

---

# Segment Anything and its Adapter

---

2023. 12. 08.

조용원

Data Mining and Quality Analytics Lab



# 발표자 소개



## ❖ 조용원(Yongwon Jo) – Ph.D. Candidate

- 고려대학교 산업경영공학과 석·박통합과정 10학기 재학 중
- Data Mining & Quality Analytics Lab (김성범 교수님)

## ❖ Research Interest

- Semantic Segmentation and Its Applications
- Semi & Self-supervised Learning for the Regression Problem
- Anomaly Detection and Segmentation

- RUSBoost
- Mask R-CNN
- Weakly-supervised Semantic Segmentation
- Human Pose Estimation
- Skeleton-based Human Activity Recognition
- SimMatch for Semi-supervised Learning



# 목차

---

- Introduction
- Segment Anything
- SAM-Adapter
- Conclusion



# Introduction

- Image segmentation

## ❖ 이미지 분할 (Image Segmentation) 이란 ?

- 사진/ 영상 내 특정 영역이 탐지하고자 하는 범주 중 어떤 범주에 속하는지 예측
- 탐지하고자 하는 범주에 대해서 사전에 정의가 필요
- 크게 세 가지 세부 분야로 구성
  - **Semantic** segmentation, **Instance** segmentation, **Panoptic** segmentation

**Semantic** Segmentation



**Instance** Segmentation



**Panoptic** Segmentation



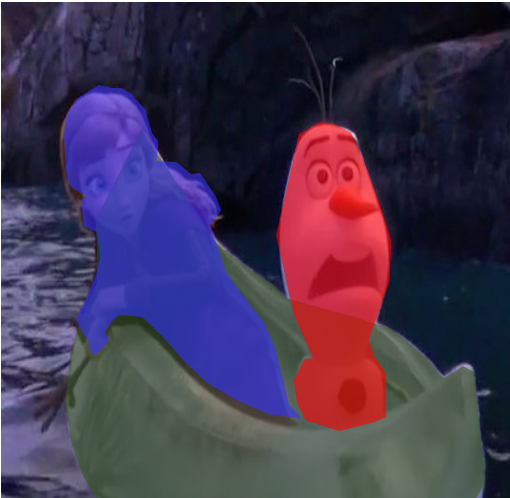
# Introduction

- Semantic segmentation

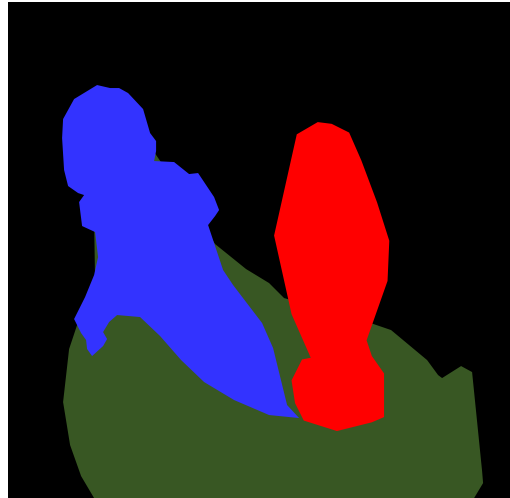
## ❖ Semantic segmentation

- 픽셀 별로 사전에 정의한 범주 중 어떤 범주에 해당하는지 '만' 예측
- 픽셀 별로 탐지하고자 하는 범주가 입력된 정답이 필요(픽셀 별 분류)
- 객체끼리 서로 다름을 인식하지 않고 단순히 픽셀 별로 어떤 범주인지만 인식
- 탐지하고자 하는 범주: 배경(0), 보트(1), 안나(2), 올라프(3)

입력 이미지



정답 이미지



Segmentation Label

0	0	0	0	0	0	0	0	0	0
0	2	2	0	0	0	0	0	0	0
0	2	2	0	0	3	3	0	0	0
0	2	2	2	0	3	3	3	0	0
0	2	2	2	0	3	3	3	0	0
0	2	2	2	1	3	3	3	0	0
0	2	1	2	2	1	3	1	0	0
0	1	1	1	2	1	3	1	1	0
0	1	1	1	1	1	3	1	1	0
0	0	1	1	1	1	1	1	1	0

Semantic Segmentation 모델



# Introduction

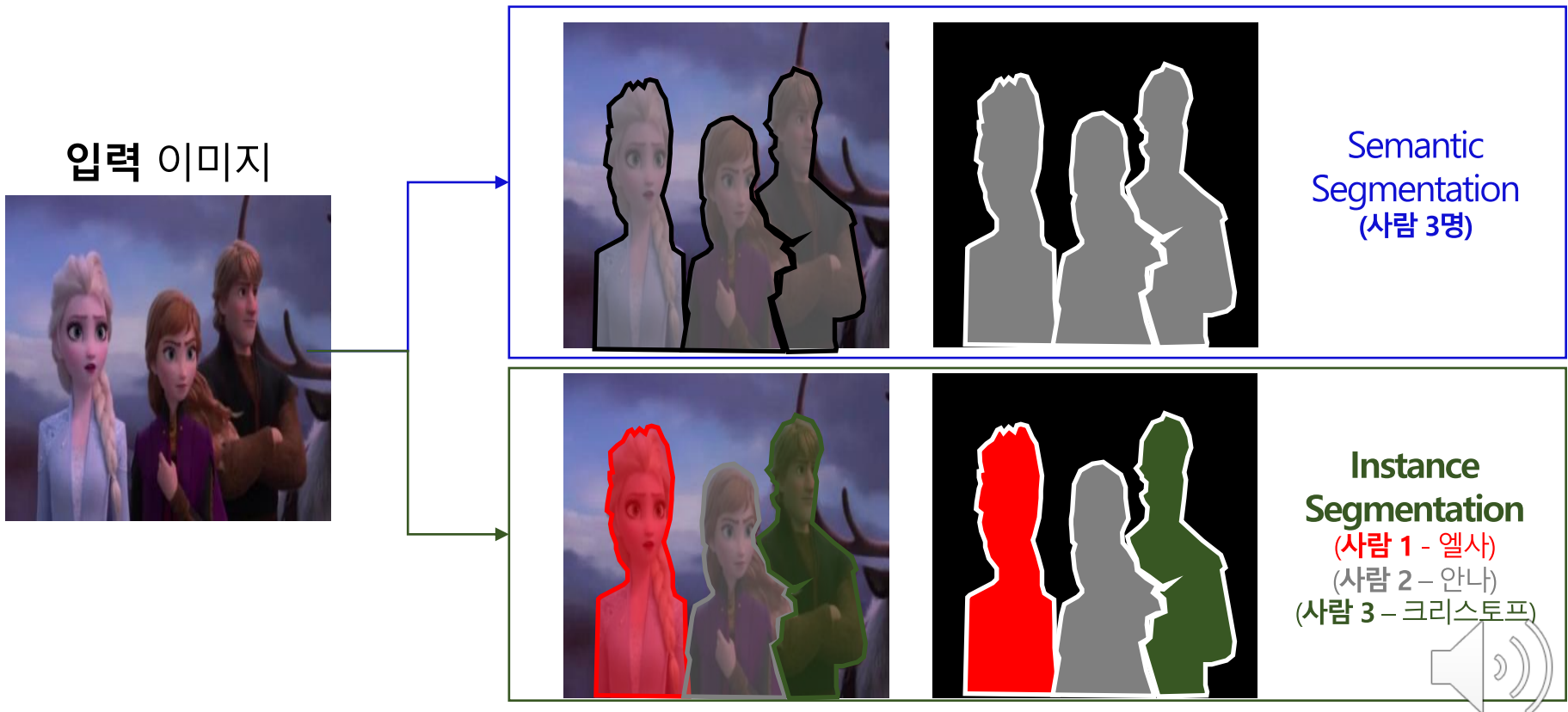
- Instance segmentation



엘사    안나  
크리스토프

## ❖ Instance segmentation

- 픽셀 별 분류 뿐만 아니라 객체간 차이가 있음을 인식하는 문제
  - Semantic segmentation을 수행하면서 어떠한 객체가 있음을 인식(객체간 구별)



