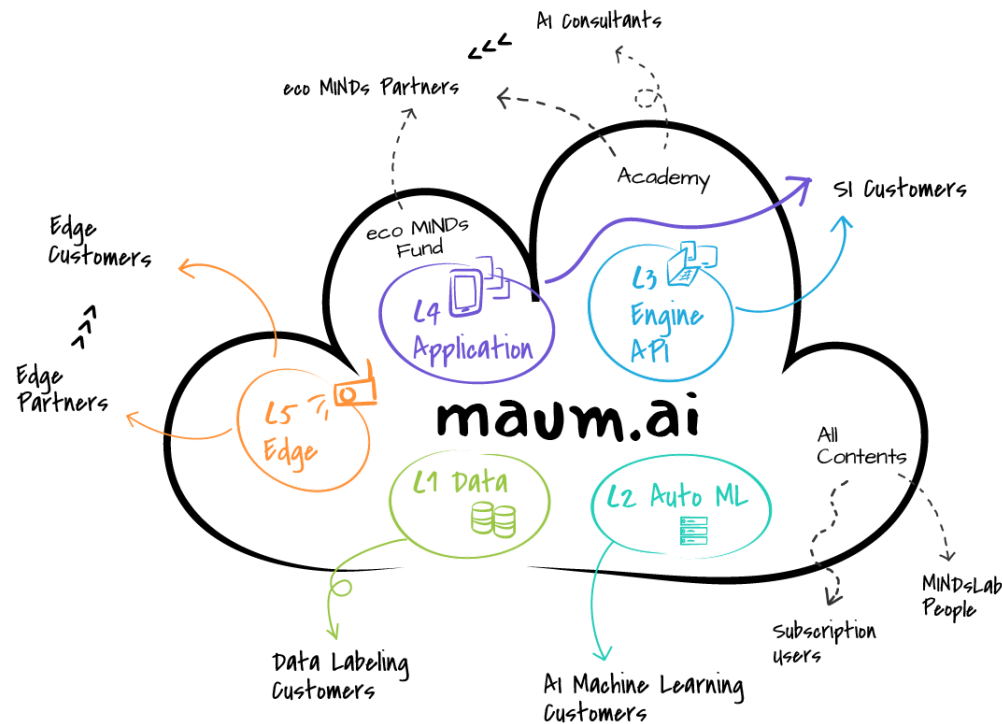


컨설턴트 교육

20201007

팀 소개

Brain-Vision



Brain-Audio

Brain-Vision

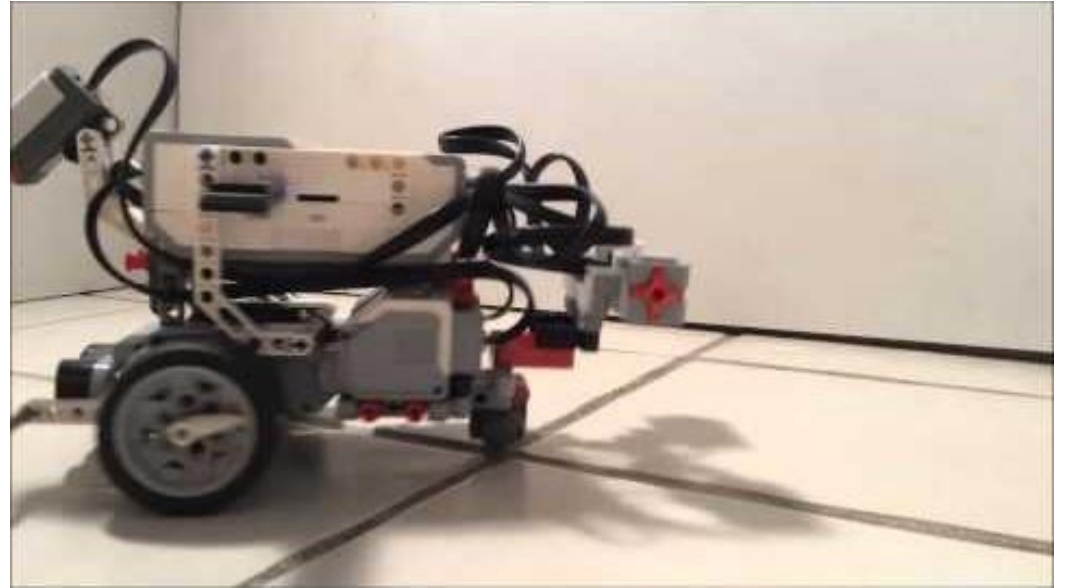
Brain-NLP

AI Consultant에게 필요한
딥러닝에 대한 인식 10분 바로잡기

NN은 Neural Network, 뇌가 아닌 신경망

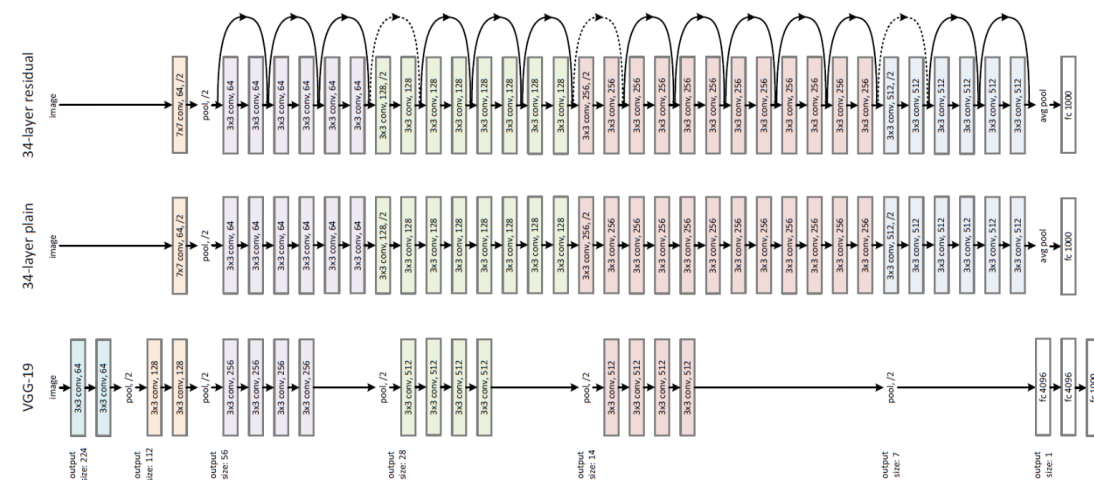
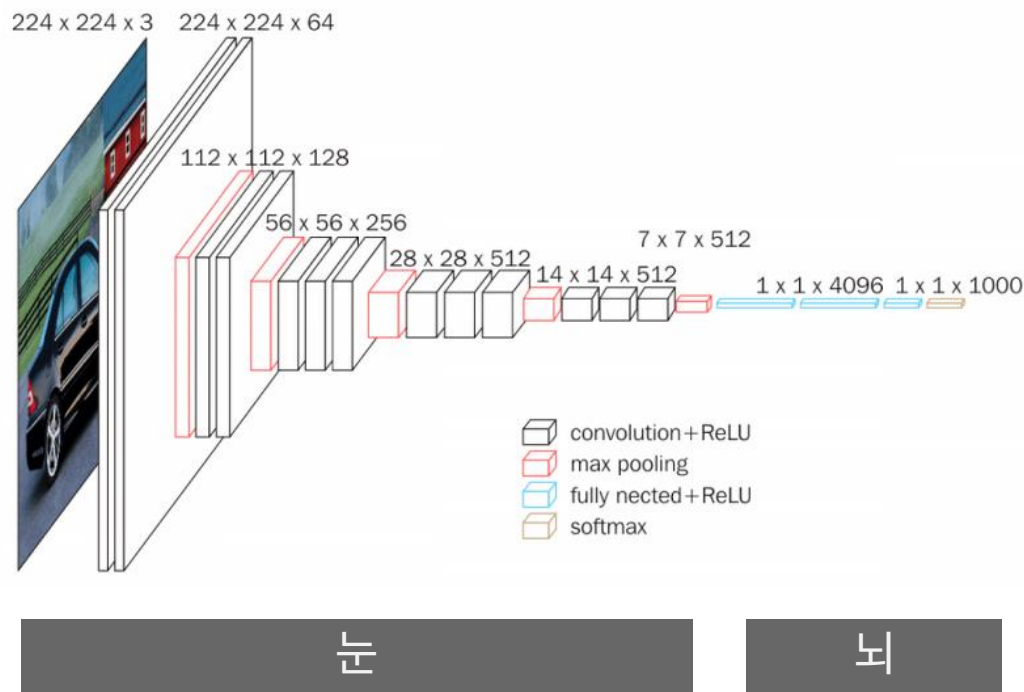
신경망으로도

- 벽이 있을 때 피할 수 있고
- 사물을 분류할 수 있고
- 큰 틀에서 인식할 수 있다.



