Alchera

알체라의 상용화 노력 및 최근 딥러닝 관련 이슈들

황영규 부대표

AI Team

Outline

- Why Deep Learning?
- Deep Learning vs Traditional Machine Learning
- Commercialization using AI
- Trends

Why Deep Learning?

- Deep Learning은
  - 기술적인 진보를 이룸
  - 이론적 및 구현적으로 쉬움

    (전통적인 Machine Learning에 비해서)

시각지능의 역사

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Goals - General

The primary goal of the project is to construct a system of programs which will divide a vidisector picture into regions such as

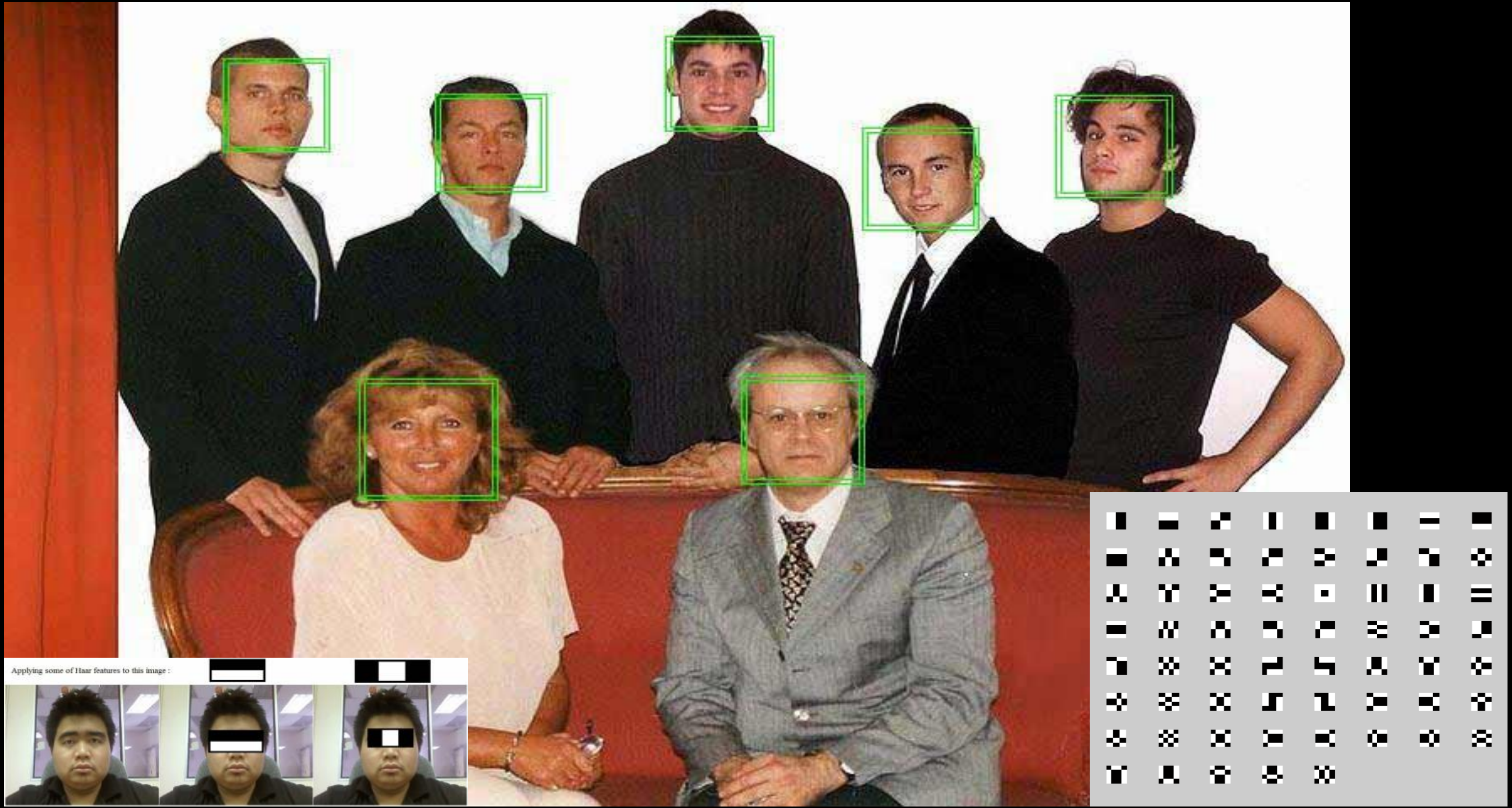
likely objects

likely background areas

chaos.

Open된 환경에서 영상인식 상용화 (1/2)

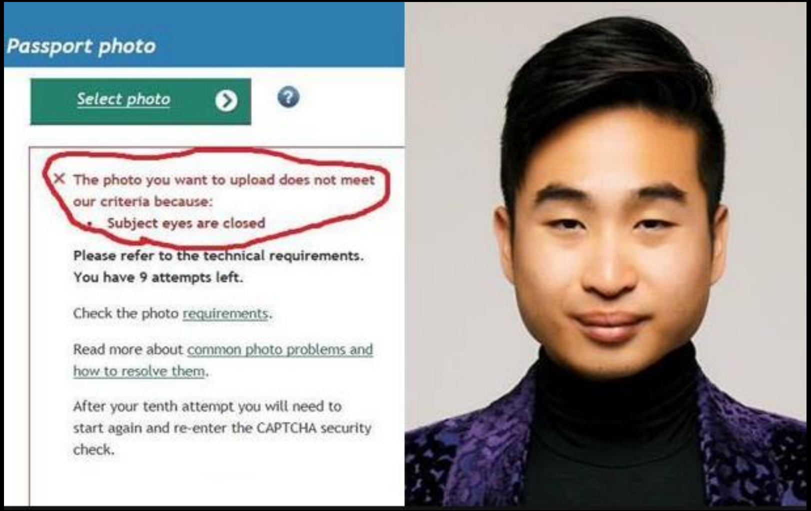
Face Detection, Viola & Jones, 2001



Open된 환경에서 영상인식 상용화 (2/2)

그렇다면 얼굴 검출 이후에는??

- 제한된 환경에서만 적용



딥러닝 이전에는

Alchera

고려대학교한민홍교수 만나봅시다/ 무인자동차를 만드는

자동차생활 피플 0 7,761 좋아요 9,714 2000.03.26 00:00

2000.03.26

2007년 DARPA-Challenge, CMU

딥러닝 이후 (2016년)

자율주행을 위한 Device

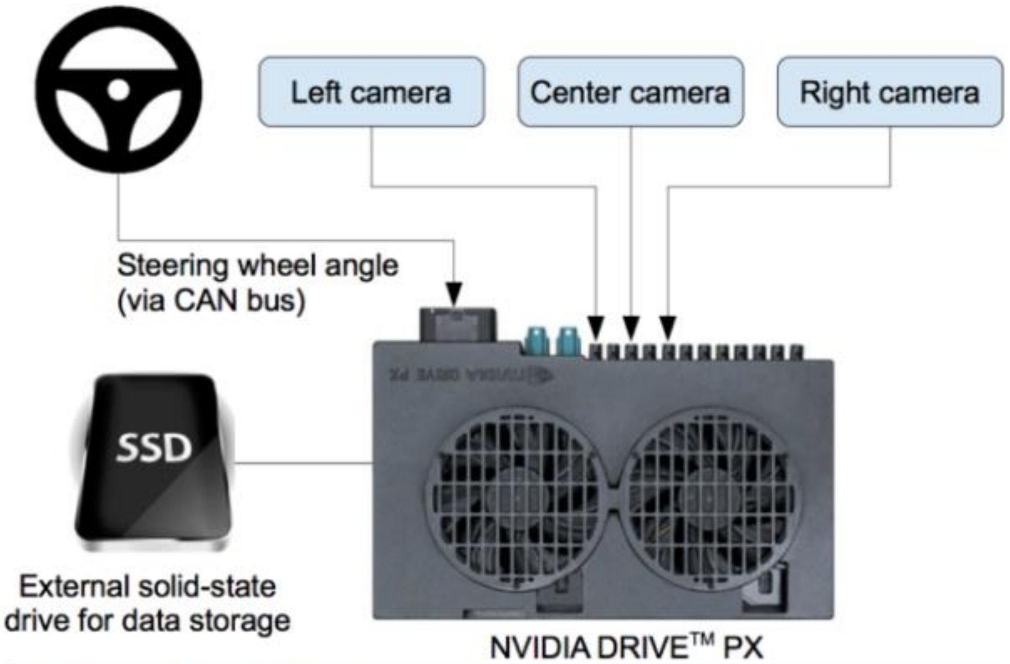


Figure 2: High-level view of the data collection system.