# Introduction

- Panoptic segmentation

Semantic Segmentation

+

Instance Segmentation

Panoptic Segmentation

## ❖ Panoptic segmentation

- 배경에 대한 픽셀에 대해서는 **배경이라고 분류 진행(픽셀에 특정 범주 할당-Semantic)**
  - 파란색은 하늘을 의미하는 픽셀이며 회색은 사람을 의미하는 픽셀
- 객체들끼리 서로 다름을 인식하며 픽셀별 분류 진행(객체간 구별-Instance)
  - 빨간색은 엘사, 회색은 안나, 초록색은 크리스토퍼를 의미하며 사람간 구별
- 즉, Semantic segmentation을 수행하며 Instance segmentation 동시 수행

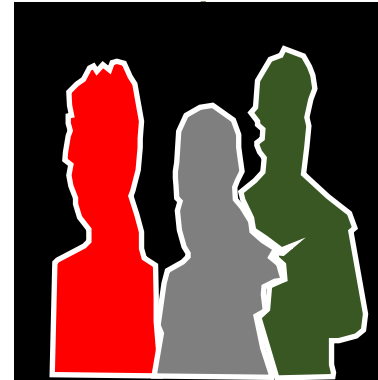
입력 이미지



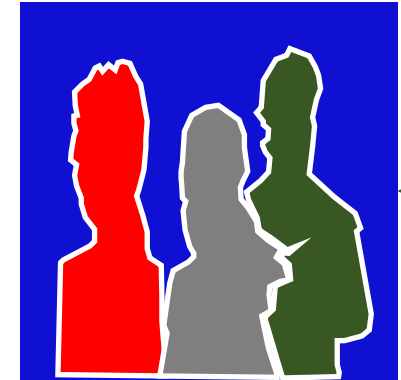
Semantic segmentation



Instance segmentation

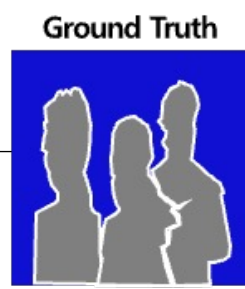
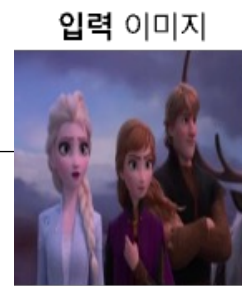


Panoptic segmentation



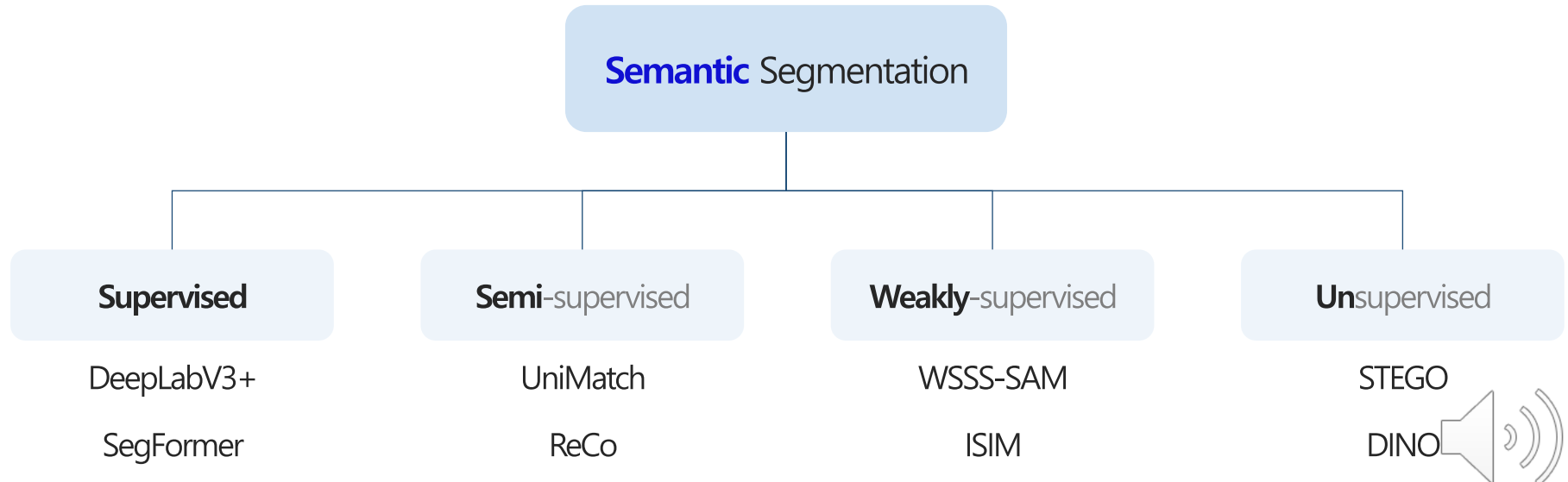
# Introduction

- How to train semantic segmentation models?



## ❖ Semantic segmentation(SS) 모델 학습 방식

- 기계학습 모델 학습 방식과 유사하게 Ground Truth(GT) 사용 방식에 따라 구분
  - **Supervised** SS: 많은 (입력 이미지-GT) 쌍을 사용해 모델 학습
  - **Semi-supervised** SS: 적은 (입력 이미지-GT) 쌍과 많은 입력 이미지를 사용해 모델 학습
  - **Weakly-supervised** SS: 많은 (입력 이미지-Weak label)을 사용해 모델 학습
  - **Unsupervised** SS: 입력 이미지만 사용해 모델 학습
- 입력 이미지에 대한 GT 생성 어려움을 해결하기 위해 연구되고 있는 분야





# Introduction

- Weakly-supervised Semantic Segmentation

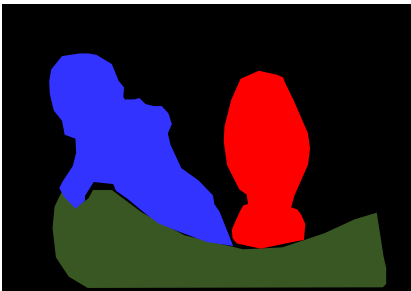
## ❖ Weakly-supervised Semantic Segmentation