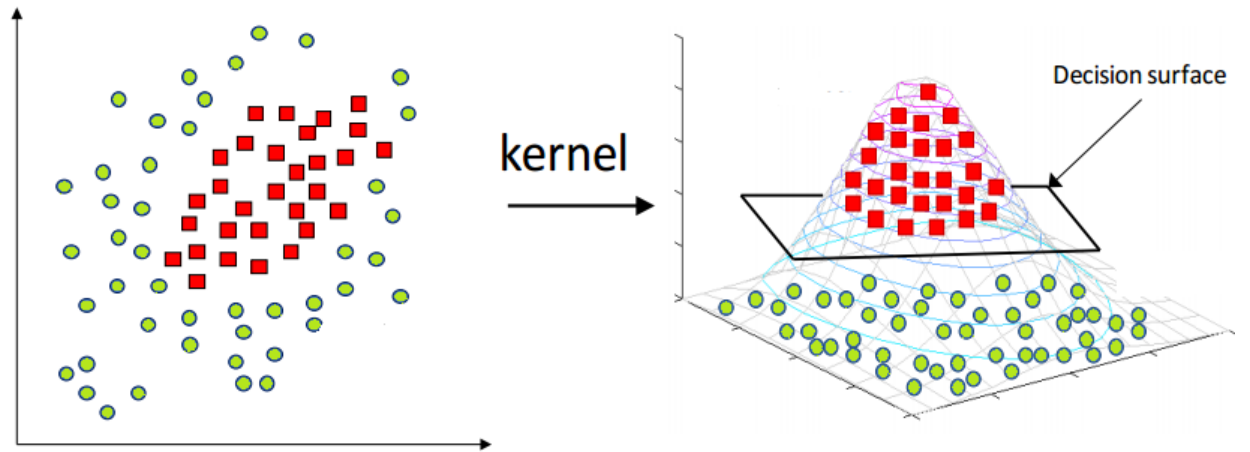
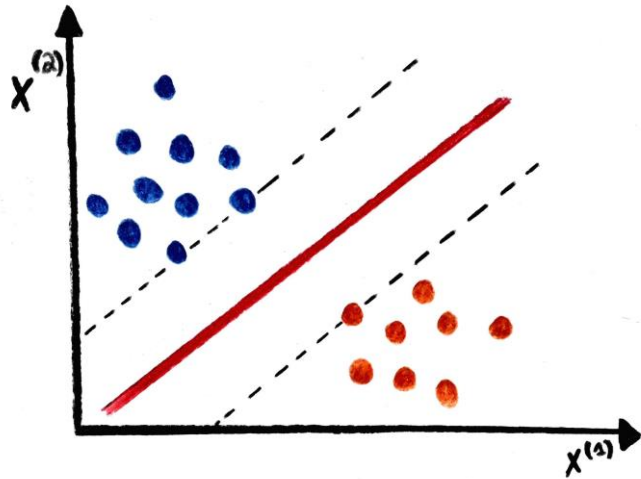
예쁜꼬마선충(*C.elegans*) 신경망을 모사하여 만든 로봇

Image Classification (이미지 분류) 모델 생각해보기

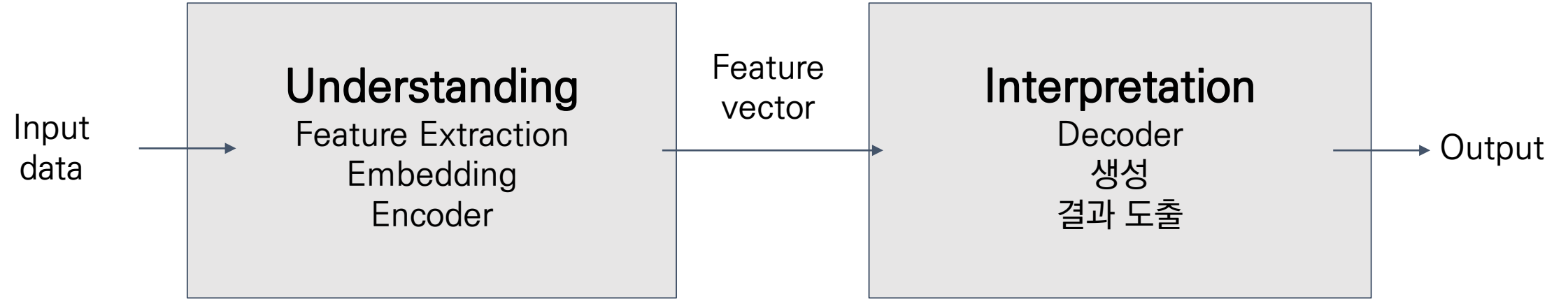


ResNet도 눈에 해당하지,
뇌에 해당하는 걸 발전시킨 것은 아니다.

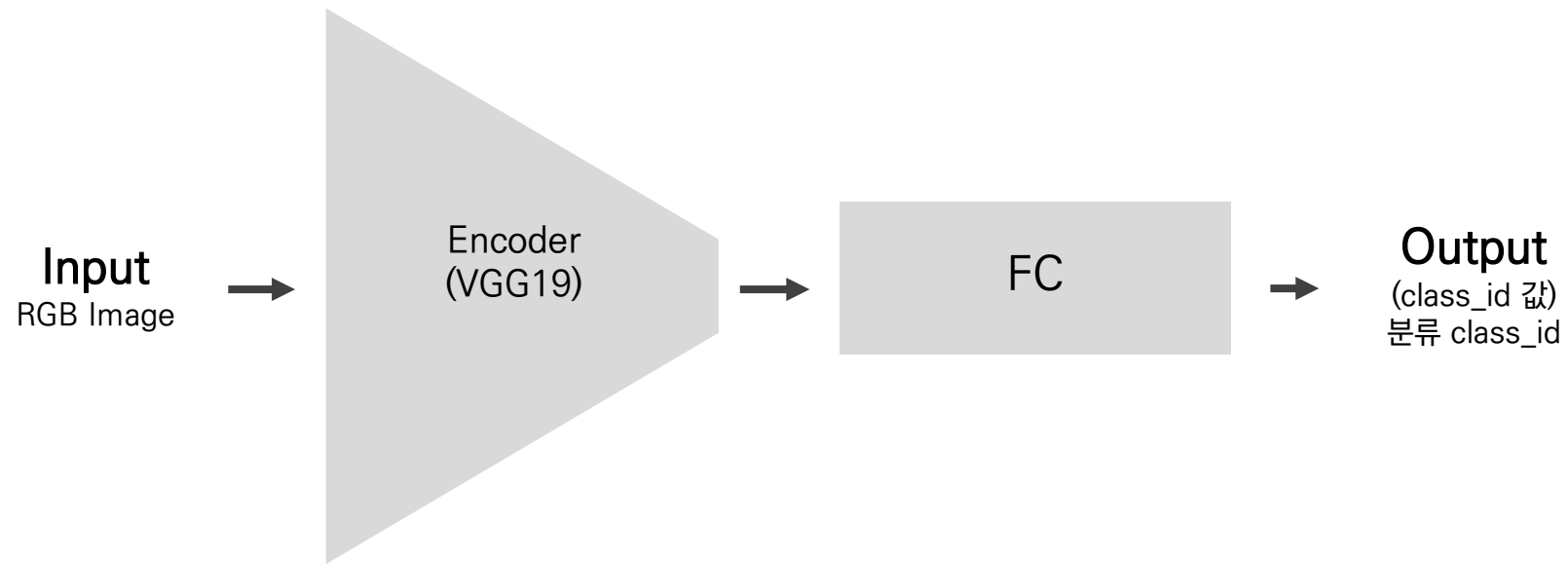
고차원으로 보내버리기 (by 비선형)