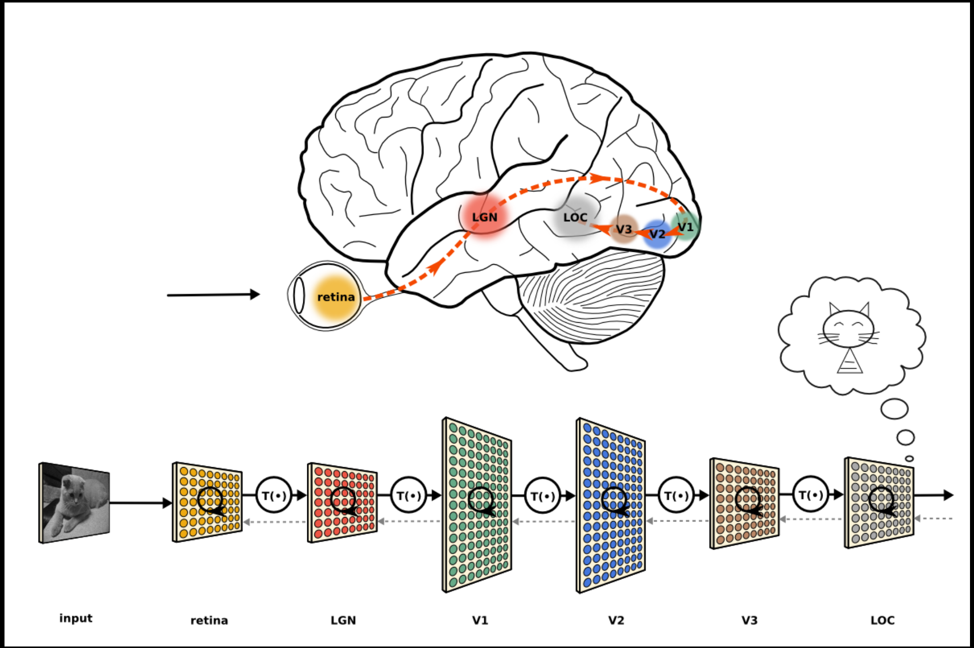
Why Deep Learning

- Deep Learning은
  - 기술적인 진보를 이룸 ➔ 잘되고
  - 이론적 및 구현적으로 쉬움 ➔ Cost Effective

    (전통적인 Machine Learning에 비해서)

Definition of Deep Learning

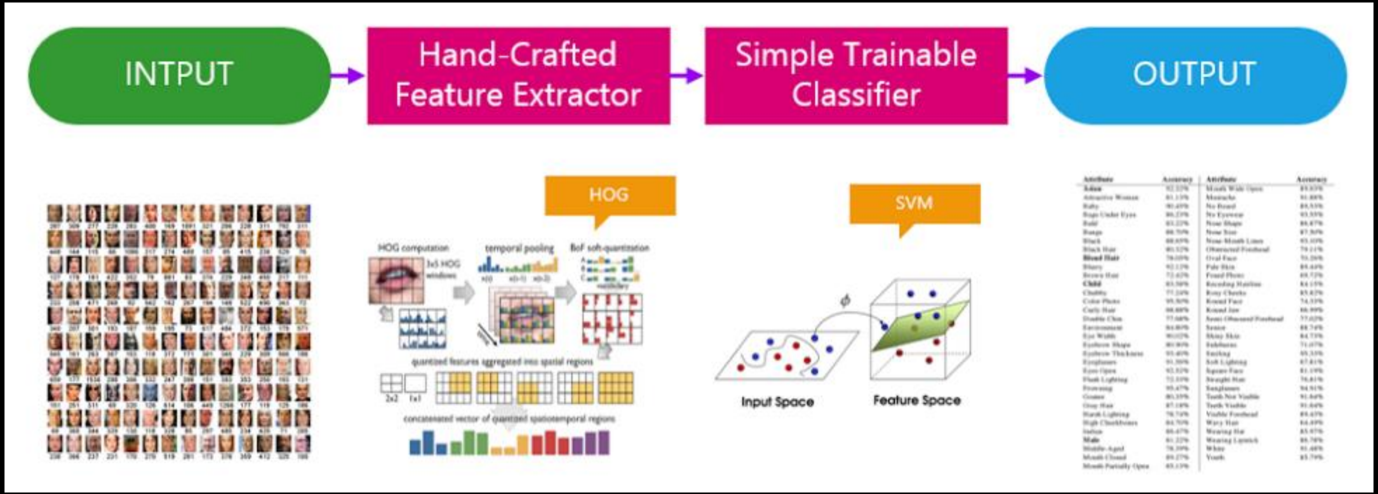
- Deep Neural Network ➔ Brain-inspired AI