- 상대적으로 작은(Weakly) 정보로 Semantic Segmentation을 수행하는 것이 목표
  - 상대적으로 작은 정보를 Weakly supervision 이라고 부름
  - 위치 정보(Bounding Box, Point), 범주 정보를 Weakly supervision 으로 사용

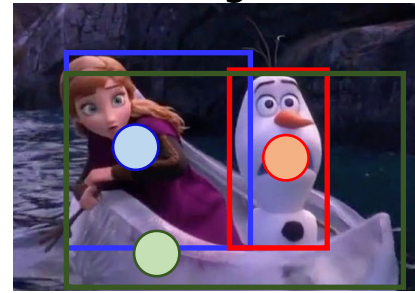
Semantic Segmentation



올라프    안나    보트



Weakly Supervised  
Semantic Segmentation



# Introduction

- ChatGPT (Foundation model of NLP)

## ❖ ChatGPT(Chatting + Generative Pre-trained Transformer)

- OpenAI에서 개발한 대형 언어 모델(Large Language Model, LLM)
- 굉장히 많은 웹 텍스트 데이터로 구성된 데이터 셋을 사용해 모델 학습
  - 문장 내 빈칸에 대한 단어를 예측하도록 모델 학습 후 강화학습을 사용해 미세조정(Fine-tune) 진행





# Introduction

- ChatGPT (Foundation model of NLP)

## ❖ ChatGPT 활용 예시

- 메시지를 입력하면 이에 대한 답변을 출력해주며 대화 생성을 진행
- ChatGPT 기술에 따르면 메시지를 Prompt라 명명하고 Prompt에 대한 답변(Response)을 생성

 **You**  
What is the semantic segmentation task in computer vision?

 **ChatGPT**  
Semantic segmentation is a computer vision task that involves classifying and labeling each pixel in an image with a corresponding class label. The goal is to understand the meaning or semantics of each part of the image. In other words, the objective is to partition an image into meaningful segments and assign a specific class label to every pixel based on its content.

Unlike other image segmentation tasks that might focus on grouping pixels into regions based on color or texture, semantic segmentation goes a step further by associating each segment with a specific object or class. This fine-grained level of labeling is crucial for applications where a detailed understanding of the visual scene is required.

Here are some common use cases for semantic segmentation:

Prompt

Response

# 목차

---

- Introduction
- **Segment Anything**
- SAM-Adapter
- Conclusion



# Segment Anything

- Segment Anything(2023, arXiv)

## ❖ Segment Anything (SAM)

- Meta (과거 Facebook)에서 공개한 이미지 데이터를 위한 ChatGPT

Panoptic Segmentation 정의

## Segment Anything

Alexander Kirillov<sup>1,2,4</sup>

Eric Mintun<sup>2</sup>

Nikhila Ravi<sup>1,2</sup>

Hanzi Mao<sup>2</sup>

Chloe Rolland<sup>3</sup>

Laura Gustafson<sup>3</sup>

Tete Xiao<sup>3</sup>

Spencer Whitehead

Alexander C. Berg

Wan-Yen Lo

Piotr Dollár<sup>4</sup>

Ross Girshick<sup>4</sup>

<sup>1</sup>project lead

<sup>2</sup>joint first author

<sup>3</sup>equal contribution

<sup>4</sup>directional lead

Focal Loss, Faster R-CNN,  
Mask R-CNN

Meta AI Research, FAIR

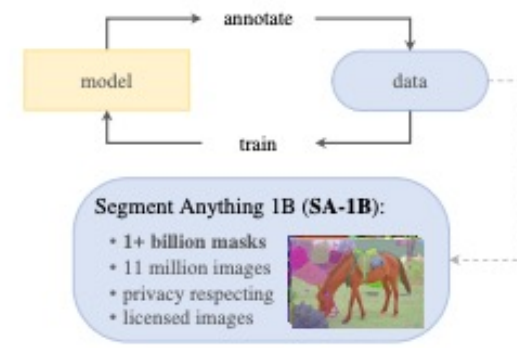
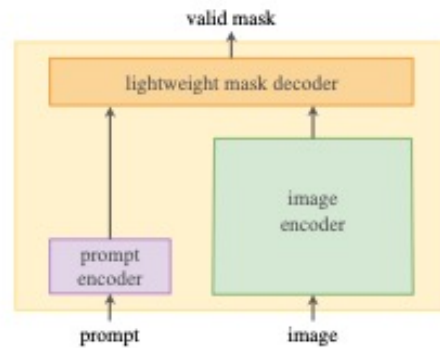
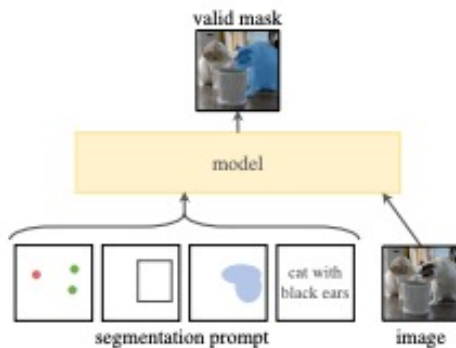


Figure 1: We aim to build a foundation model for segmentation by introducing three interconnected components: a promptable segmentation *task*, a segmentation *model* (SAM) that powers data annotation and enables zero-shot transfer to a range of tasks via prompt engineering, and a *data* engine for collecting SA-1B, our dataset of over 1 billion masks.