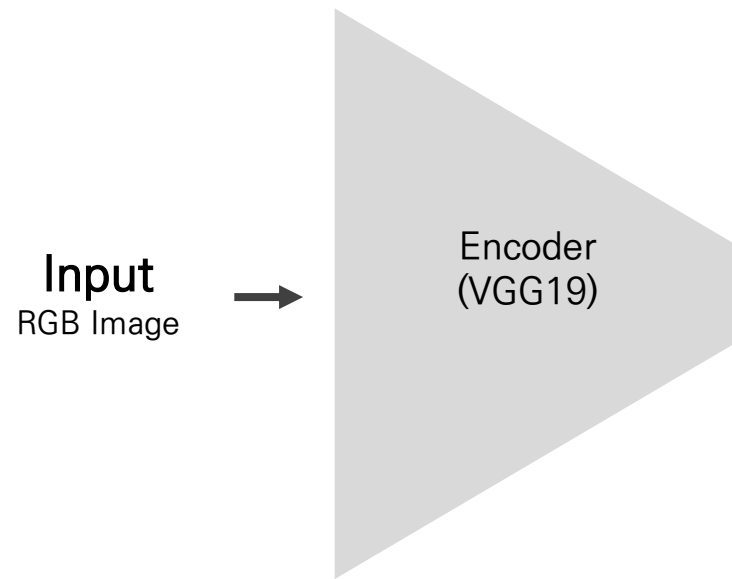
Vision 엔진들을 이해하기 가장 쉬운 방법



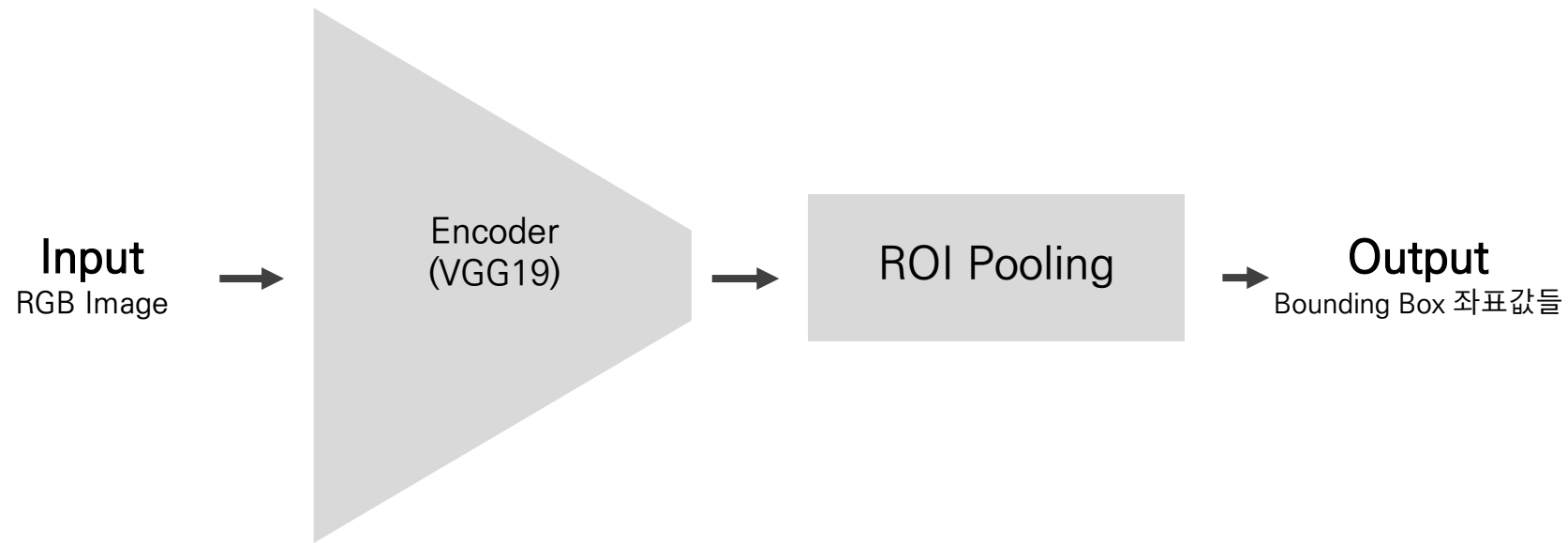
Architecture Image Classification



Architecture

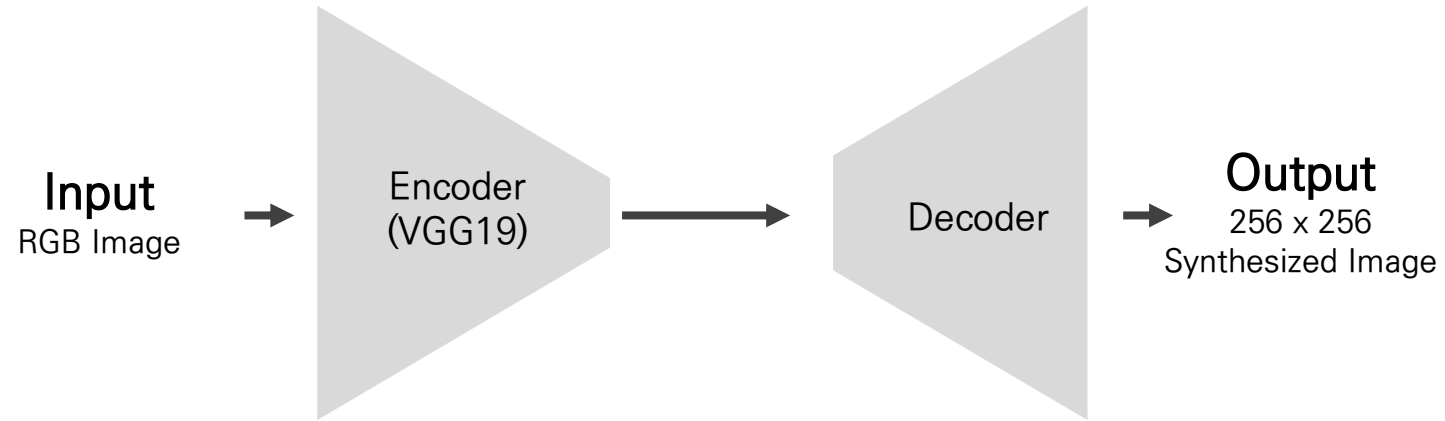


Architecture Object Detection



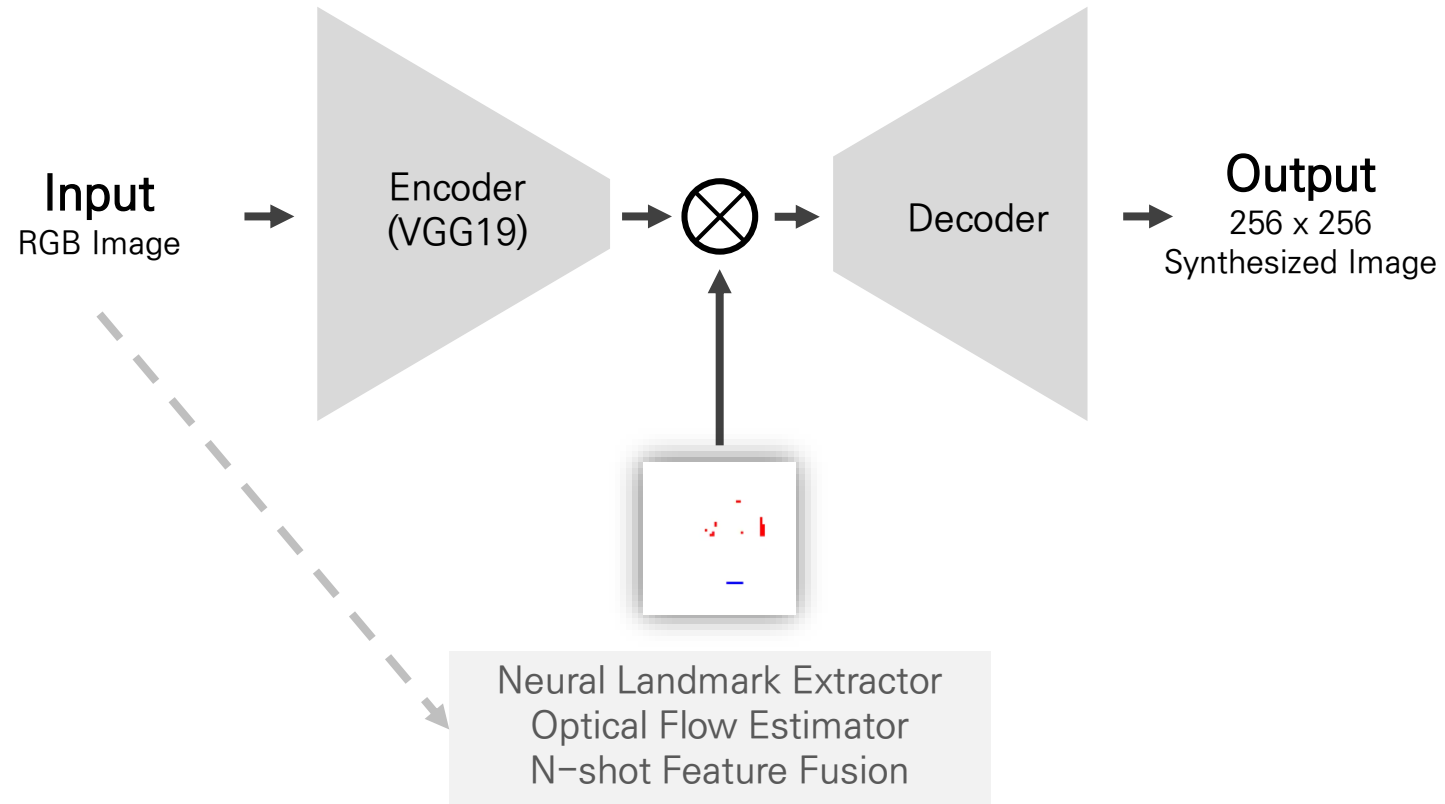
Architecture

Image Generator

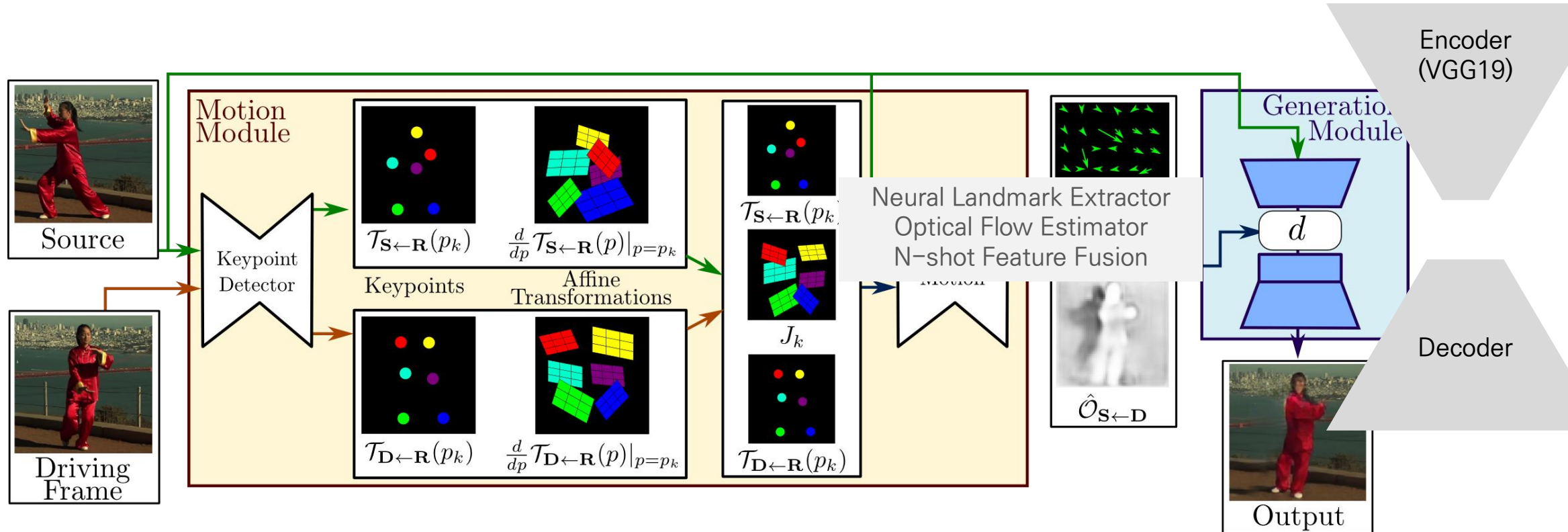