Brain inspired technology

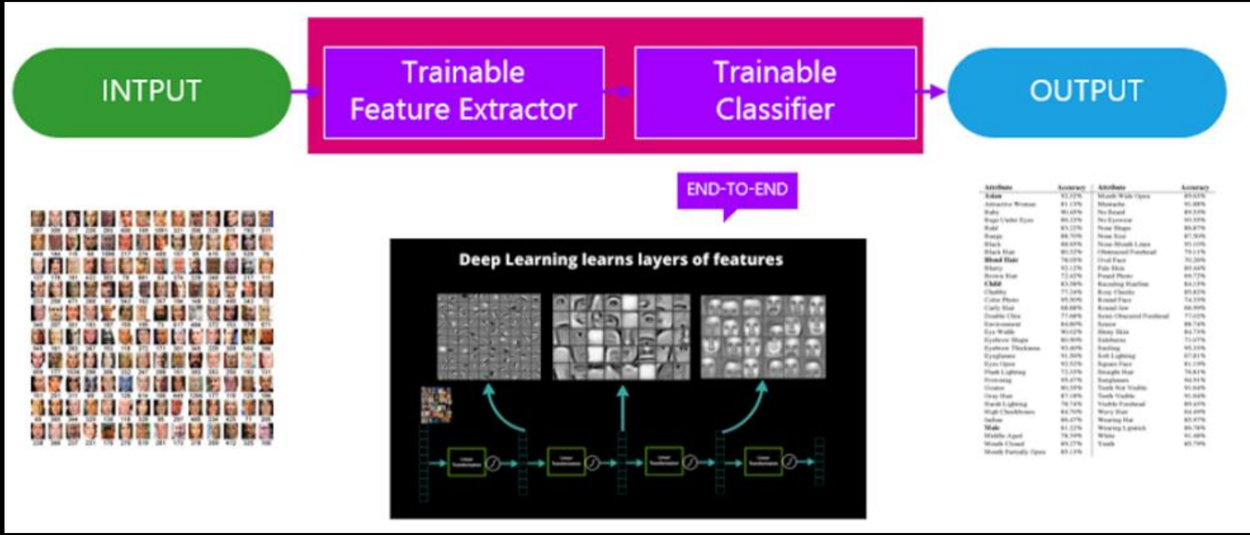
Traditional ML vs Deep Learning

Traditional Machine Learning



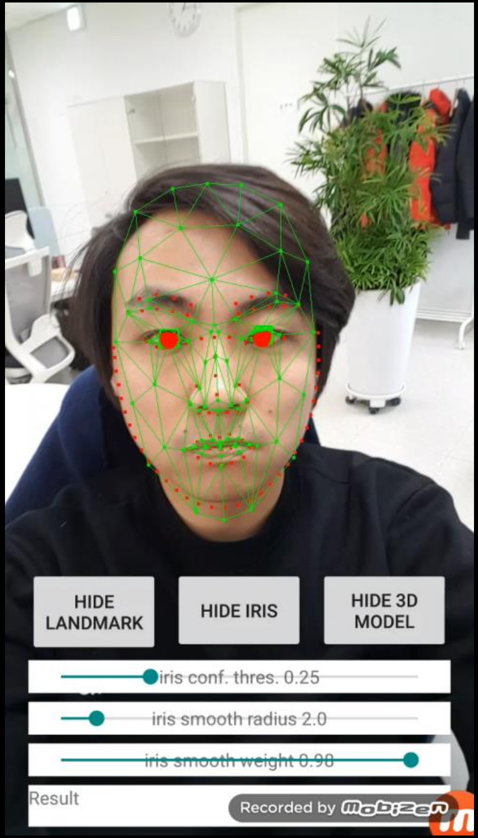
Traditional ML vs Deep Learning

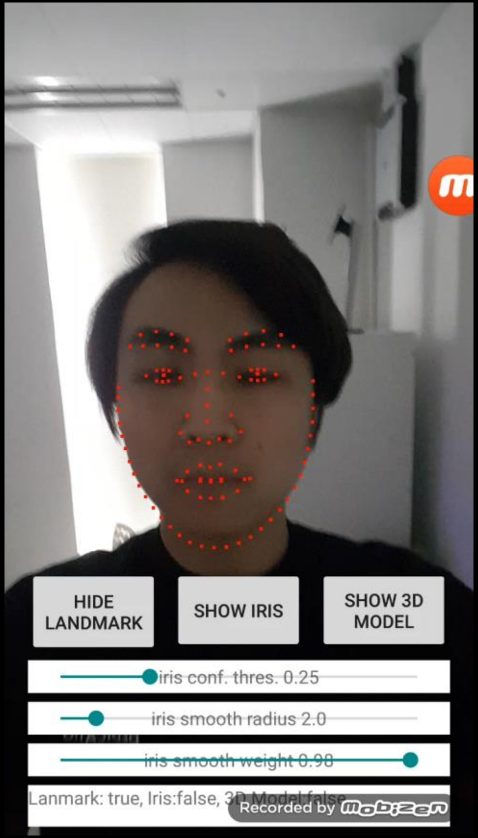
Deep Learning



AI의 상용화 Example

1. 딥러닝 기반 얼굴검출 + Landmark





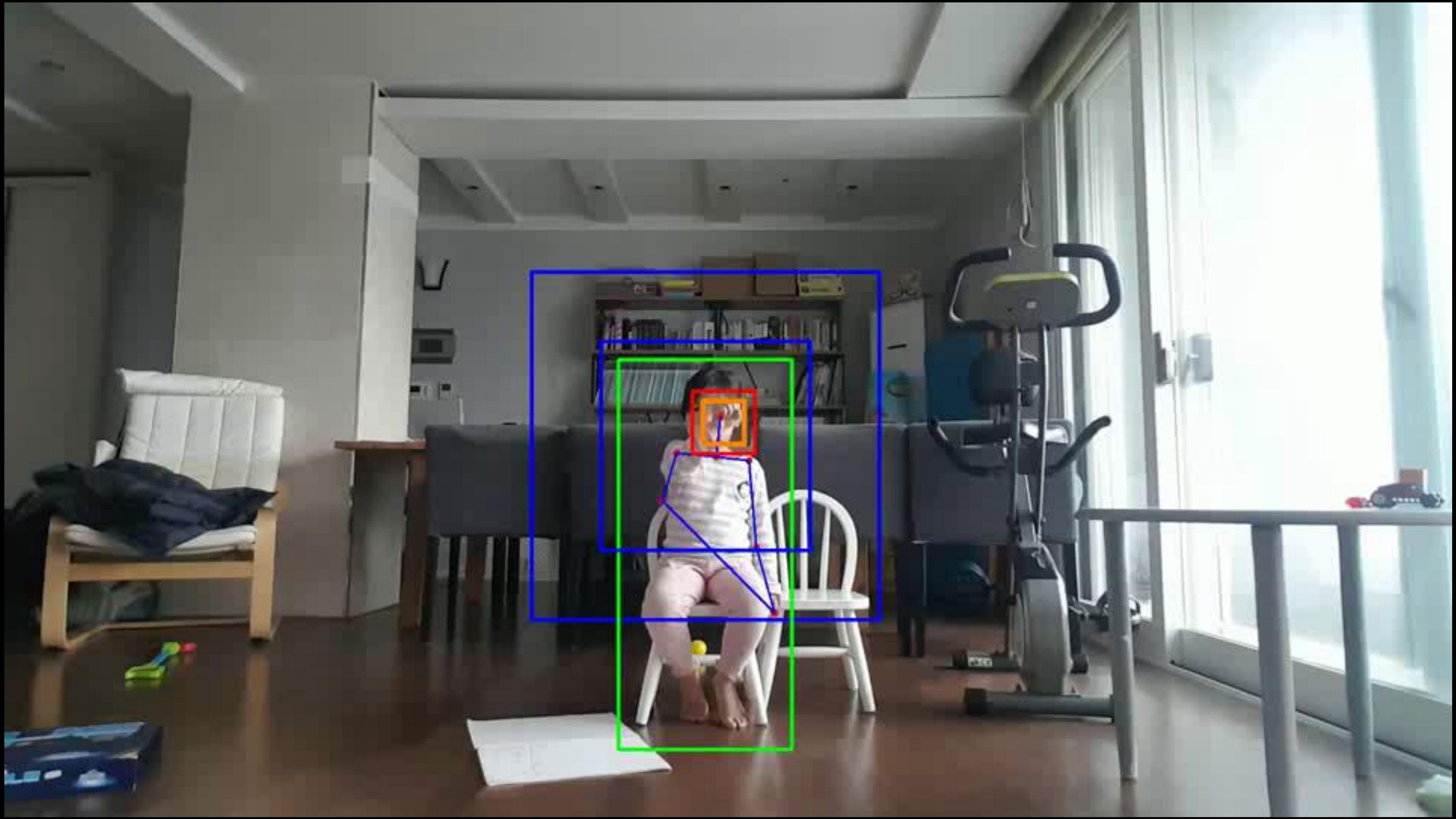
1. 딥러닝 기반 3D 얼굴 형상 추출

Alchera

1. 딥러닝 기반 얼굴 인식

- Face Identification
  : 99.8% @ LFW (6000쌍의 얼굴 영상)
  : 95.8% @ Megaface DB (1백만명의 얼굴영상)

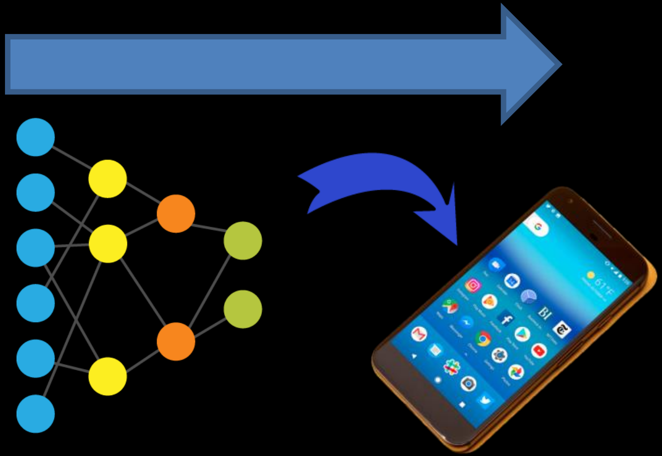
2. 딥러닝 기반 손모양 검출



2. 딥러닝 기반 3D Hand Skeleton 검출

- AI 기술을 이용하여 단일 카메라로 3D 정보 추출
- 저비용의 카메라 혹은 Device에서 실시간으로 작동가능

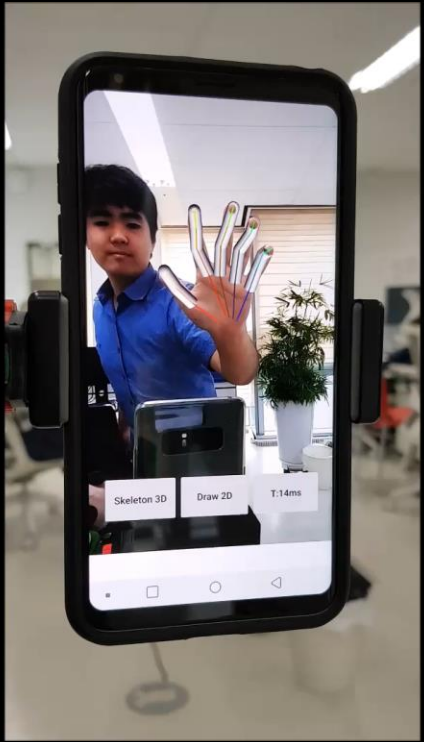
입력 영상



Mobile Deep Learning Framework
(Fast Computation and Low Power Consumption)

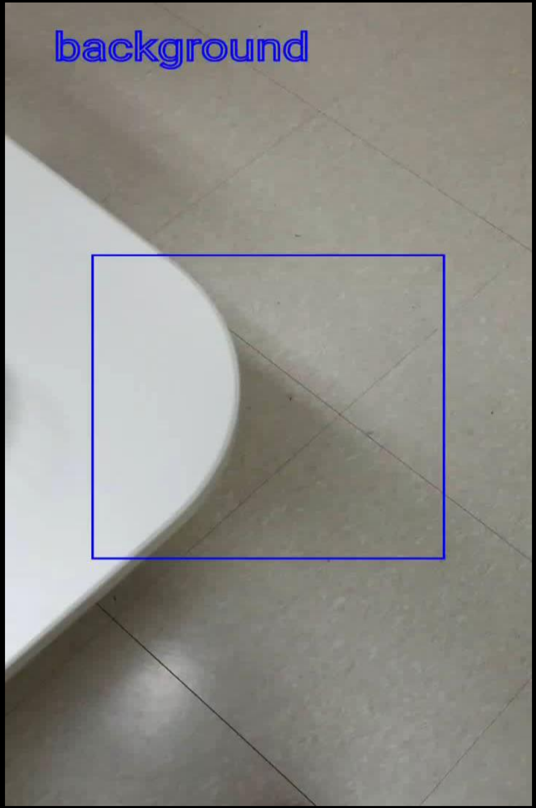
3D Hand Skeleton

2. 딥러닝 기반 3D Hand Skeleton 검출



3. 딥러닝 기반 사물인식

- 20가지 종류의 과일 인식

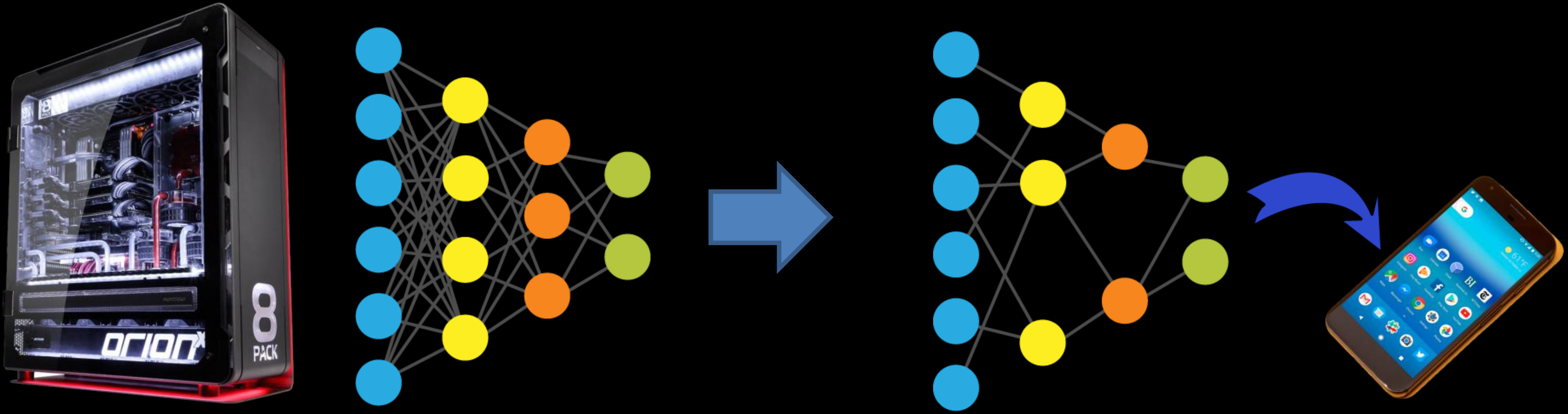


4. 딥러닝 기반 객체 검출 및 인식

- 7가지 종류의 객체의 위치 및 종류 인식 (4K영상에서 30x30까지 검출 가능)

5. 딥러닝 경량화 및 고속화

- 고성능 PC에서 동작하는 딥러닝 엔진을 저사양의 모바일 Device에서도 실시간으로 동작하게 함.



Deep Learning Framework

Mobile Deep Learning Framework
(Fast Computation and Low Power Consumption)

딥러닝 상용화 추세

AI goes to 상용화

Alchera

But.. 상용화는 막대한 돈과 노동력이 필수

상용화에 제일 중요한 것..

COST EFFECTIVE

상용화에 제일 중요한 것..