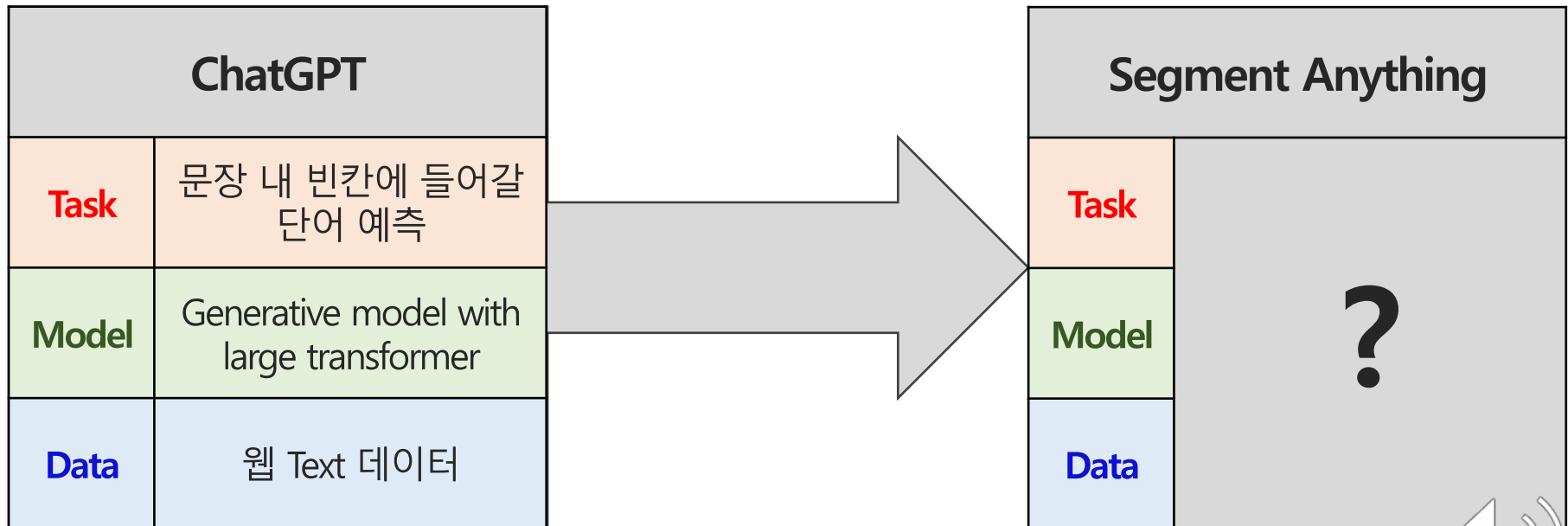


# Segment Anything

- Motivation from ChatGPT

## ❖ SAM 저자들이 가진 의문

- Segmentation을 위한 ChatGPT와 같은 근본(Foundation) 모델을 만들 수 없을까?
  - What **task** will enable zero-shot generalization? (학습 방식)
  - What is the corresponding **model** architecture? (모델 구조)
  - What **data** can power this task and model? (데이터)

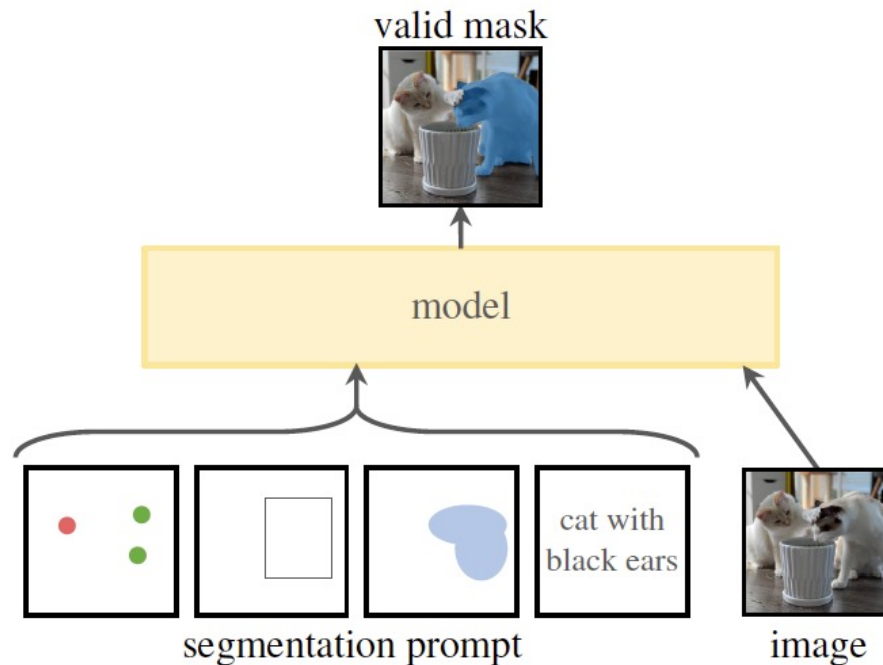


# Segment Anything

- Task of SAM

## ❖ What **task** will enable zero-shot generalization? (Task 정의)

- 원하는 객체만 추출할 수 있도록 Prompt segmentation이라는 문제 정의
- 입력 이미지와 Prompt를 입력 받아 Prompt가 설명하는 객체를 추출하는 문제
- Prompt로 Weakly-supervision을 사용하고 객체에 대한 Segmentation 결과를 Mask로 정의



(a) **Task:** promptable segmentation

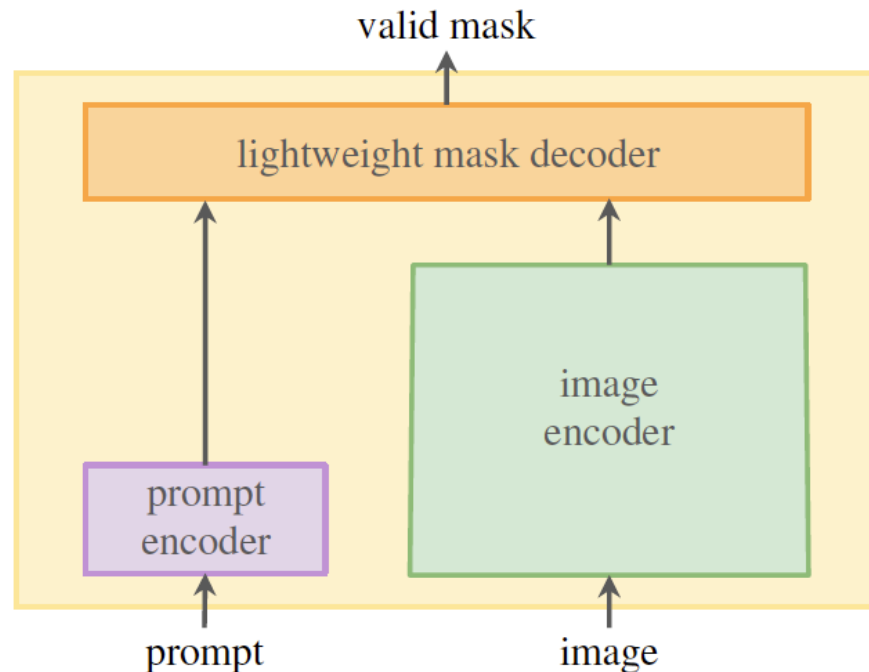


# Segment Anything

- Model of SAM

## ❖ What is the corresponding **model** architecture? (Segment Anything Model)

- 두 Encoder와 단일 Mask decoder로 구성
- 입력 이미지와 Prompt에 대한 특징을 추출할 수 있는 서로 다른 Encoder 구성
- 두 Encoder에서 나온 표현 벡터를 결합해 원하는 객체만 추출하는 Decoder 존재