Architecture

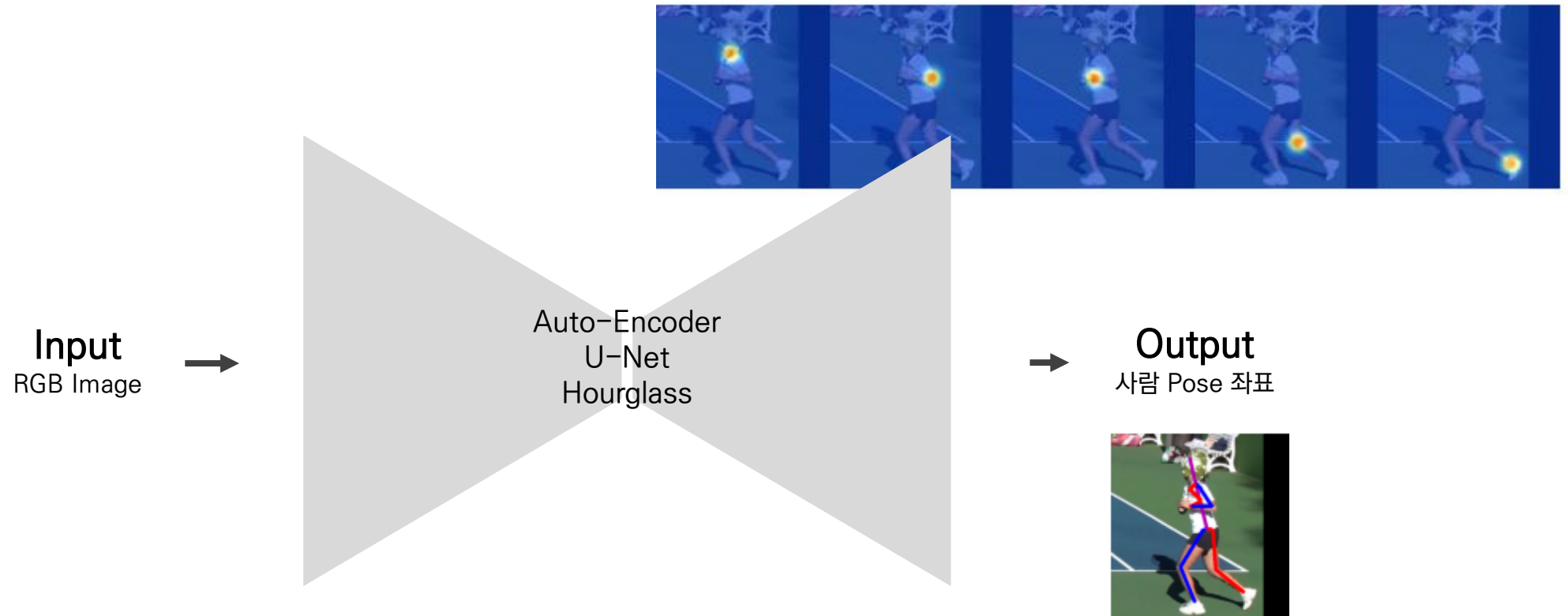
Image Generator



Baseline 논문의 Figure



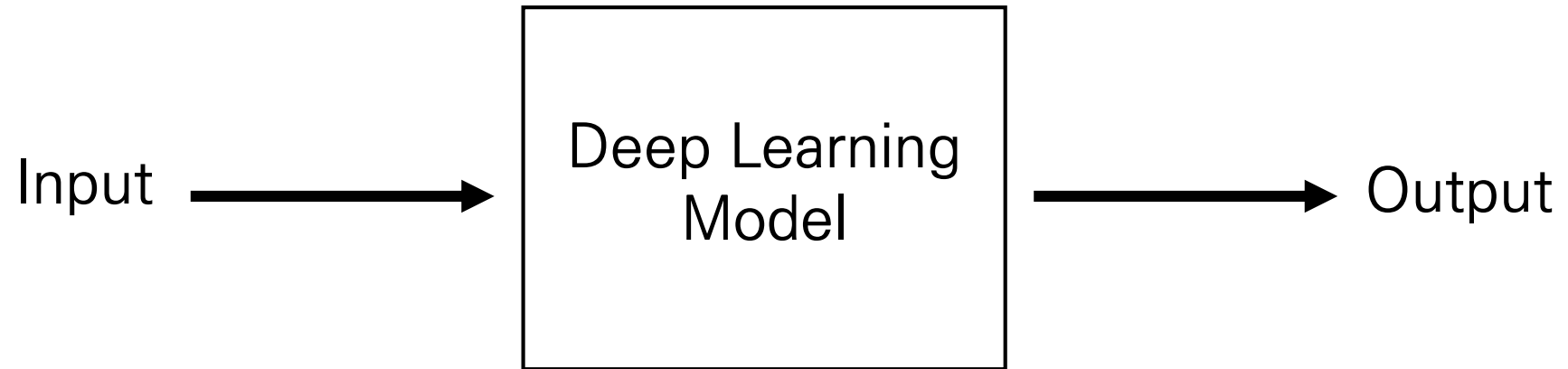
Architecture Pose Estimation



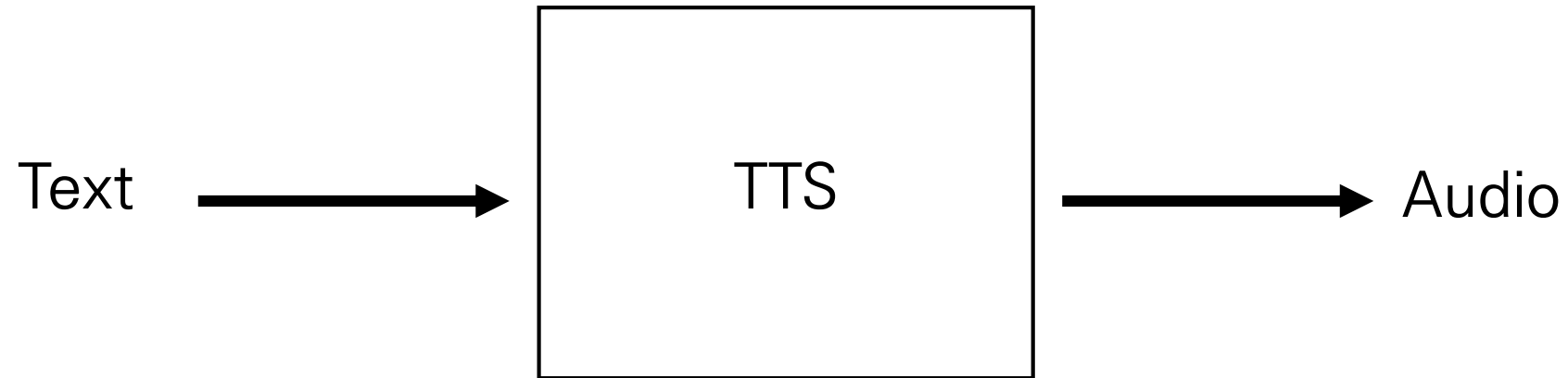
나비 모양이라고 다 Generation 하는 Task는 아니다.
Image Segmentation, Pose Estimation은
Decoder에 해당하는 구조가 필요하다.

AI 컨설턴트가 데이터를 보면
모델이 보인다!