딥러닝 개발은 프로세스

Data

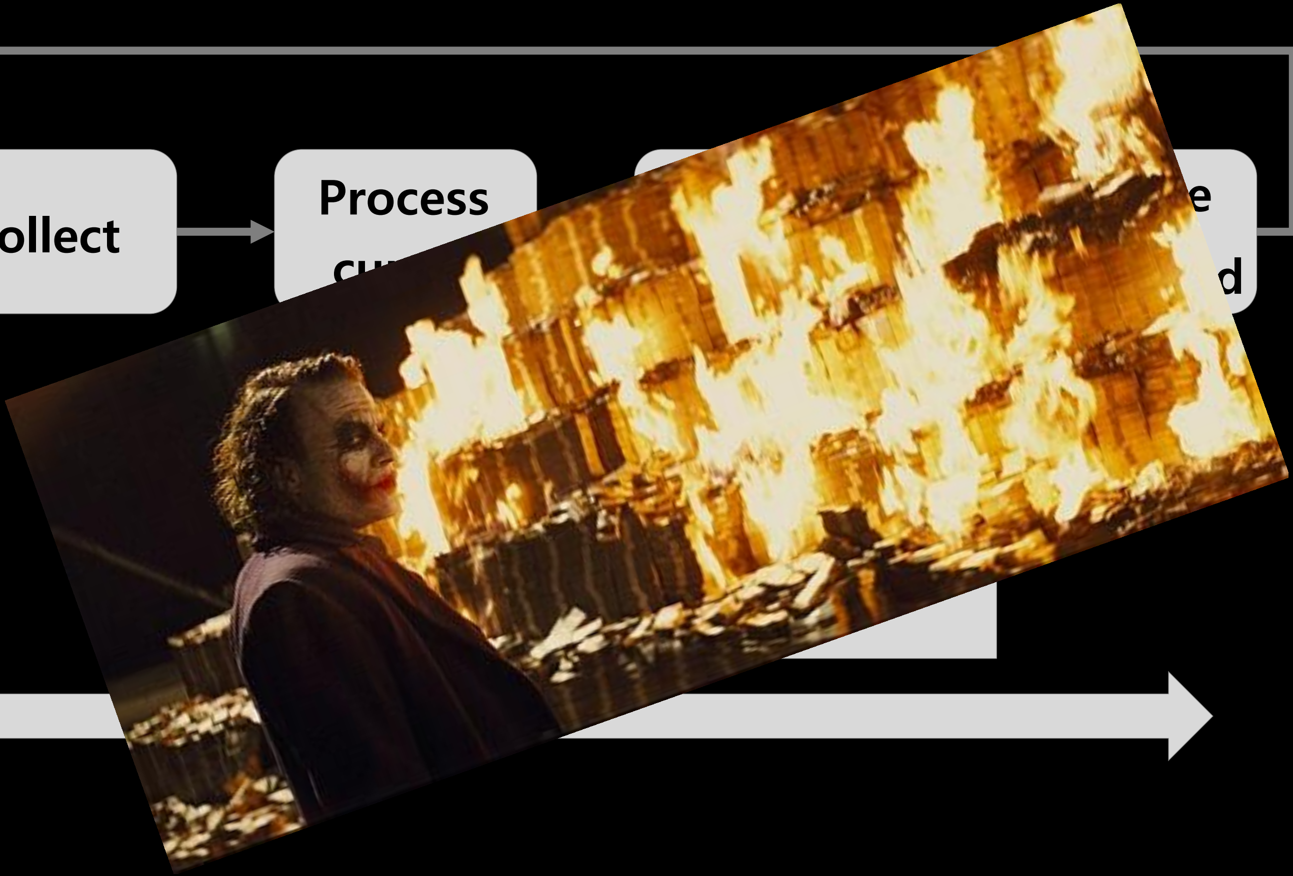
Training

Inference

Path to Production

딥러닝을 위한 데이터 수집

Collect → Process



데이터의 중요도

일반적인 검출 Labeling

데이터의 중요도

일반적인 검출 Labeling & 공들인 Labeling

데이터 수집은 또다른 AI의 적용 분야

- As Is

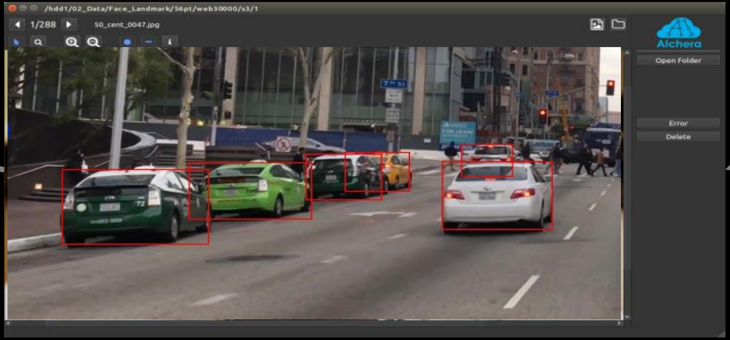
서비스 (Product)

Train/test data

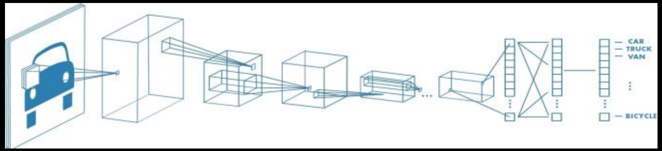
Datasets

Labels

Manually selected data



Deep Learning



데이터 수집은 또다른 AI의 적용 분야

- To Be

서비스 (Product)

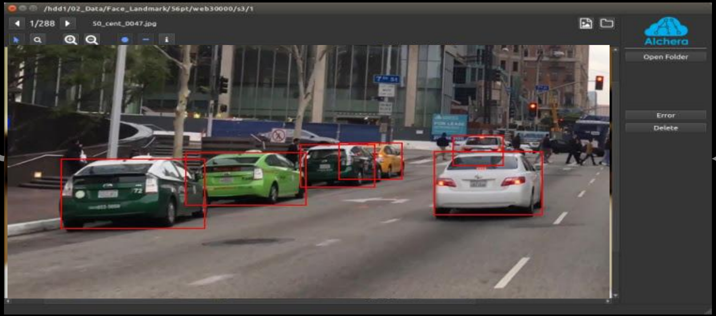
Train/test data

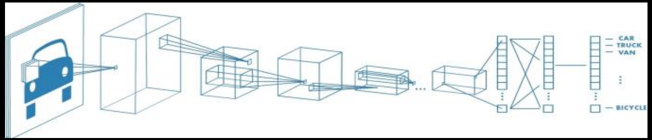
Datasets

Labels

Deep Learning

Intelligently & Cost effectively selected data





What about labeling?

- 학습에 large-scale 데이터셋이 필요
  - Labeling은 시간이 드는 작업

- 컴퓨터에서 만든 이미지는 realistic domain에서 not generalize
  - 이유는 도메인이 두 이미지의 도메인이 다르기 때문

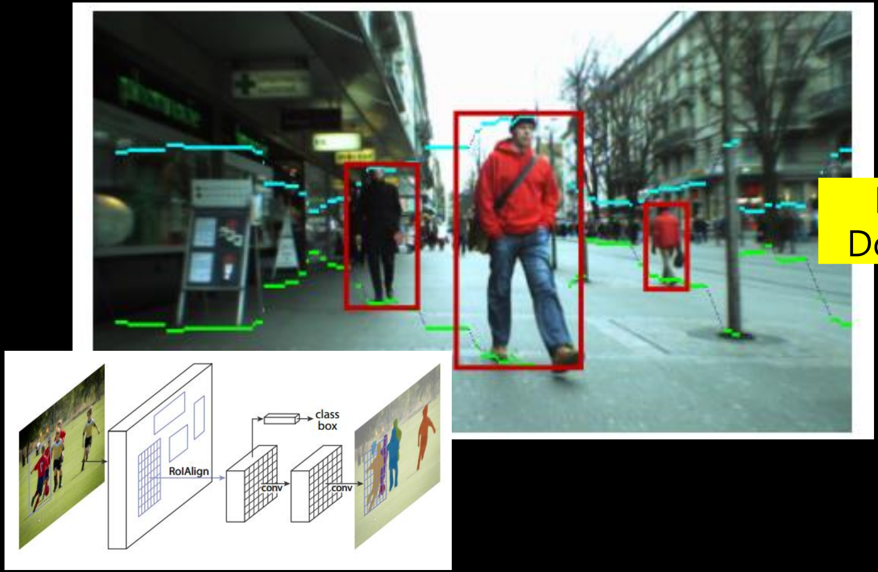
- Labeling된 도메인에서 학습된 네트워크가 labeling이 없어 지도학습을 하지 못한 다른 도메인에서도 적용할 수 있어야 한다.