(b) **Model: Segment Anything Model (SAM)**



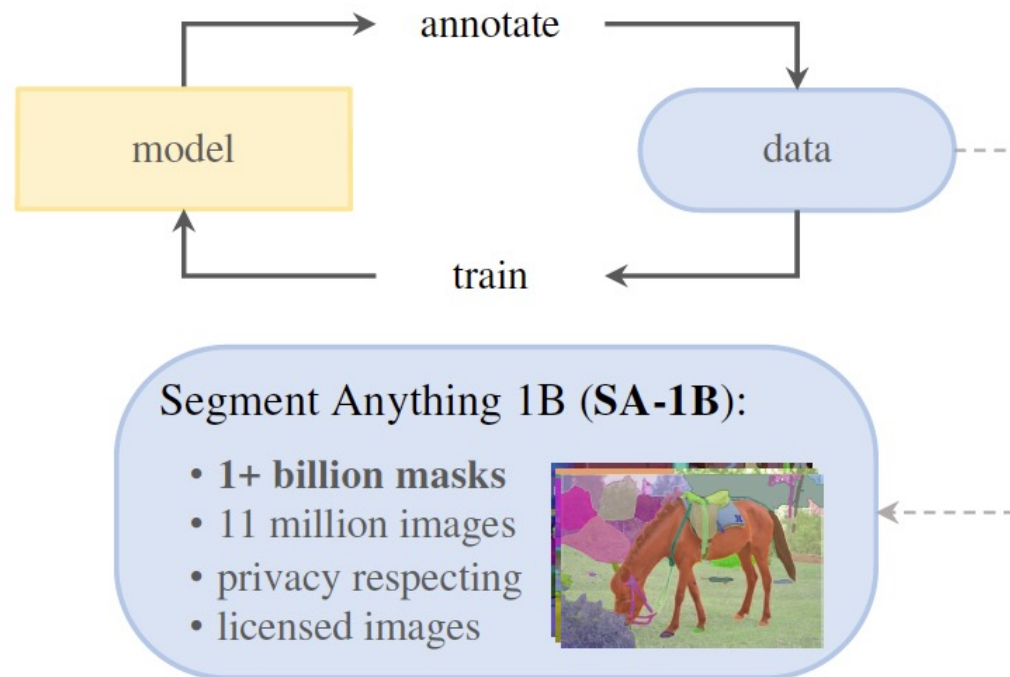


# Segment Anything

- Data of SAM

## ❖ What **data** can power this task and model? (데이터)

- 1,100 만장 입력 이미지와 이미지 내 객체에 대한 Mask 10억장 이상을 사용
- 입력 이미지는 사진 작가들에게 직접 받아 License 및 사생활 보호 처리 완료된 이미지
- 어떻게 Mask 10억장 이상을 만든 것일까? (Annotate ?)

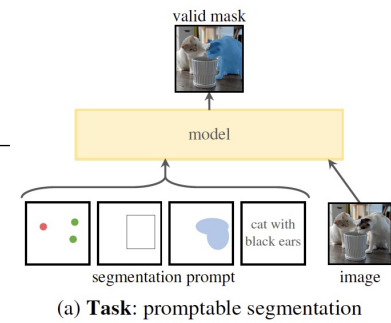


(c) **Data:** data engine (top) & dataset (bottom)



# Segment Anything

- Definition of Promptable Segmentation



## ❖ Promptable Segmentation

- 입력 데이터(X): 입력 이미지와 탐지하려는 객체에 대한 Prompt
- 출력 데이터(Y): 원하는 객체에 대한 Segmentation 결과
- Prompt가 애매모호하게 입력되었음에도 불구하고 객체에 대한 타당한 Mask 도출 필요

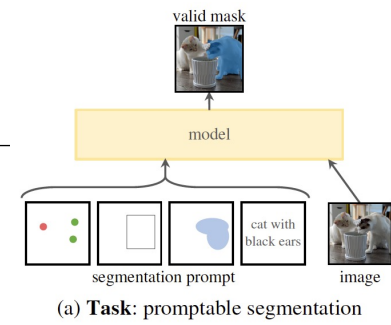


- 타조
  - 머리/ 상체/ 타조 전체
- 가방 내 지퍼
  - 지퍼/ 가방/ 가방 메고 있는 사람
- 팔목
  - 사람/ 시계 제외한 팔목/ 모든 팔목
- 'Z' 글자
  - Z/ ZURICH/ 모든 문자



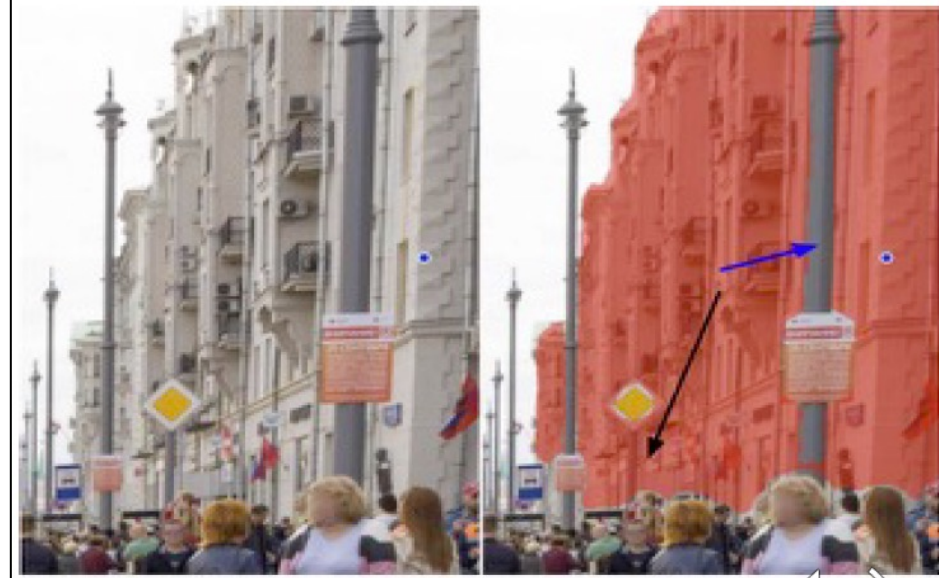
# Segment Anything

- Ambiguity result



## ❖ 애매모호한 객체 인식(Ambiguous mask)

- Prompt가 애매모호하게 입력되었음에도 불구하고 객체에 대한 타당한 Mask 도출 필요
  - 도마뱀 인식을 위해 파란점을 찍었으나 주변 새가 같이 인식
  - 건물을 인식하기 위해 파란점을 찍었으나 폴대가 건물이라 인식하는 경우
- 해당 문제 발생을 줄이기 위해 모델 학습은 어떻게 진행할까?