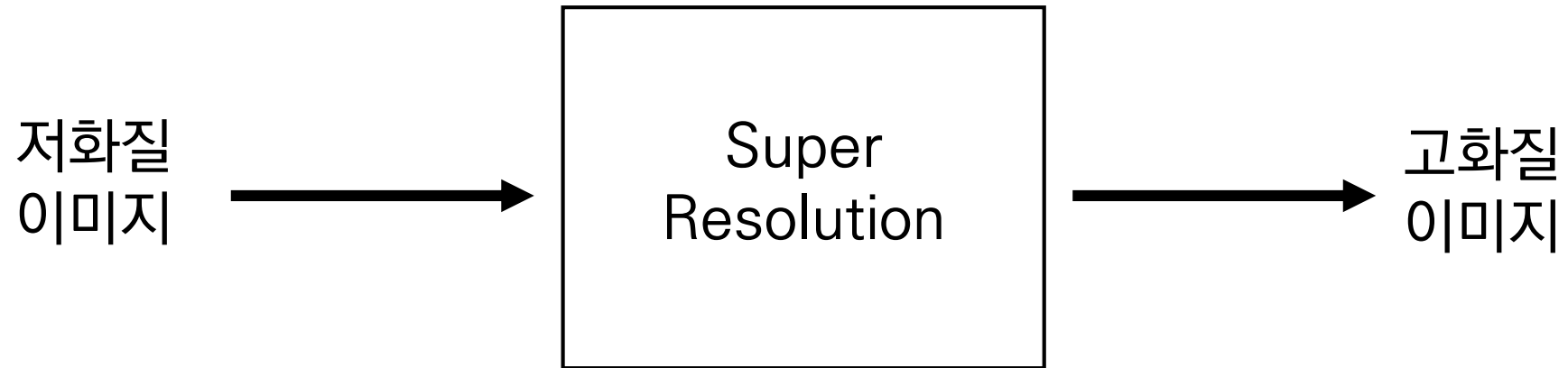
High Level



High Level



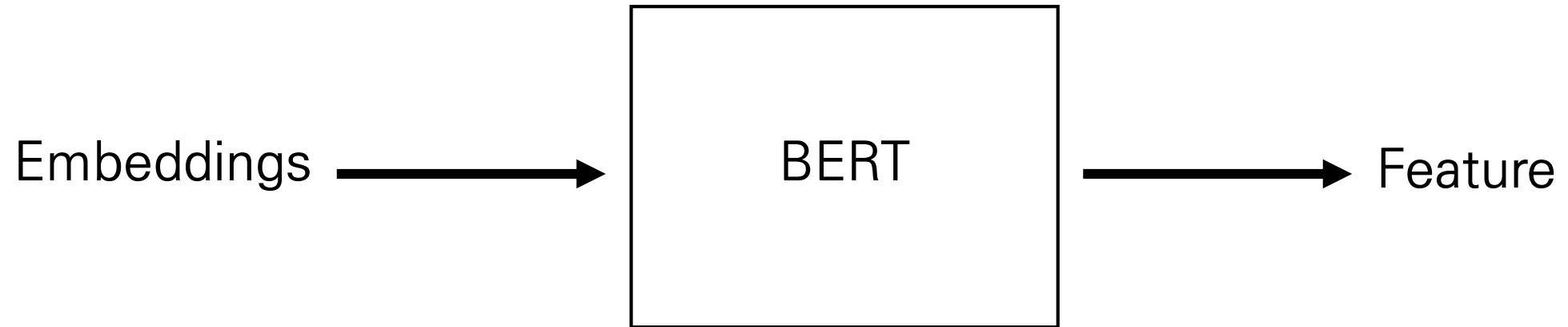
High Level



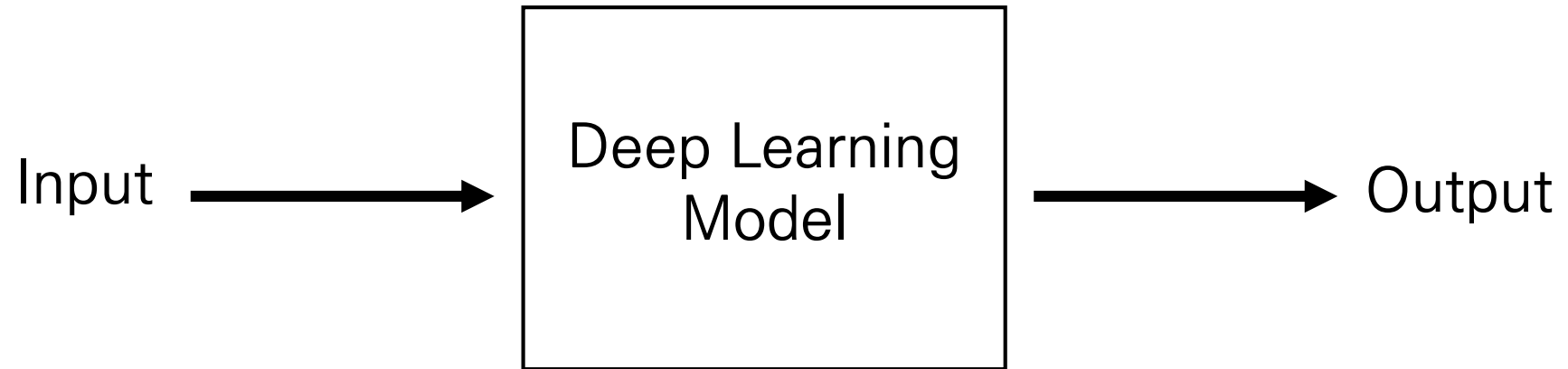
High Level



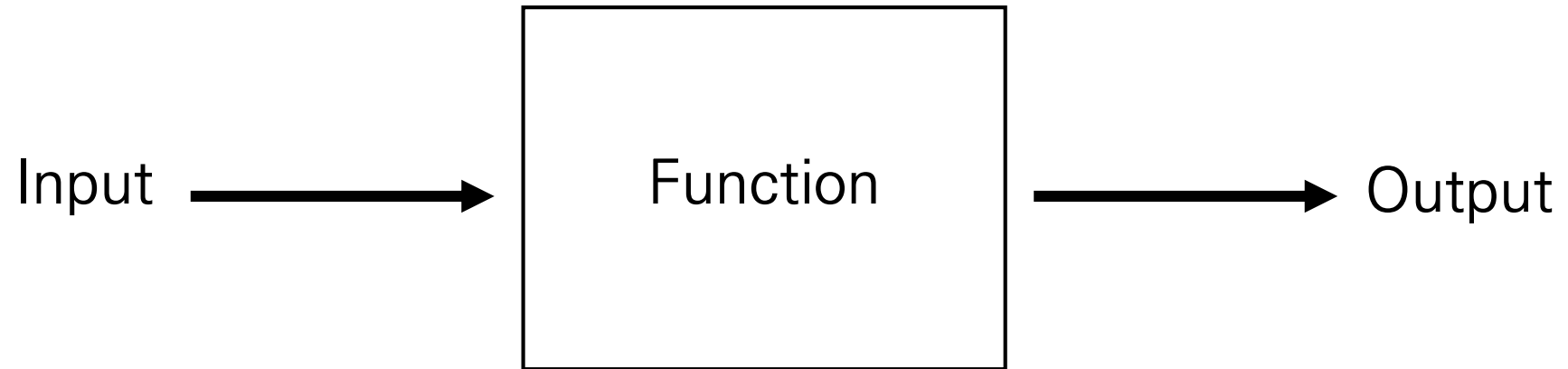
High Level



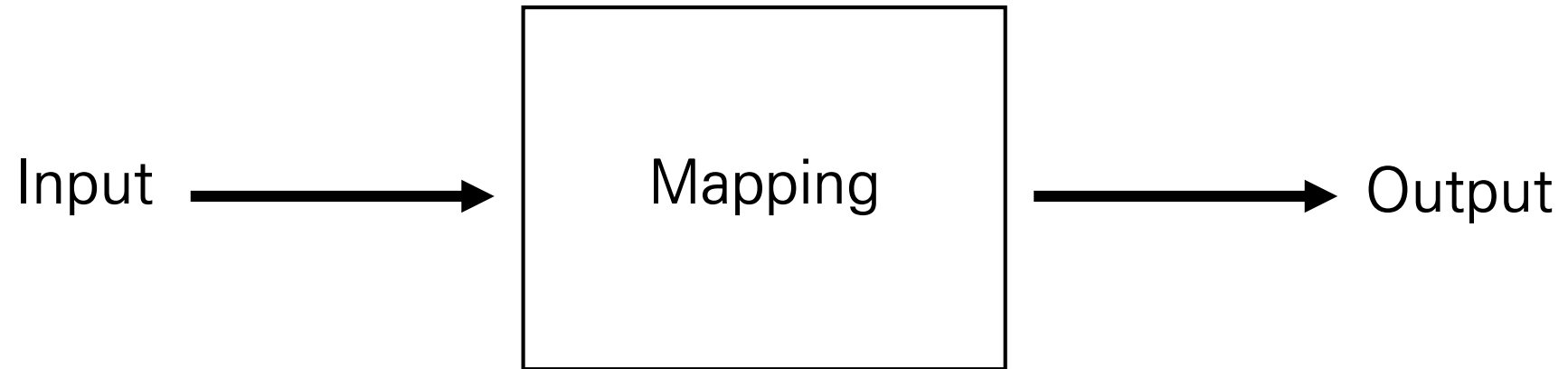
High Level



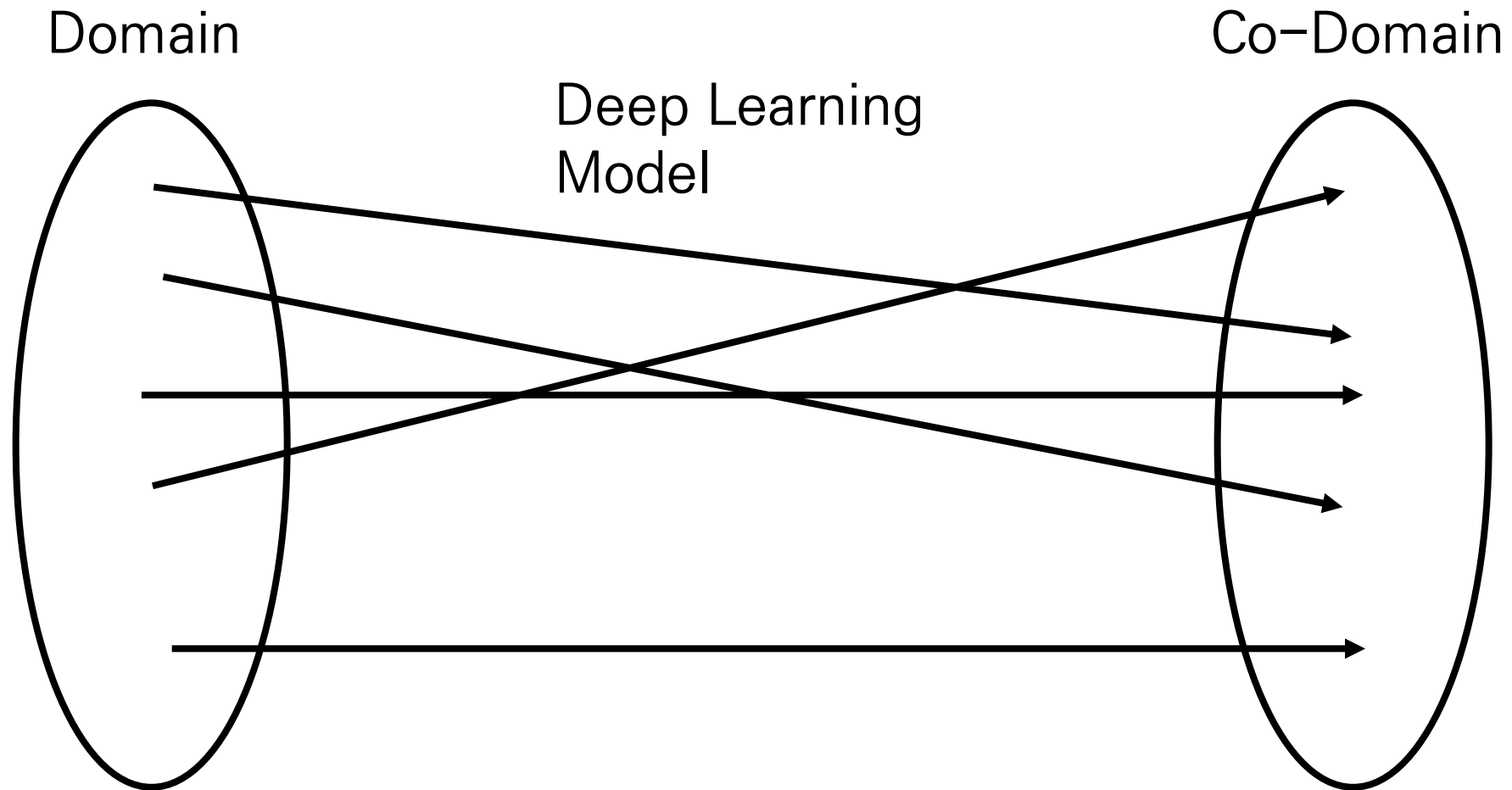
High Level



High Level

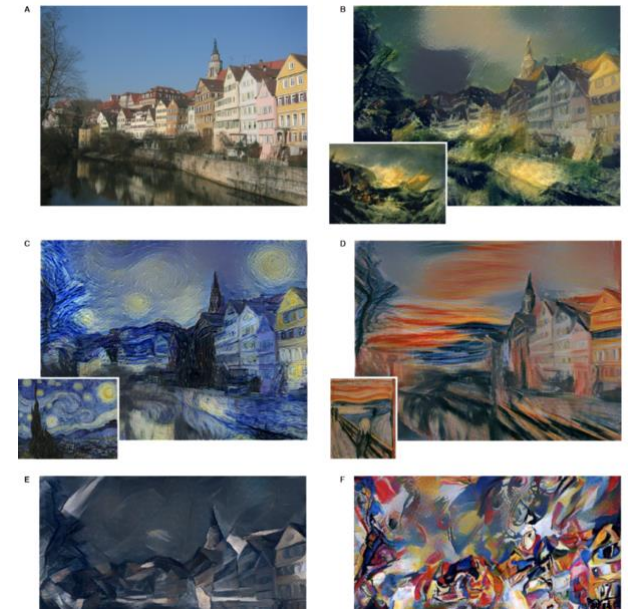
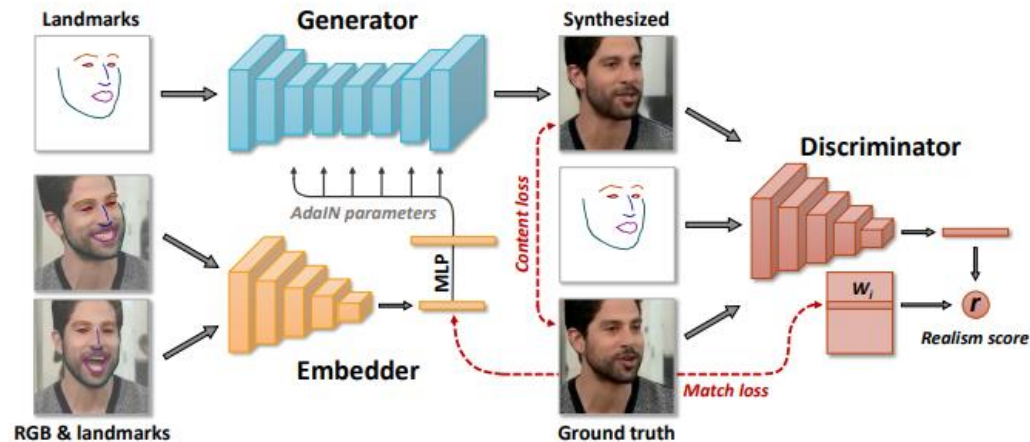


DL as a Function



Task 데이터에 대한 1차적인 파악

- Many-to-One
 - Classification, Object Detection 등 Image Recognition Task
- One-to-Many