  → **Domain Adaptation**

https://www.youtube.com/watch?v=w2pwxv8rFkU

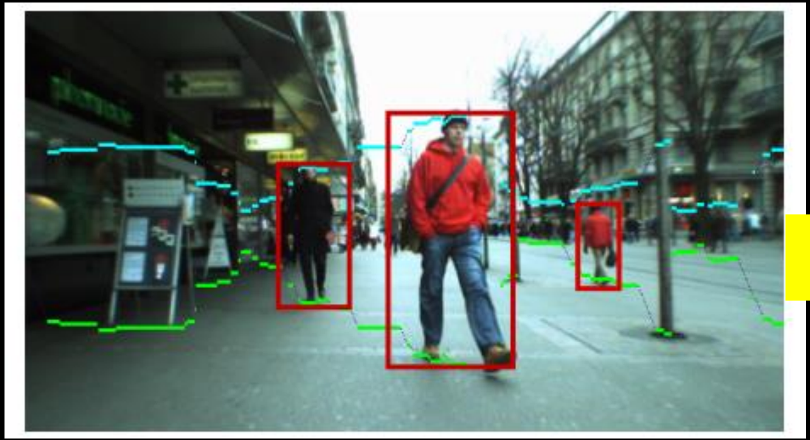
Major limitation of Deep Learning

No data efficiency!! : It needs millions of images



New Domain

Major limitation of Deep Learning



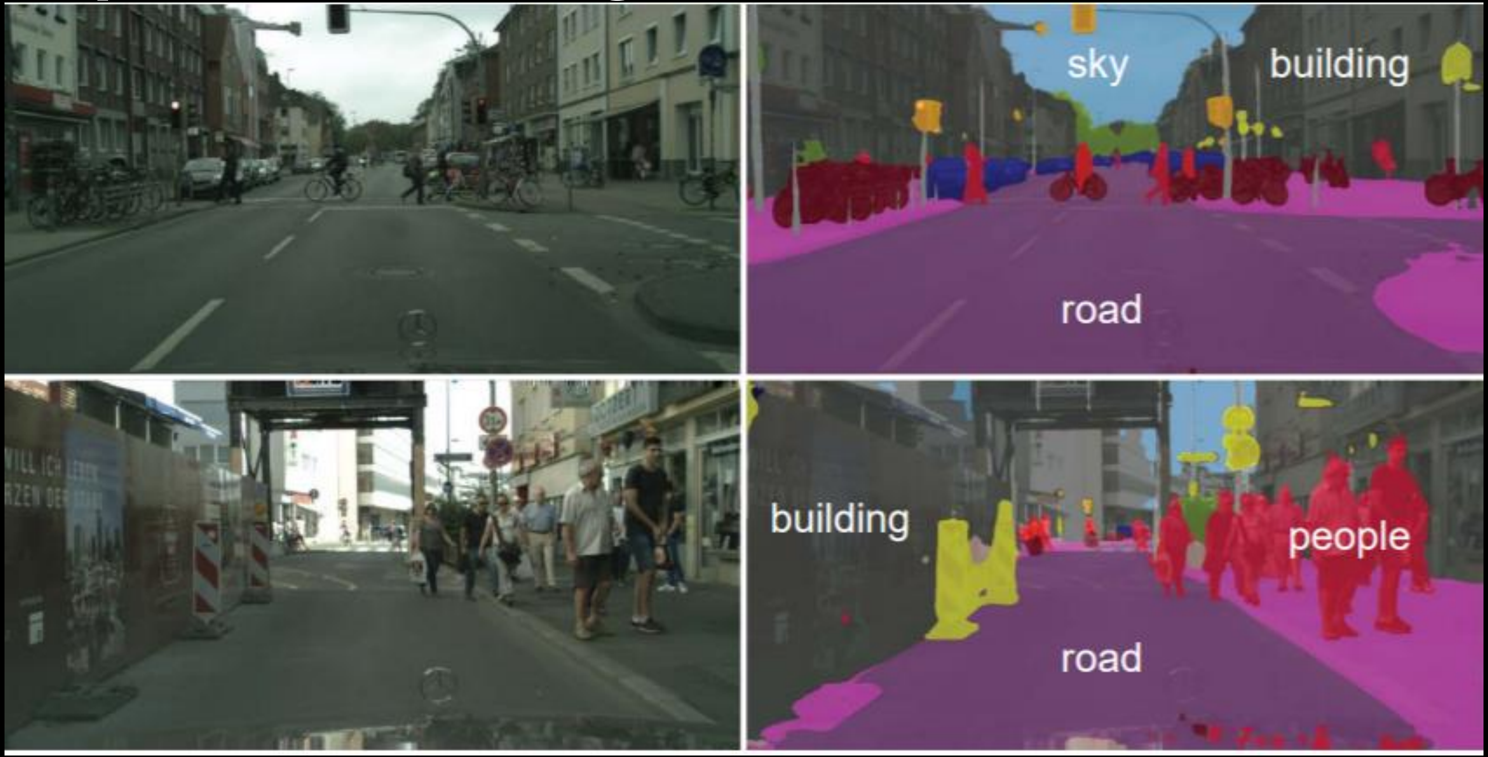
My net is trained on

New Domain

I am asked to acquire and label those data

"Dataset Bias"
"Domain Shift"
"Domain Adaptation"
"Domain Transfer"

Example: Scene Segmentation

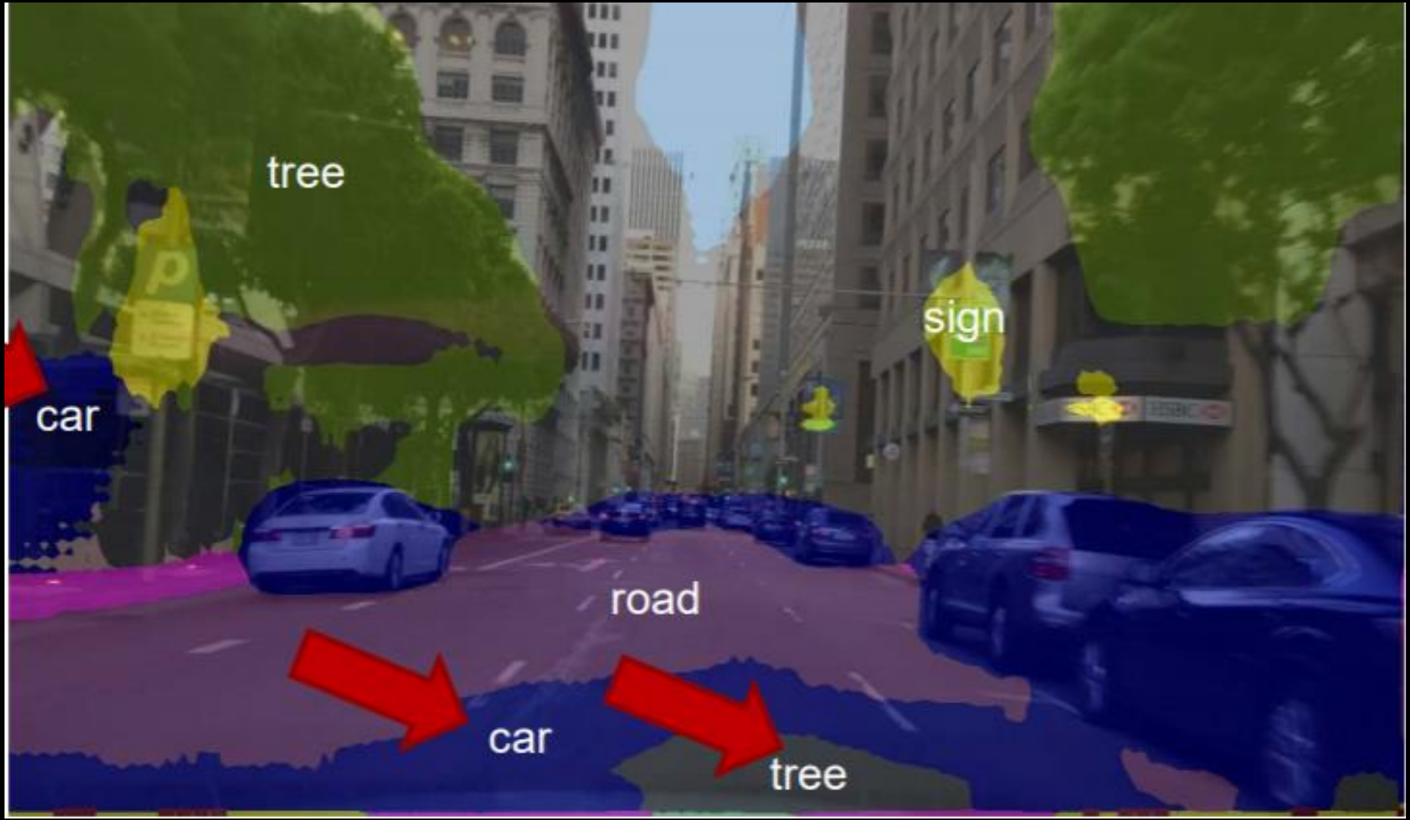


Trained on Cityscape Dataset

Tested on Cityscape Dataset

FCNs in the Wild: Pixel-level Adversarial and Constraint-based Adaptation, Judy Hoffman, Dequan Wang, Fisher Yu, Trevor Darrell, Arxiv 2016

Domain Shift : Cityscape to SF

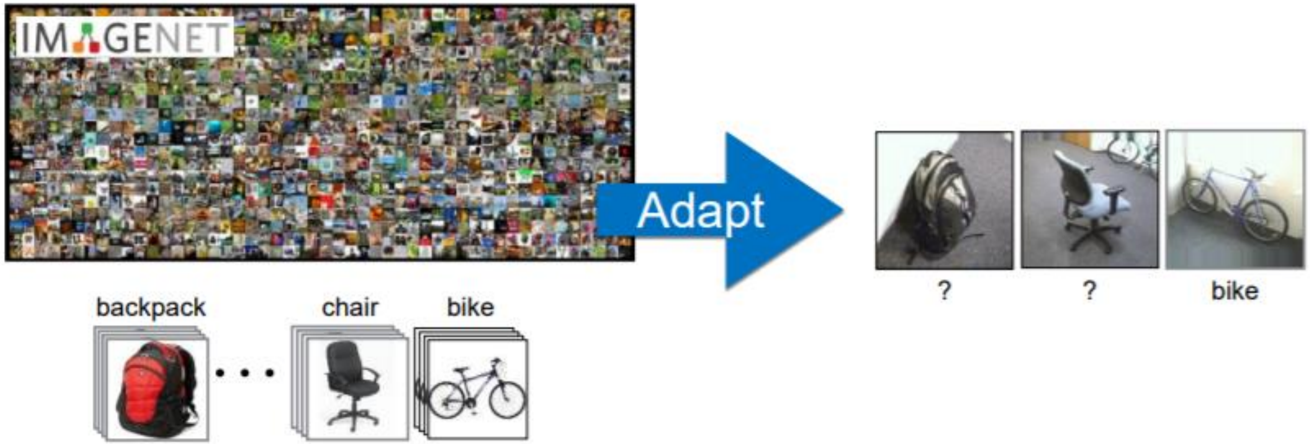


FCNs in the Wild: Pixel-level Adversarial and Constraint-based Adaptation, Judy Hoffman, Dequan Wang, Fisher Yu, Trevor Darrell, Arxiv 2016

No tunnel in the Cityscape

FCNs in the Wild: Pixel-level Adversarial and Constraint-based Adaptation, Judy Hoffman, Dequan Wang, Fisher Yu, Trevor Darrell, Arxiv 2016

