linear projections that are maximally correlated:

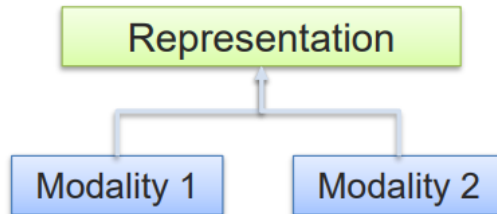


# Multimodal Deep Learning

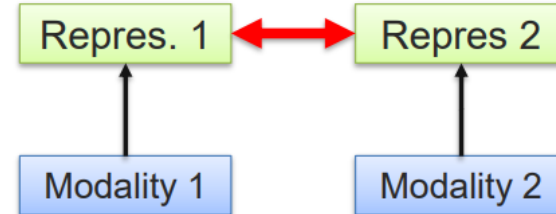
## ❖ Core Challenge I: Representation

- Definition: Learning how to represent and summarize multimodal data in away that exploits the complementarity and redundancy.
- Learn linear projections that are maximally correlated:

### Ⓐ Joint representations:



### Ⓑ Coordinated representations:

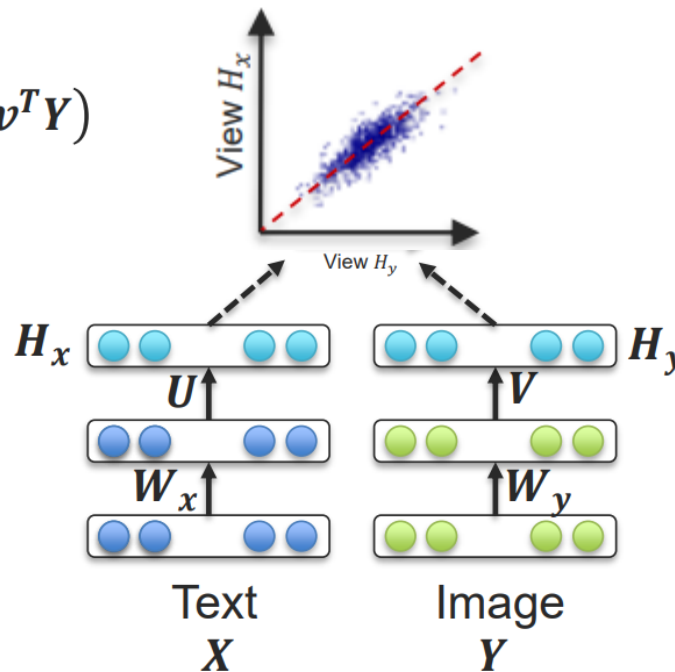


# Multimodal Deep Learning

## ❖ Core Challenge I: Representation

- Coordinated Representation: Deep CCA
- Learn linear projections that are maximally correlated:

$$(u^*, v^*) = \underset{u, v}{\operatorname{argmax}} \operatorname{corr}(u^T X, v^T Y)$$

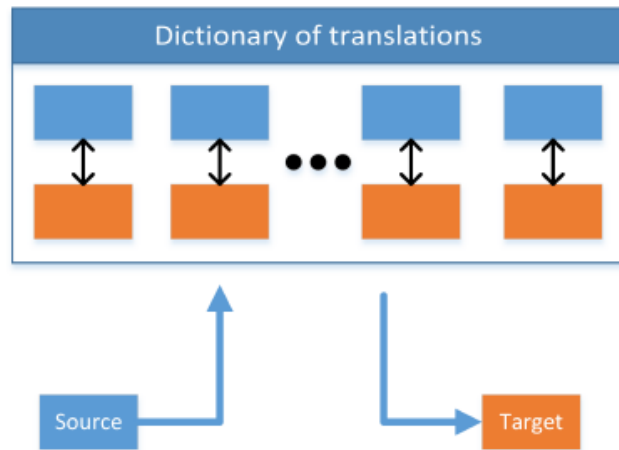




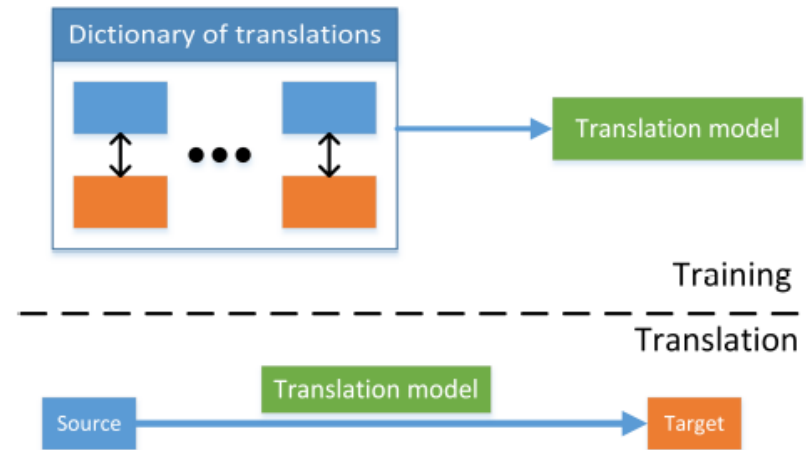
# Multimodal Deep Learning

## ❖ Core Challenge 2: Translation

- Definition: Given an entity in one modality the task is to generate the same entity in a different modality.