 - Style Transfer 등 Image Generation

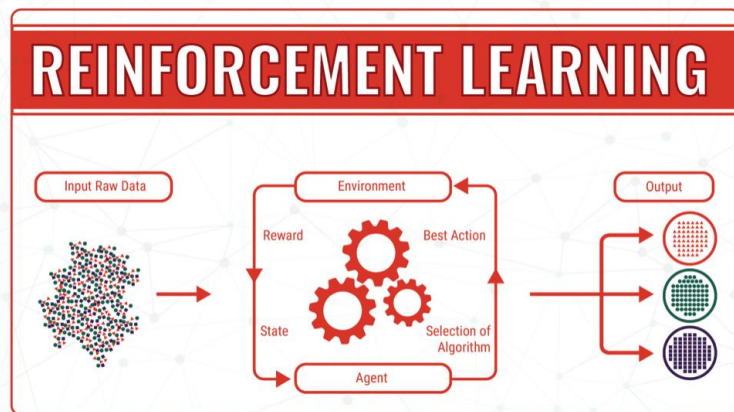
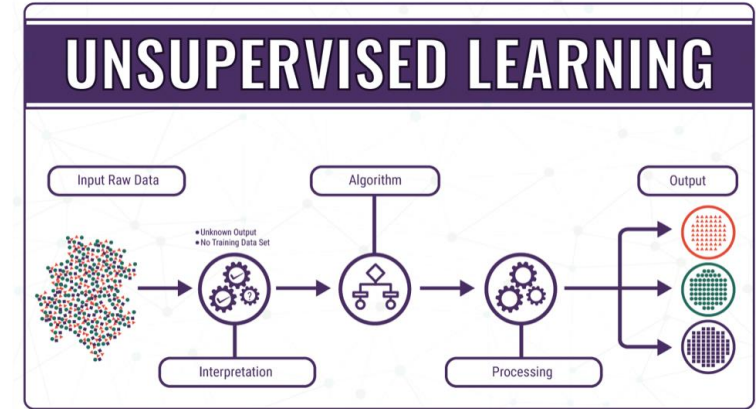
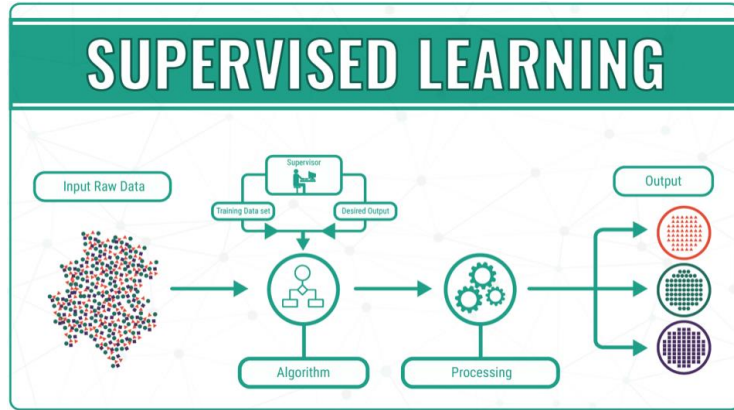


고객사들이 가지고 있는 데이터의 유형과
이에 따른 방향 Guide 제시

고객사 데이터의 3가지 형태

- Paired Data
 - Input 데이터 샘플과 이로부터 나와야 하는 데이터 샘플이 서로 쌍을 잡고 있다.
 - Ex. (OCR을 위한 이미지, 이미지에서 Text 위치와 Text 라벨)
- Unpaired Data
 - Input 데이터들끼리 모여서 엄청 많고, 나와야 하는 형태의 데이터들끼리 엄청 많다.
 - Ex. 사람 얼굴 이미지 잔뜩 + 애니메이션 얼굴 이미지 잔뜩
- Only Input Data
 - Input 데이터들끼리 모여서 엄청 많다.
 - Ex. 걸어 놓고 찍은 옷 사진 1,000장 (하고 싶은 건 아바타에 옷 입히기)

절대 해서는 안되는 말 : 강화학습으로 할 수 있어요



Unpaired Data라면 강화학습? No.