Domain Adaptation



Source Domain $\sim P_S(X, Y)$ $\neq$ Target Domain $\sim P_T(Z, H)$

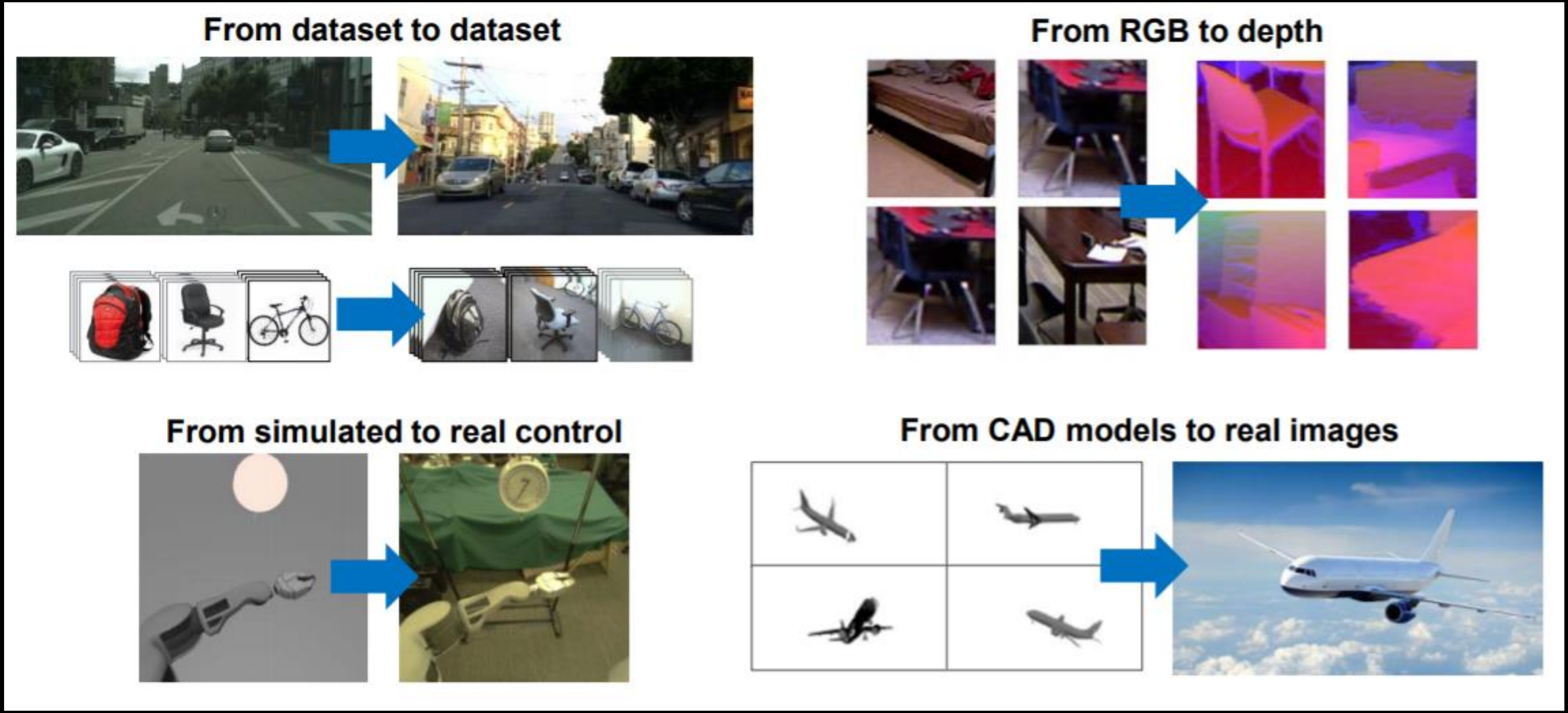
lots of **labeled** data

unlabeled or limited labels

$$D_S = \{(\mathbf{x}_i, y_i), \forall i \in \{1, \ldots, N\}\}$$

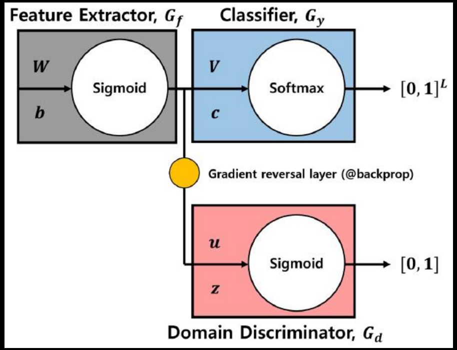
$$D_T = \{(\mathbf{z}_j, \mathbf{?}), \forall j \in \{1, \ldots, M\}\}$$

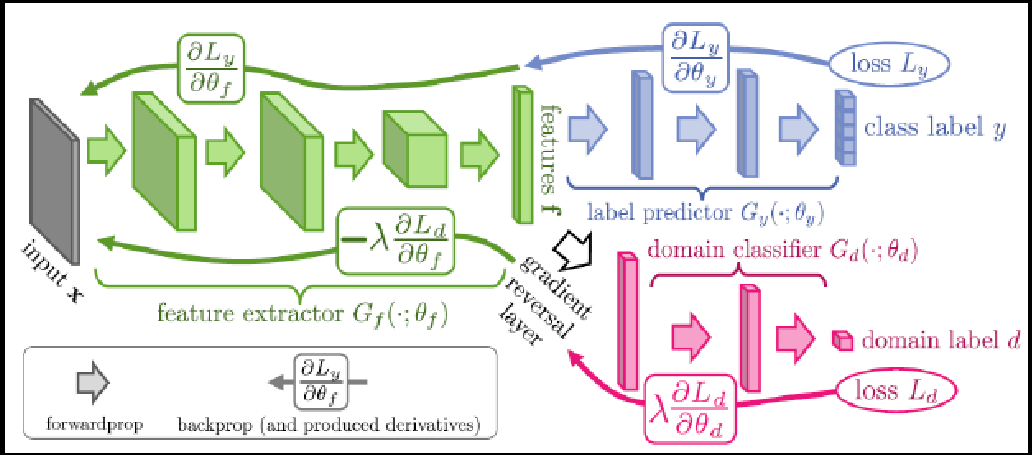
Domain Adaptation : Application



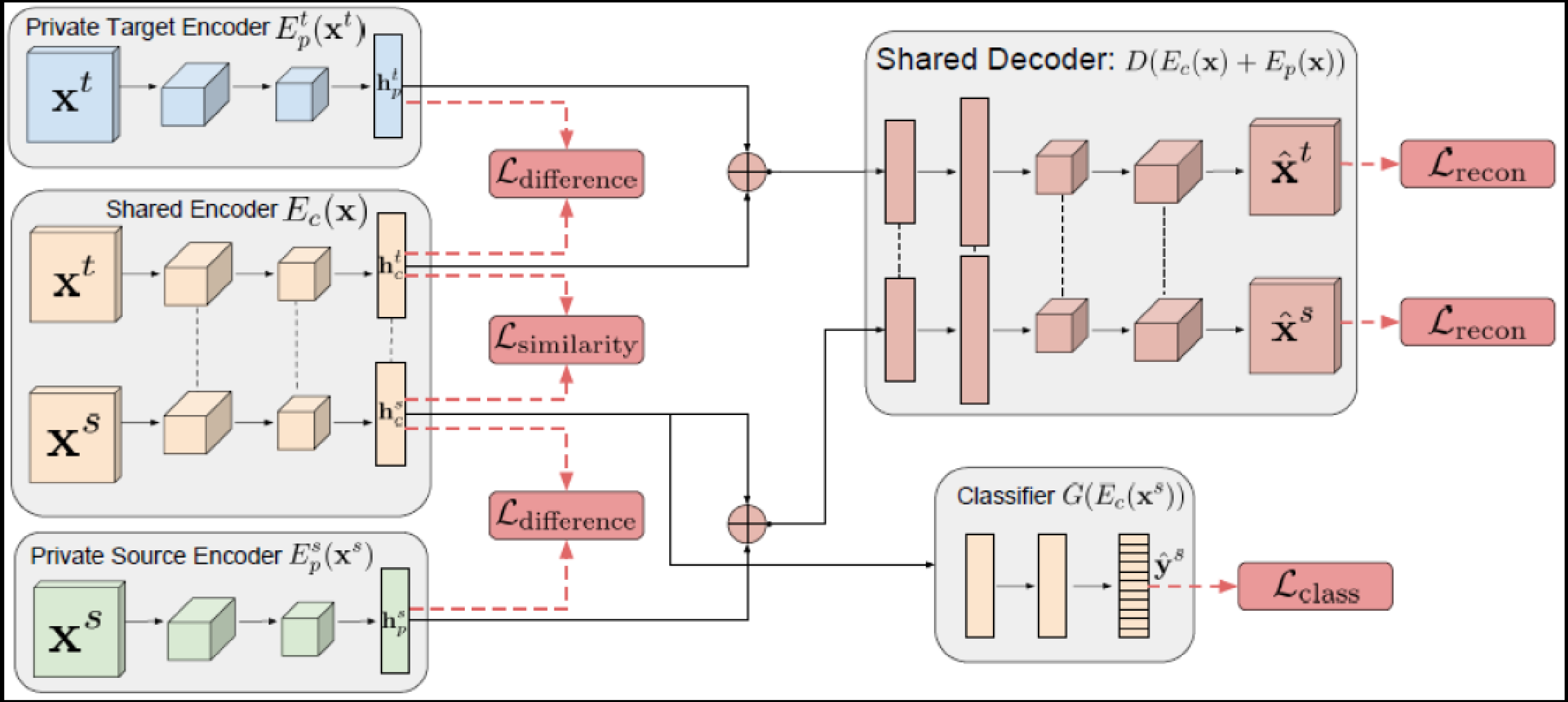
Domain Adversarial Neural Network

- Domain classifier
  - Source/Target 중 어느 도메인에서 만든 feature인지 구분
  - 학습 시, 두 도메인이 동일한 분포 내의 feature를 생성하도록 유도
- Label predictor
  - 학습된 feature를 통해 class 구분