(a) Example-based

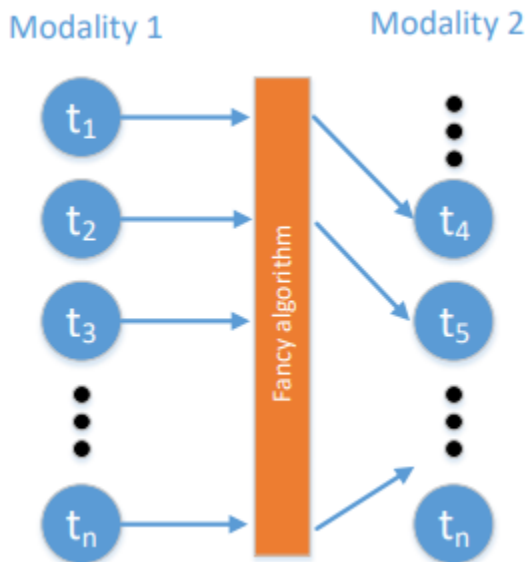


(b) Generative

# Multimodal Deep Learning

## ❖ Core Challenge 3: Alignment

- Definition: Identify the direct relations between (sub)elements from two or more different modalities.



### A Explicit Alignment

The goal is to directly find correspondences between elements of different modalities

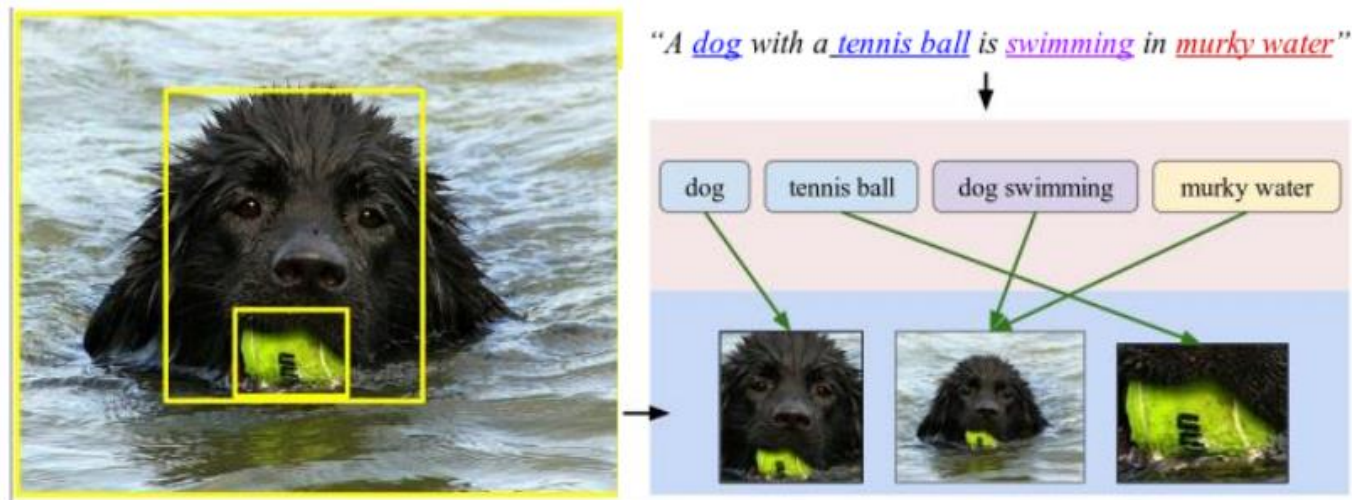
### B Implicit Alignment

Uses internally latent alignment of modalities in order to better solve a different problem

# Multimodal Deep Learning

## ❖ Core Challenge 3: Alignment

- Definition: Identify the direct relations between (sub)elements from two or more different modalities.
- Implicit Alignment



# Multimodal Deep Learning

## ❖ Core Challenge 4: Fusion

- Definition: To join information from two or more modalities to perform a prediction task.

## ❖ Core Challenge 5: Co-learning

- Definition: Co-learning is aiding the modeling of a (**resource poor**) modality by exploiting knowledge from another (**resource rich**) modality.