1. Supervised로 유사한 도메인에 학습된 모델들이 있는지 파악한 후 전이 학습 진행
2. 데이터에 라벨링을 하도록 하여 1.로 진행
3. (결과 담보 불가) Unsupervised 방식으로 진행
4. (현실적으로 불가) 강화학습

전 데이터가 많습니다. Vs 전 데이터가 없습니다.

- 우선 진짜 많은 것인지 확인해봐야 함.
 - 배경만 바뀌가면서 contents는 그대로인 채로 만든 데이터
 - 라벨링이 엉망진창으로 되어 있는 데이터 (라벨링을 다시 할 수 있는 rule도 없다.)
- 데이터가 많은 기준
 - 10^4 (=10,000) 개의 샘플 이상
- 데이터가 적다의 기준
 - 10^3 (=1,000) 개의 샘플 이상
 - 성능을 장담하기 어려움

각 상황 별 대처 방법

- 전 데이터가 많습니다 + Paired Data
 - 샘플 데이터 확보 후 바로 진행
- 전 데이터가 많습니다 + Unpaired Data
 - 라벨링 및 Pair로 만들 수 있는 지 확인
- 전 데이터가 적습니다 + Paired Data (Labeling Quality가 좋다는 가정)
 - 데이터 단독으로 학습 모델 구축 불가
 - Public Dataset 중에 유사한 Task를 다루고 있어 보강할 수 있는 여지가 있는 지 확인해봐야 함
 - Data Synthesis가 가능한 Task인지 봐야 함
- 전 데이터가 적습니다 + Unpaired Data
 - 그냥 다 해달라는 고객사

각 상황 별 대처 방법

- 전 데이터가 많습니다 + Paired Data
 - 샘플 데이터 확보 후 바로 진행
- 전 데이터가 많습니다 + Unpaired Data
 - 라벨링 및 Pair로 만들 수 있는 지 확인
- 전 데이터가 적습니다 + Paired Data (Labeling Quality가 좋다는 가정)
 - 데이터 단독으로 학습 모델 구축 불가
 - Public Dataset 중에 유사한 Task를 다루고 있어 보강할 수 있는 여지가 있는 지 확인해봐야 함
 - Data Synthesis가 가능한 Task인지 봐야 함
- 전 데이터가 적습니다 + Unpaired Data
 - 그냥 다 해달라는 고객사

쉬는 시간 10분

Cloud API 상 엔진들을 보며
어떤 데이터가 필요할 지 생각해봅시다.

Video Super Resolution

Task 목적

화질이 낮은 video 파일에 대해 2x 등 해상도 높은 video 영상으로 만들 수 있다.

필요한 데이터

- 궁극적으로 필요한 건?
 - 저화질 영상 → 고화질 영상
- 전처리 단계
 - 저화질 영상과 고화질 영상 Pair Data를 찾는다.
 - 없다면, 고화질 영상을 저화질로 낮추고 Noise를 부여한 다음 고화질이 되도록 학습한다.
- 고객사 단계

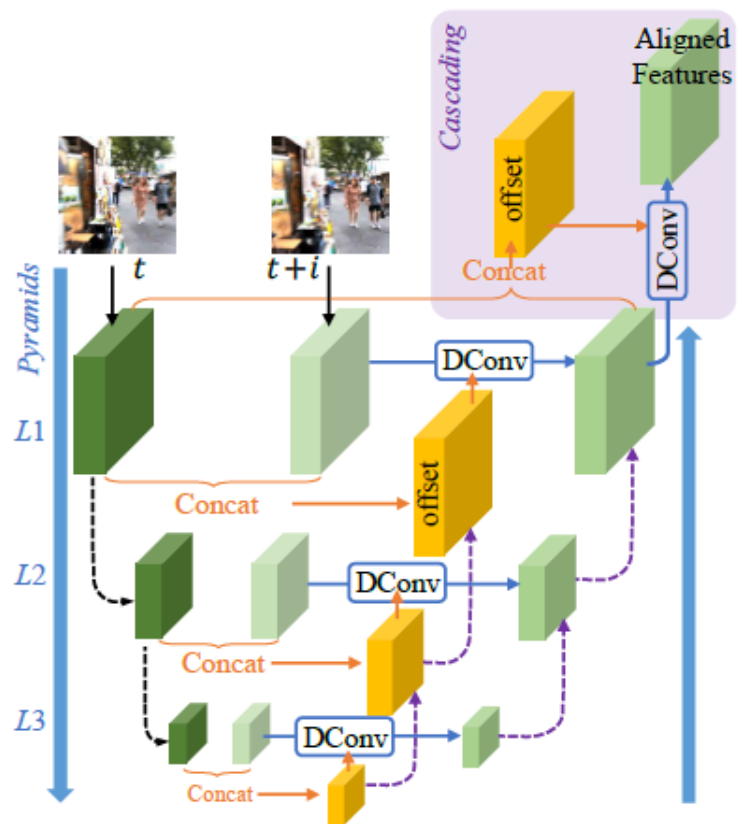


Figure 3. PCD alignment module with Pyramid, Cascading and Deformable convolution.

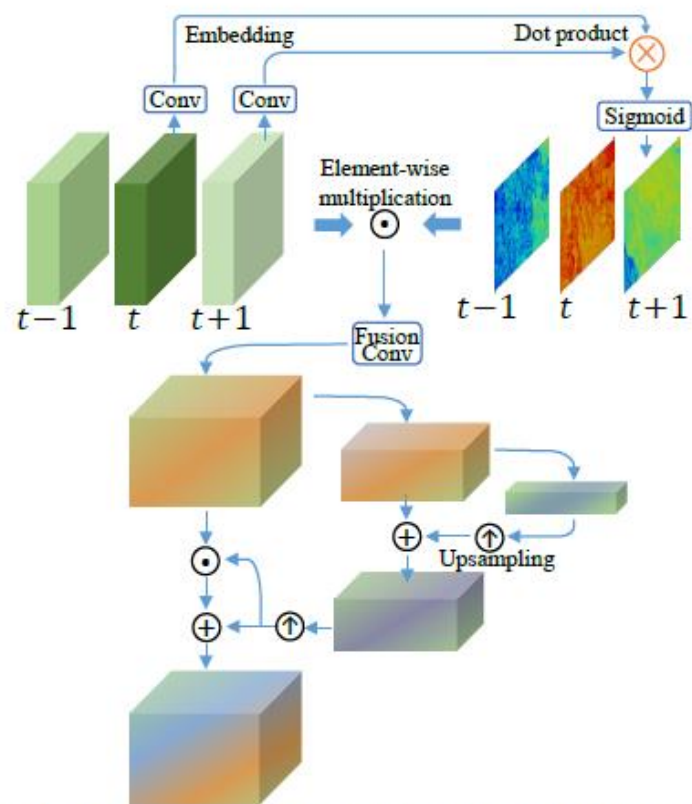


Figure 4. TSA fusion module with Temporal and Spatial Attention.