Domain Separation Networks

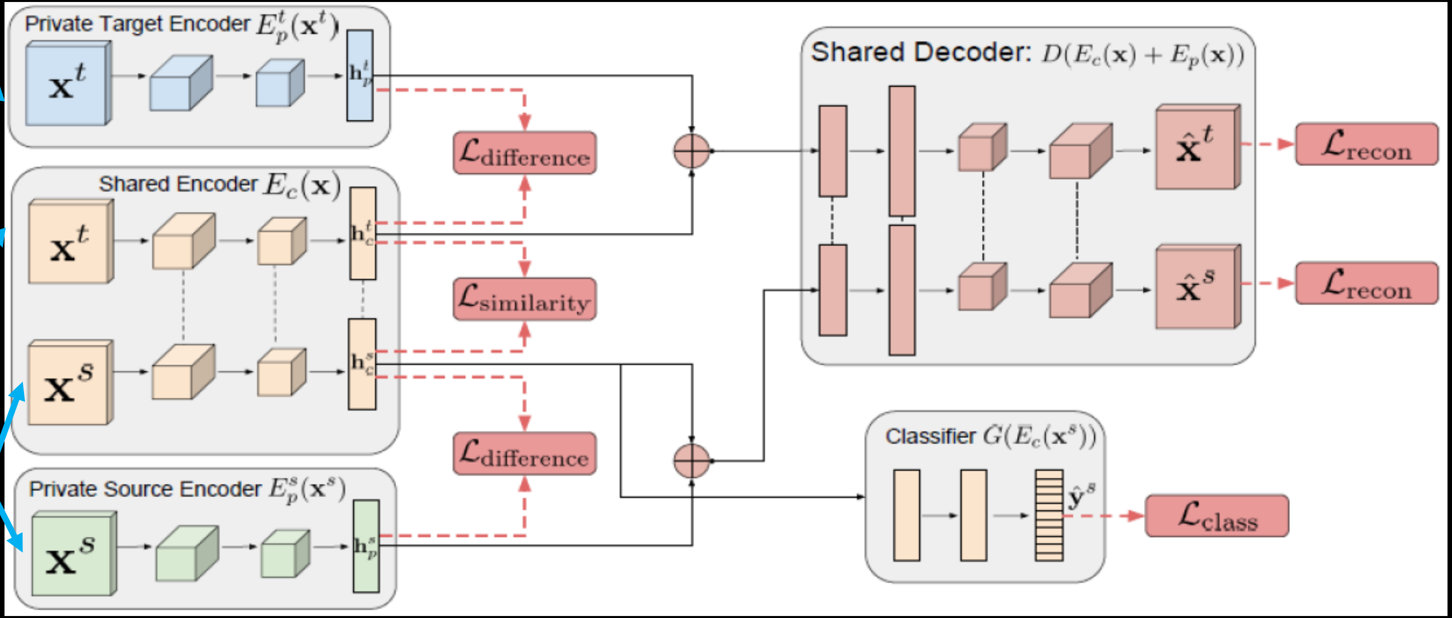


$$\mathcal{L} = \mathcal{L}_{\text{task}} + \alpha\,\mathcal{L}_{\text{recon}} + \beta\,\mathcal{L}_{\text{difference}} + \gamma\,\mathcal{L}_{\text{similarity}}$$

Domain Separation Networks

- Private encoder → unique feature to each domain = low-level
- Shared encoder → domain invariant feature = high-level

Target domain (w/o label)



Source domain (w/ label)

Losses

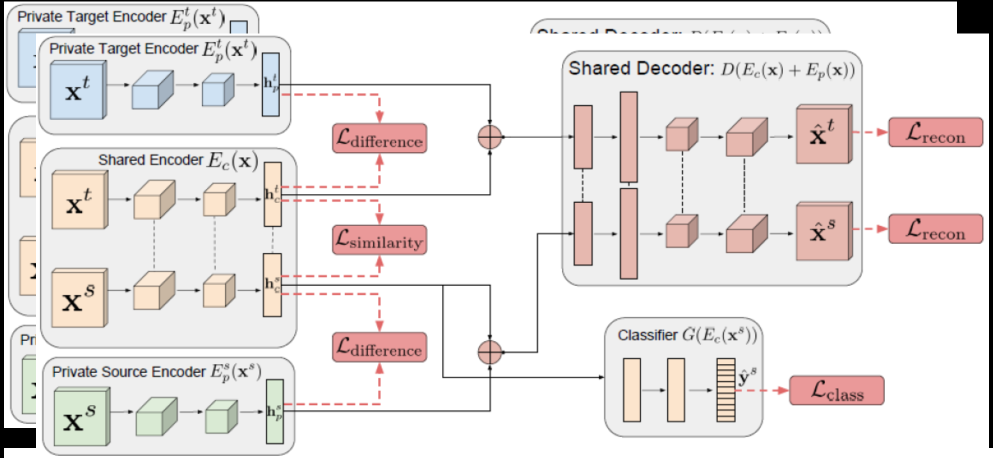
$$\mathcal{L}_{\text{task}} = -\sum_{i=0}^{N_s} \mathbf{y}_i^s \cdot \log \hat{\mathbf{y}}_i^s$$

Task loss : Cross entropy loss

$$\mathcal{L}_{\text{recon}} = \sum_{i=1}^{N_s} \mathcal{L}_{\text{si\_mse}}(\mathbf{x}_i^s, \hat{\mathbf{x}}_i^s) + \sum_{i=1}^{N_t} \mathcal{L}_{\text{si\_mse}}(\mathbf{x}_i^t, \hat{\mathbf{x}}_i^t)$$

Reconstruction loss : scale-invariant mean square error

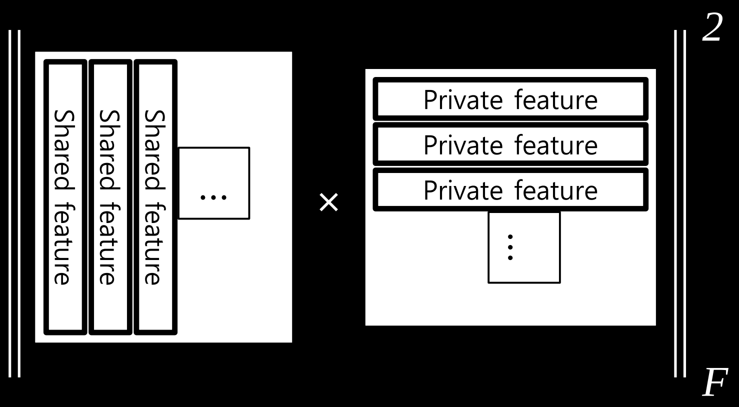
$$\mathcal{L}_{\text{si\_mse}}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{k}\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 - \frac{1}{k^2}([\mathbf{x} - \hat{\mathbf{x}}] \cdot \mathbf{1}_k)^2$$



$$\mathcal{L} = \mathcal{L}_{\text{task}} + \alpha\,\mathcal{L}_{\text{recon}} + \beta\,\mathcal{L}_{\text{difference}} + \gamma\,\mathcal{L}_{\text{similarity}}$$

Losses

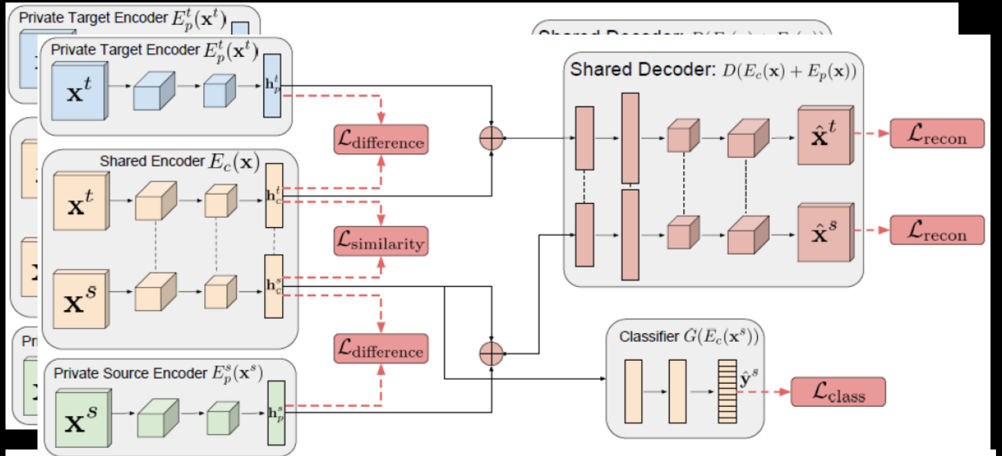
$$\mathcal{L}_{\text{difference}} = \left\| {\mathbf{H}_c^s}^\top \mathbf{H}_p^s \right\|_F^2 + \left\| {\mathbf{H}_c^t}^\top \mathbf{H}_p^t \right\|_F^2$$



Difference loss : Frobenius norm

$$\|A\|_F \equiv \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2}$$

- **Shared / private feature가 orthogonal하도록 유도**
  → Uncorrelate, Independence



$$\mathcal{L} = \mathcal{L}_{\text{task}} + \alpha\, \mathcal{L}_{\text{recon}} + \beta\, \mathcal{L}_{\text{difference}} + \gamma\, \mathcal{L}_{\text{similarity}}$$

Result on Cityscape to SF adaptation

Before Domain Adaptation

After Domain Adaptation

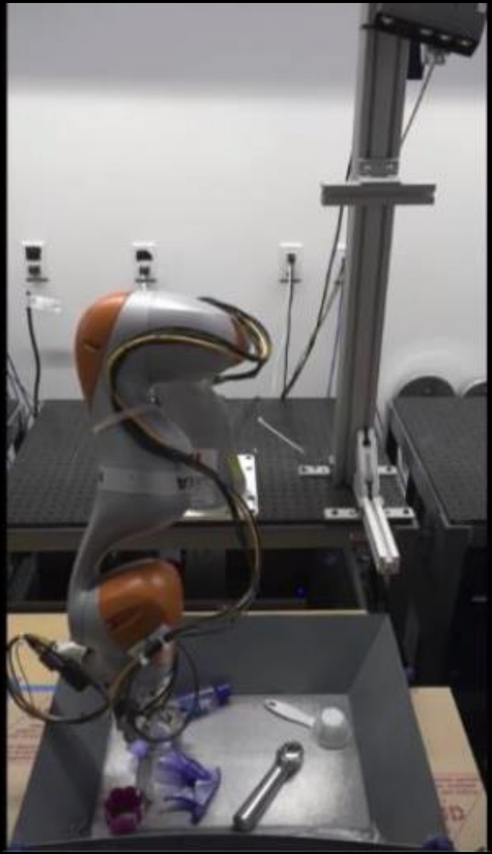
Another DA example

Using Simulation and Domain Adaptation to Improve Efficiency of Deep Robotic Grasping

Konstantinos Bousmalis[*,1], Alex Irpan[*,1], Paul Wohlhart[*,2], Yunfei Bai[2], Matthew Kelcey[1], Mrinal Kalakrishnan[2], Laura Downs[1], Julian Ibarz[1], Peter Pastor[2], Kurt Konolige[2], Sergey Levine[1], Vincent Vanhoucke[1]