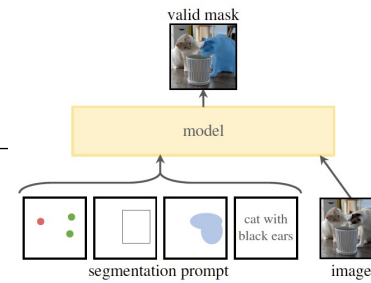


# Segment Anything

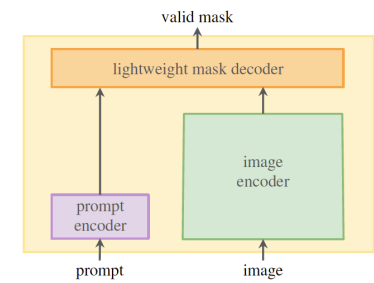
- Model architecture

## ❖ SAM 구조

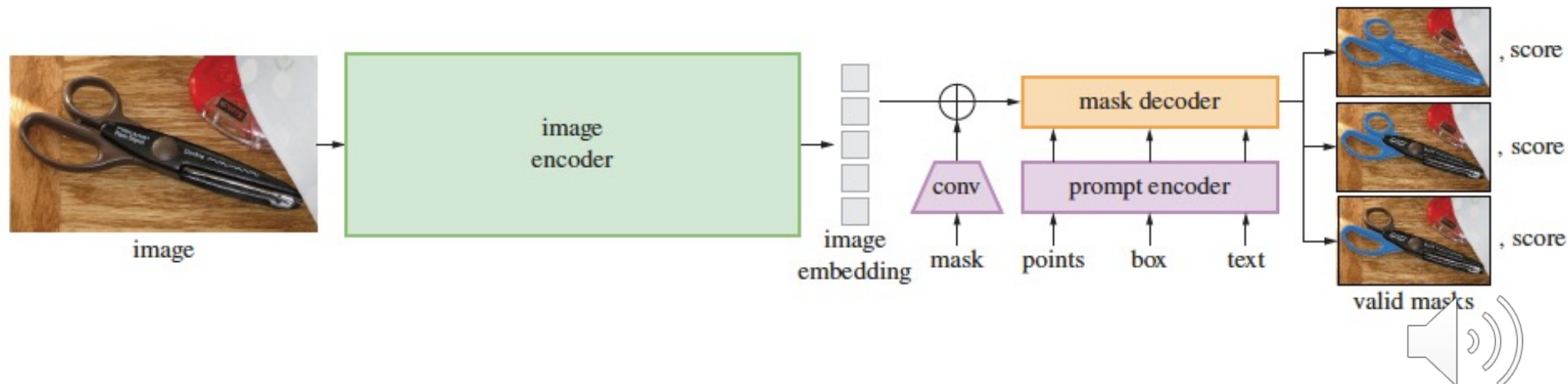
- SAM 학습을 위한 적절한 문제는 Prompt segmentation이라 정의
  - 모델 학습을 위해서 입력 이미지와 Prompt와 필요
- 입력된 서로 다른 데이터에 대한 표현을 추출할 수 있는 Encoder 필요
- 이미지와 Prompt에 대한 표현을 결합해 객체를 인식하는 Mask decoder 필요



(a) Task: promptable segmentation



(b) Model: Segment Anything Model (SAM)

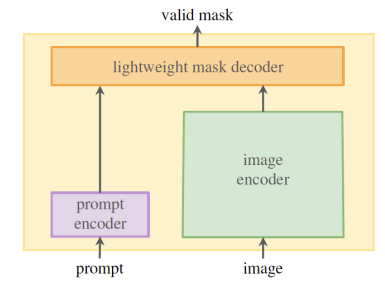


# Segment Anything

- Model architecture

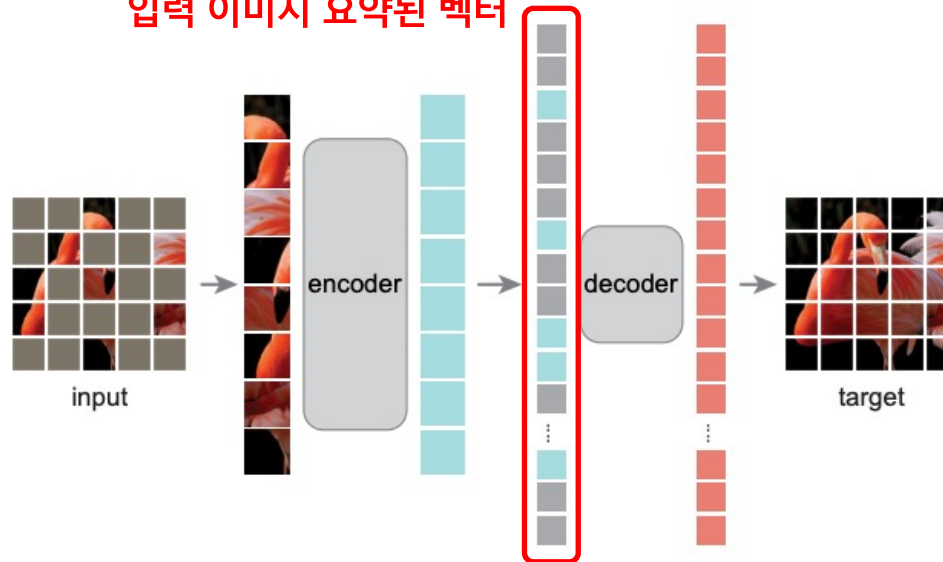
## ❖ Image encoder in SAM

- Masked Autoencoder(MAE) 라는 사전 학습 방식을 사용해 학습된 Encoder 사용
  - MAE는 Visual transformer를 기반으로 하는 Encoder, Decoder로 구성된 Autoencoder 학습
  - 입력 데이터는 이미지에 대한 여러 패치 중 25% 이며 사용 출력 데이터는 원본 이미지
  - **Encoder 산출물인 표현 벡터**는 입력 이미지에 대한 전반적인 특징이 포함된 벡터일 것
- 사전 학습된 모델 내 Encoder를 SAM 내 Image encoder로 사용



(b) Model: Segment Anything Model (SAM)

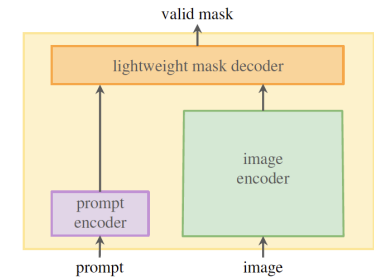
입력 이미지 요약된 벡터



- He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). Masked autoencoders are scalable vision learners. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 16000-16009).

# Segment Anything

- Model architecture

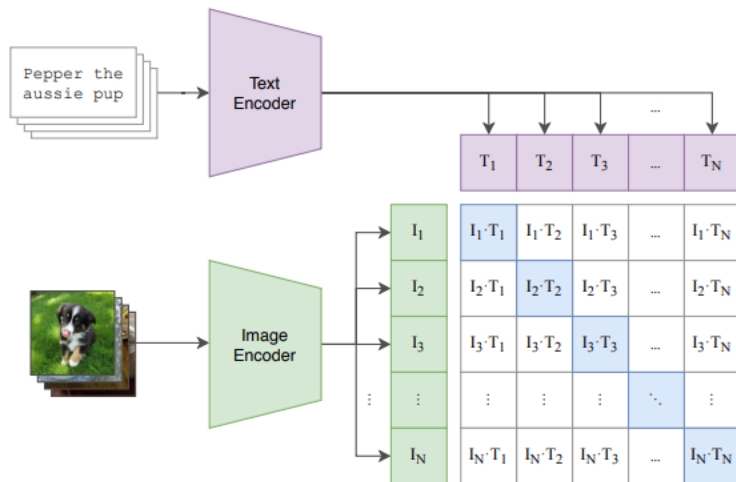


(b) Model: Segment Anything Model (SAM)