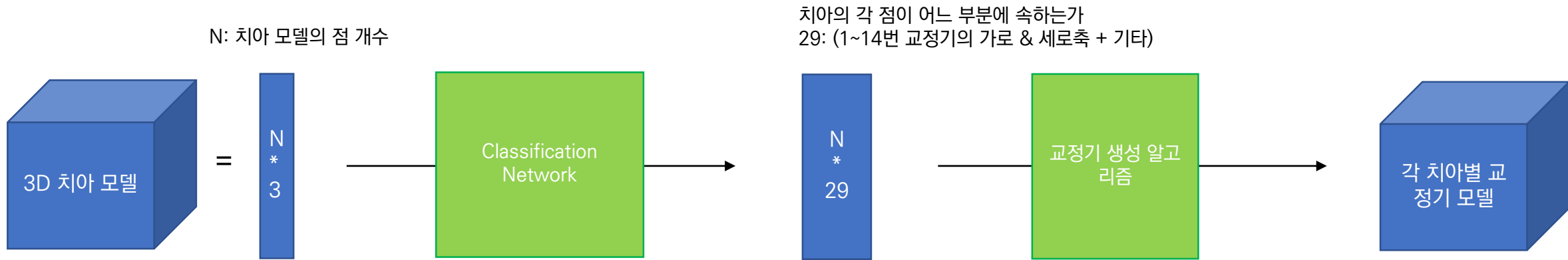
Bracket Positioning

Task 목적

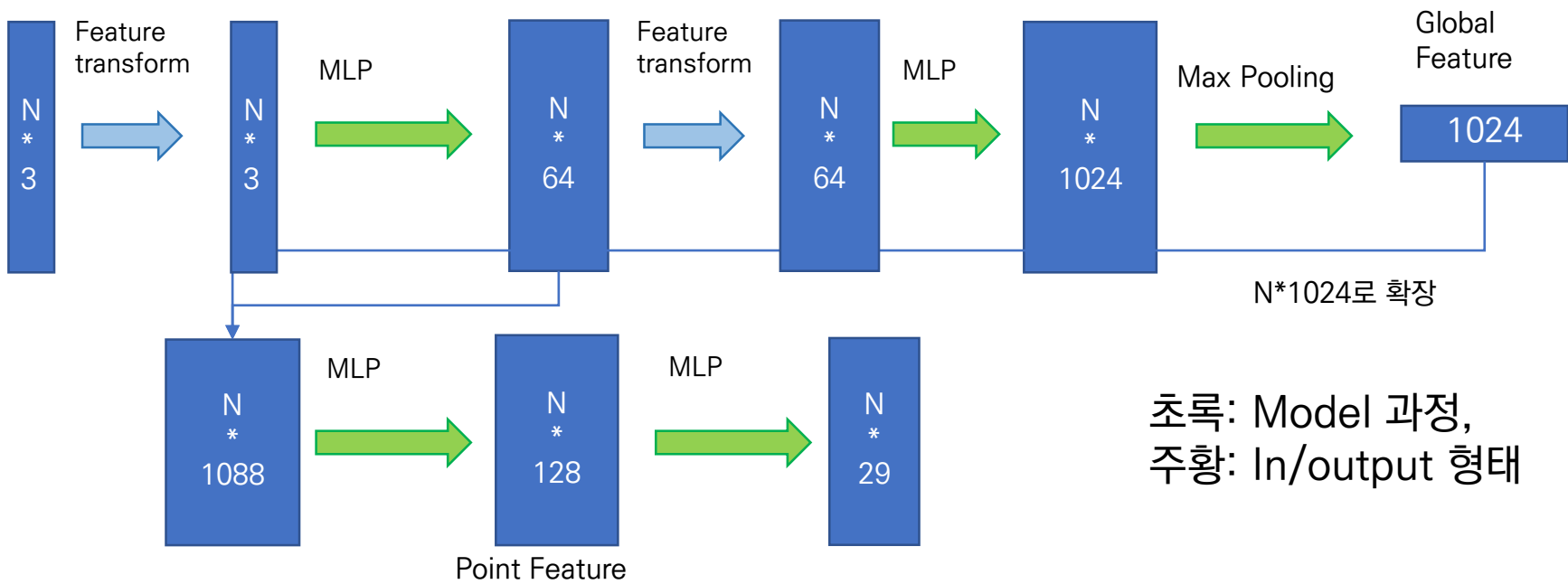
CT Scan된 치아 모델을 보고, 치아 교정기를 끼기에 적합한 위치를 찾는다.

필요한 데이터

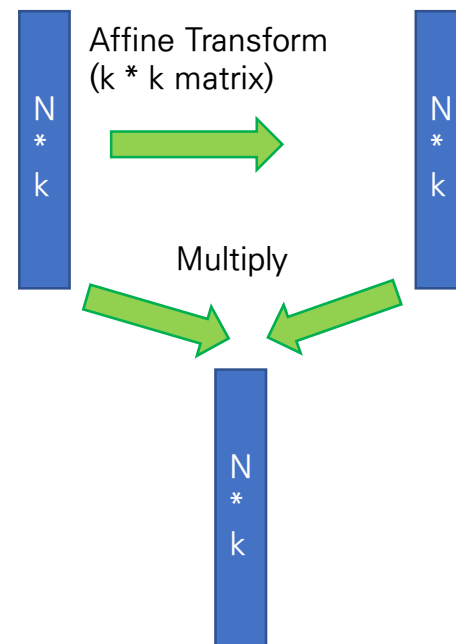
- 궁극적으로 필요한 건?
 - 치아 CT Scan한 데이터와 이 때 어디에 교정기가 부착되어야 하는 지 표시된 데이터
- 전처리 단계
 - 데이터에 라벨링을 직접 하기에는 현실적으로 어려움
 - 치아 CT Scan한 데이터에 대한 확장자 및 Mesh 표현 방식 등을 확인하여 모델에 맞게끔 전처리
- 고객사 단계



Classification Network



Feature Transform



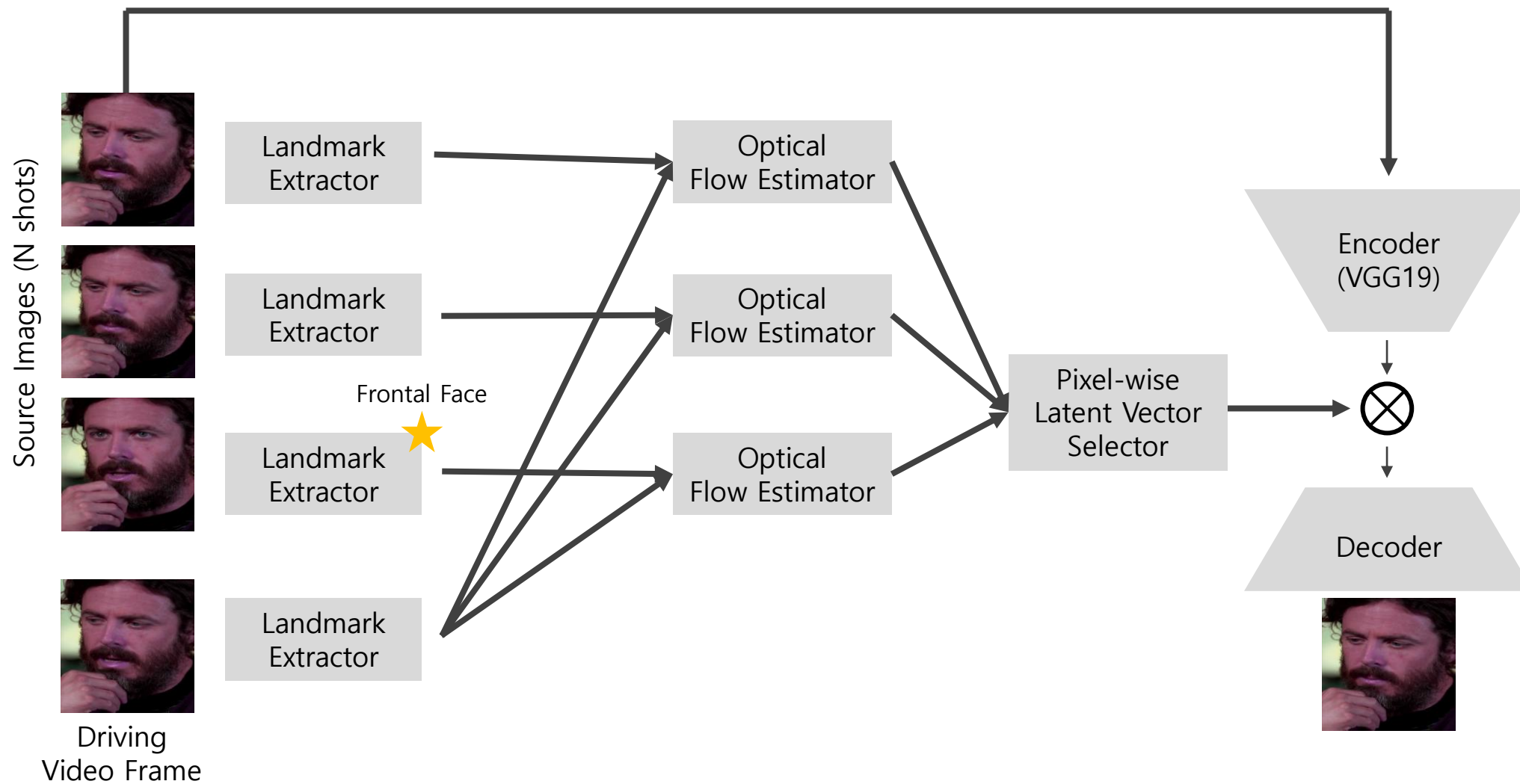
Face-to-Face Translation

Task 목적

유저의 얼굴 움직임을 포착하여, 특정 인물의 사진 한 장 또는 여러 장을 가지고 움직이는 영상을 만들 수 있다.

필요한 데이터

- 궁극적으로 필요한 건?
 - 다양한 사람들이 여러 가지 표정을 하는 얼굴이 담긴 데이터
- 전처리 단계
 - 얼굴 특징점 추출 네트워크를 통해 특징점 추출을 진행하고 이를 기반으로 해당 특징점을 기반으로 얼굴 부분 인식 및 영상 자르기
 - 다양한 사람에 대한 얼굴 데이터를 고객사에서 얻기에는 힘들.
- 고객사 단계



Lip Sync

Task 목적

특정 음성에 대해 이를 발화하는 얼굴 영상을 만든다.

필요한 데이터

- 궁극적으로 필요한 건?
 - 음성과 이를 발화하는 얼굴의 쌍 데이터
- 전처리 단계
 - 영상을 촬영하고 이를 문장 단위로 쪼개야 한다.
 - 얼굴이 감지되지 않는 부분들은 버리고, 얼굴이 감지될 경우, 해당 부분에 대해 Box가 필요하다.
- 고객사 단계

Document OCR + Text Classification

Task 목적

서류에 있는 글씨들의 위치와 텍스트를 인식하고, 해당 텍스트가 어떤 Category에 속하는 지 파악한다.

필요한 데이터

- 궁극적으로 필요한 건?
 - 서류 이미지와 해당 이미지 내 텍스트의 위치, 텍스트 내용이 담긴 라벨 파일
- 전처리 단계
 - 서류 이미지에서 텍스트 영역 감지와 텍스트 영역 인식을 별도 네트워크로 구성하여 학습을 진행해야 한다.
 - 텍스트 데이터는 가공하기가 쉬우므로, 데이터 augmentation 방법이나 공개 데이터가 있는 지 찾아본다.
- 고객사 단계