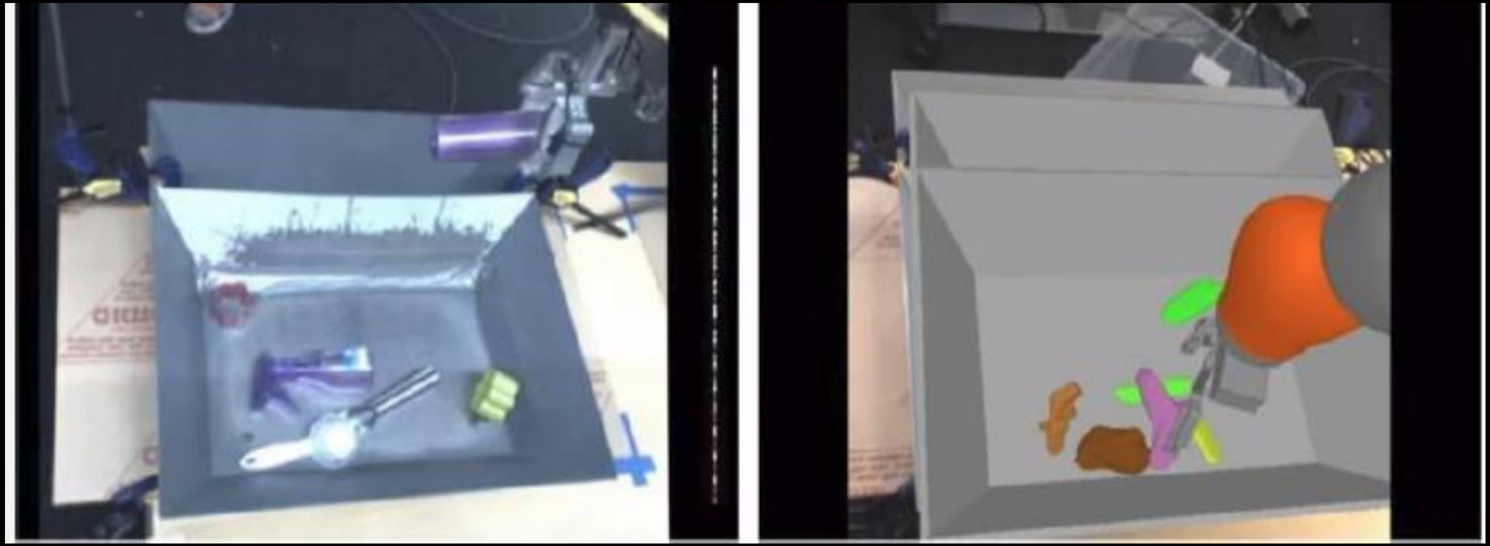
Robot Grasping

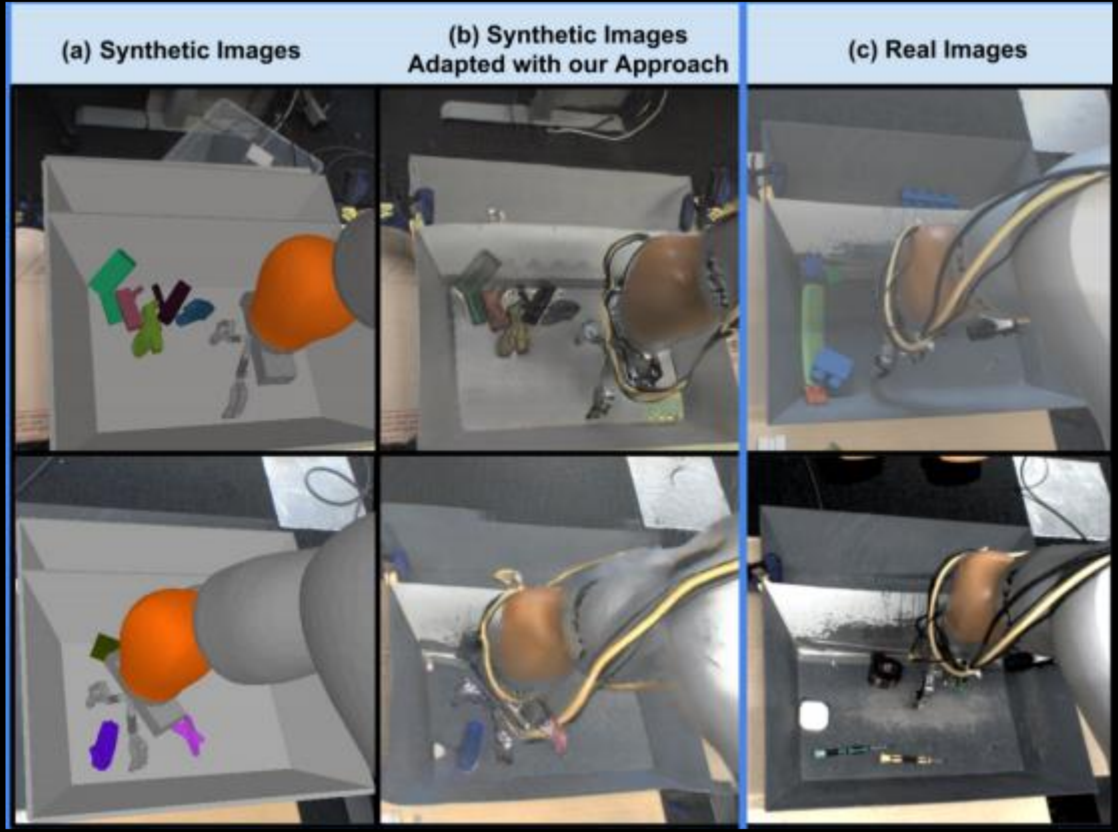
- Problem setup : Single-viewpoint RGB camera로 로봇이 특정 물체를 grasping 하는 것

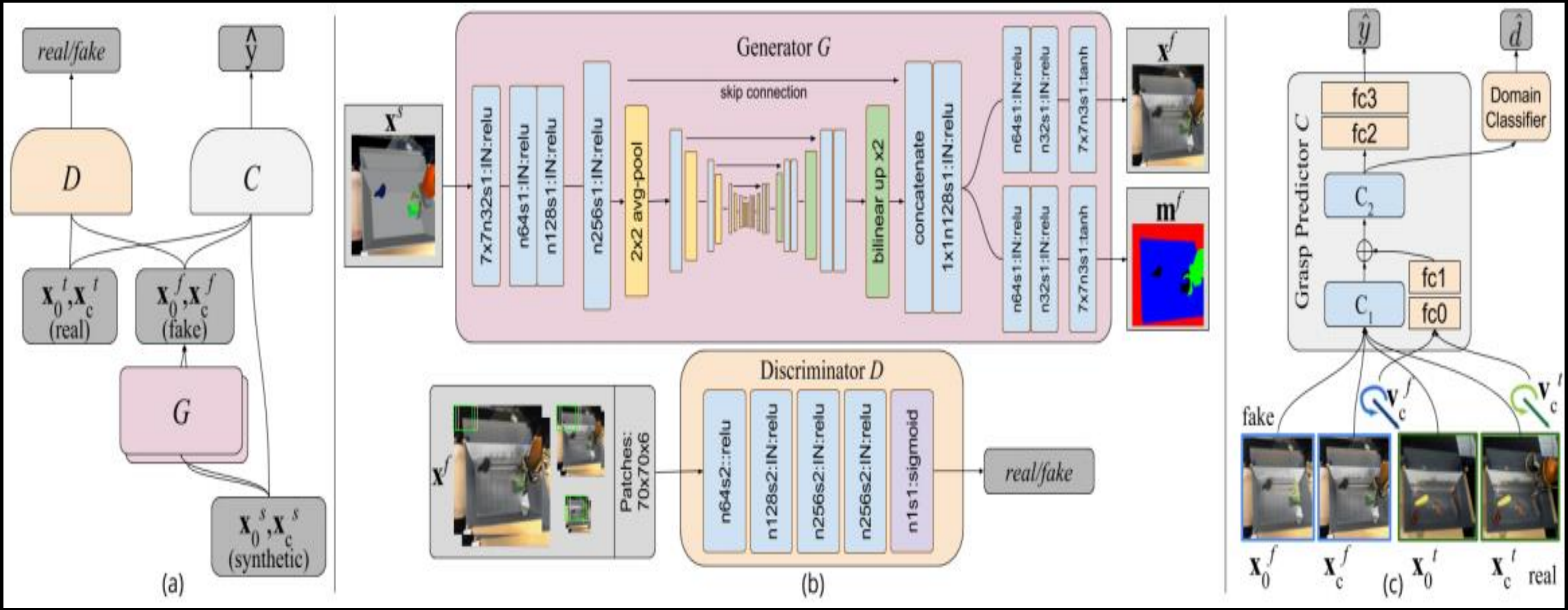


Using Simulation

- Millions of grasps in hours instead of weeks




(a) Synthetic Images
(b) Synthetic Images Adapted with our Approach
(c) Real Images



real/fake
ŷ
D
C
G
x_0^t, x_c^t (real)
x_0^f, x_c^f (fake)
x_0^s, x_c^s (synthetic)
(a)
Generator G
skip connection
x^s
7x7n32s1:IN:relu
n64s1:IN:relu
n128s1:IN:relu
n256s1:IN:relu
2x2 avg-pool
bilinear up x2
concatenate
1x1n128s1:IN:relu
n64s1:IN:relu
n32s1:IN:relu
7x7n3s1:tanh
x^f
m^f
Discriminator D
Patches: 70x70x6
n64s2::relu
n128s2:IN:relu
n256s2:IN:relu
n256s2:IN:relu
n1s1:sigmoid
real/fake
(b)
ŷ
d̂
Grasp Predictor C
fc3
fc2
C_2
C_1
fc1
fc0
Domain Classifier
fake
real
v_c^f
v_c^t
x_0^f
x_c^f
x_0^t
x_c^t
(c)

Summary

- Domain-Adaptation은 딥러닝의 Dataset Efficiency를 증가시키는 것으로 딥러닝의 중요한 연구분야

- Domain adaptation은 유사 site가 다수 존재하고 모든 Site에 Labeling된 Dataset을 확보하기 어려운 영역(특히 Smart Factory 부분)에 적용 가능

- 그 외에 다양한 입력 modality에 대하여도 적용 가능