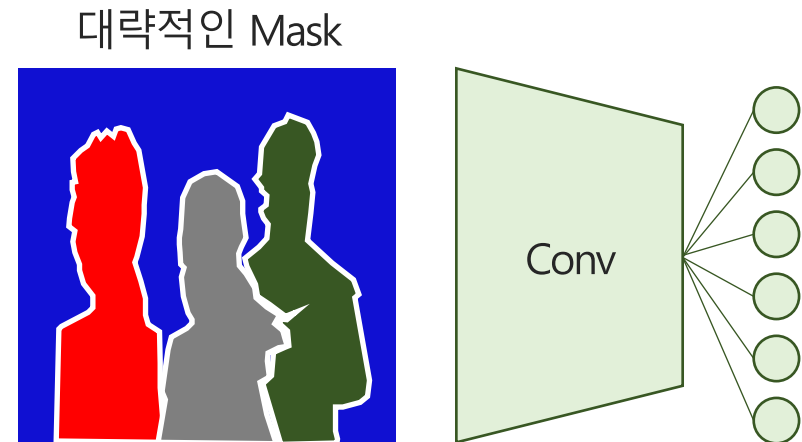
## ❖ Prompt encoder in SAM

- SAM에서 사용하는 Prompt 종류는 Point, Bounding Box, 텍스트, 대략적인 Mask
- Point, Bounding Box에 대해서는 해당 위치에 대해 Positional embedding 진행
- 텍스트에 대한 Encoder는 CLIP 방식으로 학습된 Text encoder 사용
- 대략적인 Mask에 대해서는 합성곱 연산을 사용해 해당 정보 Encoding 진행

### 텍스트 Prompt Encoder



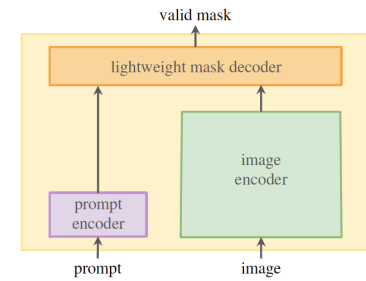
### Mask Prompt Encoder



- Radford, A, Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July) Learning transferable visual models from natural language supervision. In International conference on machine learning (pp. 8748-8763). PMLR.

# Segment Anything

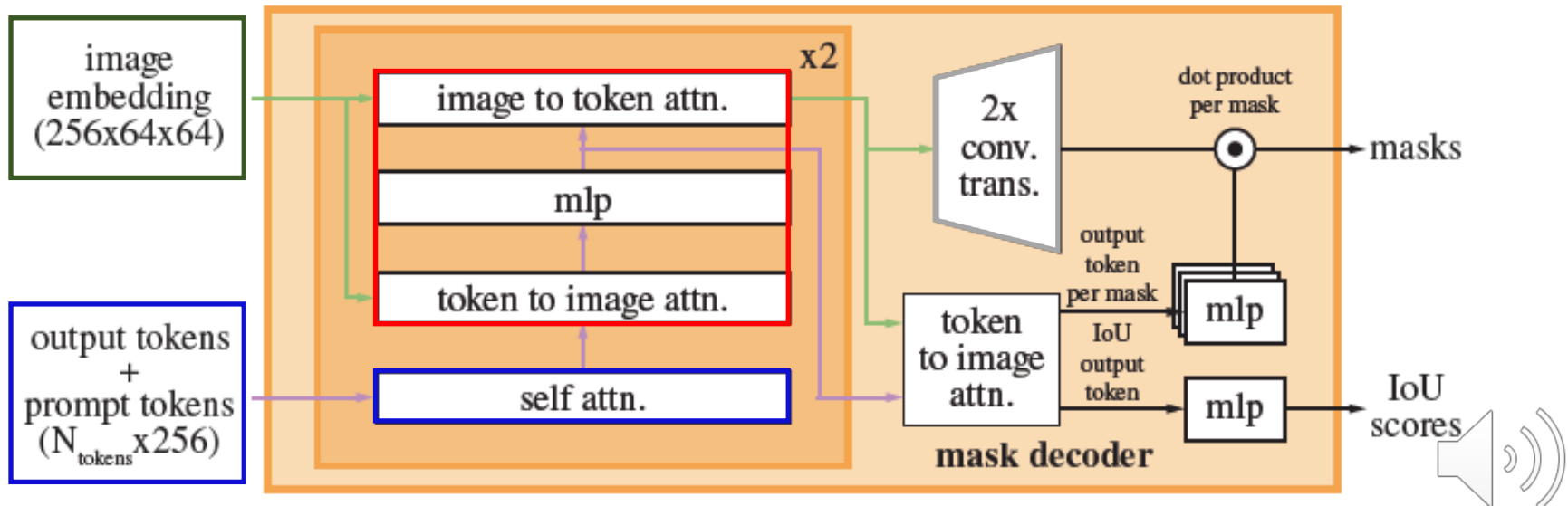
- Model architecture



(b) Model: Segment Anything Model (SAM)

## ❖ Mask Decoder in SAM

- Image encoder를 통과한 **특징**과 Prompt encoder 통과한 **특징**을 결합해 Mask Decoder 입력
- **Prompt 자체에 대한 Attention**과 **양방향 Attention**을 통해 서로 특징 융합
  - 양방향 Attention = (Image to token Attention) + (Token to Image Attention)
- 이미지에 대한 특징을 원본 사이즈로 복원(**Transposed Convolution**)해 여러 Mask 생성
  - 애매 모호한 Mask(Ambiguity) 제거를 위해 서로 다른 세 개 Mask 생성

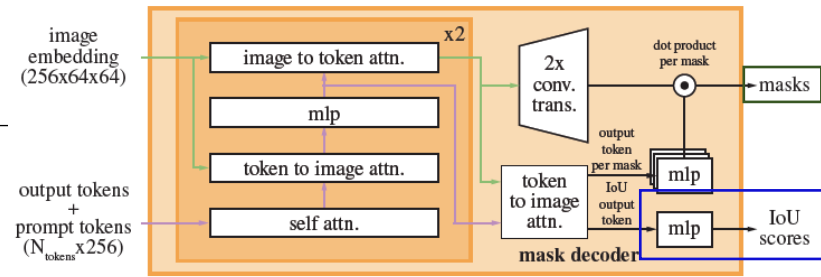


# Segment Anything

- Loss function of SAM

## ❖ SAM 학습을 위한 손실 함수 산출

- Mask Decoder에서 산출한 세 개 Mask와 실제 Mask 사이 손실 함수 산출
  - Mask에 대한 손실 함수 = (20\*Focal 손실 함수 + 1\*Dice 손실 함수)
- Mask 별 IoU 값 예측을 위한 MLP 출력 값과 실제 IoU 사이 평균 제곱 오차 손실 함수 산출
- 가장 큰 예측 IoU를 가진 Mask에 대한 손실 함수만 모델 학습에 활용(Ambiguity 제거)



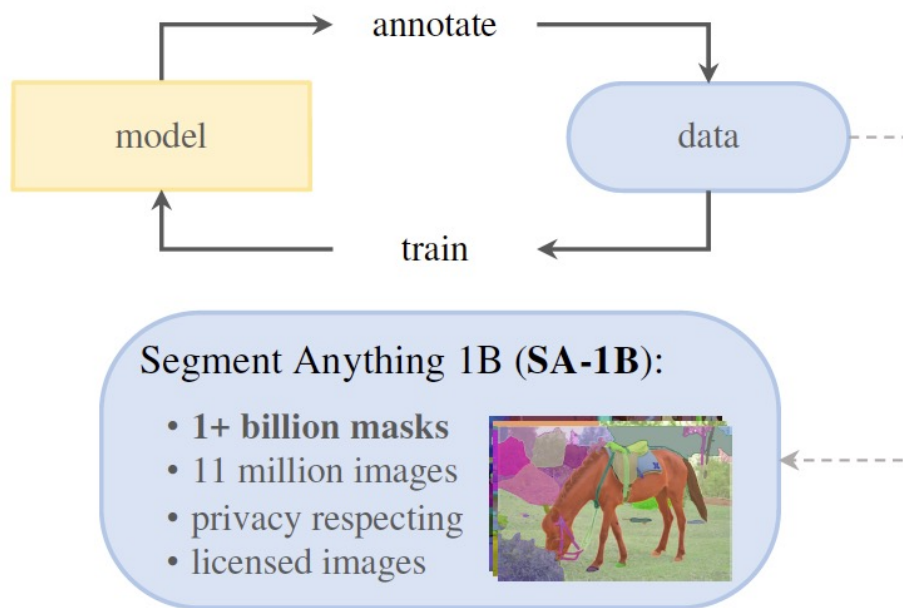


# Segment Anything

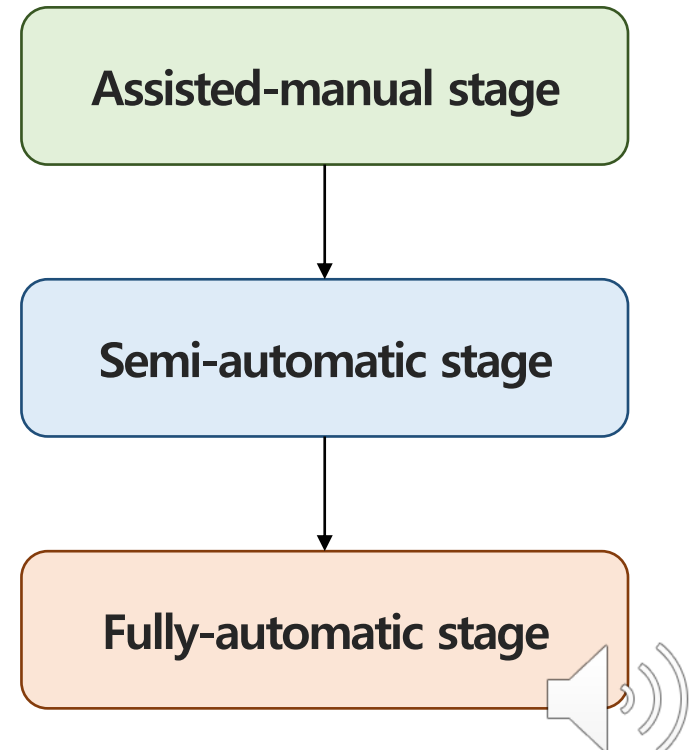
- Data Engine

## ❖ SAM 학습을 위한 데이터 생성

- 1,100만장 입력 이미지와 이에 대한 1억 개 이상 Mask를 어떻게 다 만든 것일까?
- 사람이 해당 Mask를 직접 레이블링 했을까?



(c) **Data:** data engine (top) & dataset (bottom)



# Segment Anything

- Data Engine

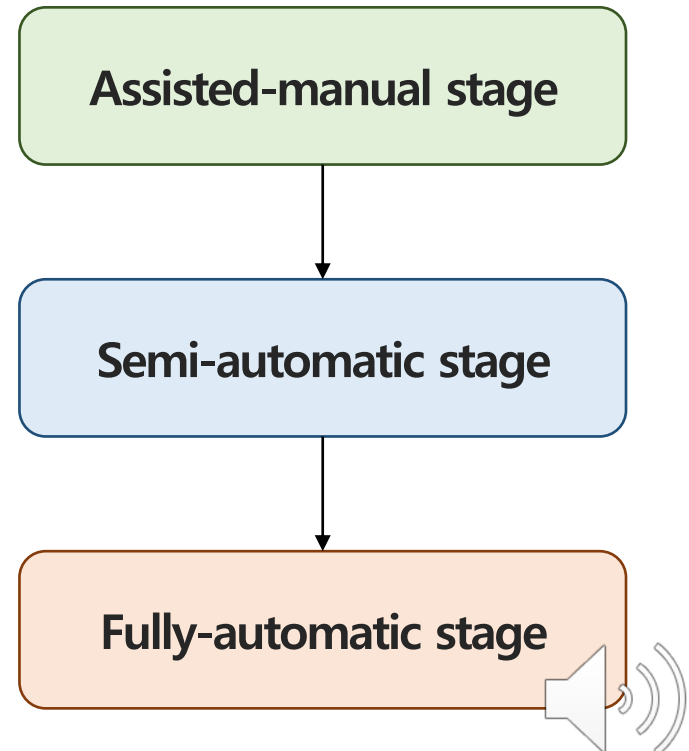


## ❖ Assisted-manual stage & Semi-automatic stage

- (Assisted) Labeling 전문가가 특정 이미지에 대한 Point prompt 및 Mask 생성
- (Semi) 모델 예측 결과를 Prompt로 입력해 더 정확한 Mask 생성 및 전문가 보정
- 즉, (Assisted+Semi) stage까지는 사람 개입이 필요한 데이터 생성 단계

전문가 직접 생성(Prompt, Mask)

모델이 예측한 Prompt → 모델 Mask 예측  
+  
전문가 Mask 보정

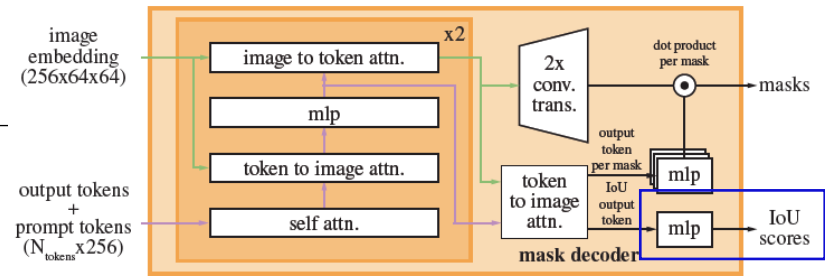


# Segment Anything

- Data Engine

## ❖ Fully-automatic stage

- 마지막 단계에서는 사람 개입 없이 엄청난 Mask를 생성(1억 개 이상 Mask 중 99% 이상)
- 총 1024개(32 X 32 Grid) Point prompt를 생성해 각각 점에 대한 Mask를 생성
- Point prompt에 대한 여러 Mask 중 IoU 스코어가 가장 높은 데이터를 추출해 결과물 생성



Assisted-manual stage

Semi-automatic stage

Fully-automatic stage

- <https://thecho7.tistory.com/entry/%EB%85%BC%EB%AC%B8-%EB%A6%AC%EB%B7%B0-Segment-Anything-%EC%84%A4%EB%AA%85-%EC%BD%94%EB%93%9C-%ED%8F%AC%ED%95%88>

# Segment Anything

- Data Engine

## ❖ Fully-automatic stage

- Fully-automatic stage를 사용해 생성한 Mask 예시

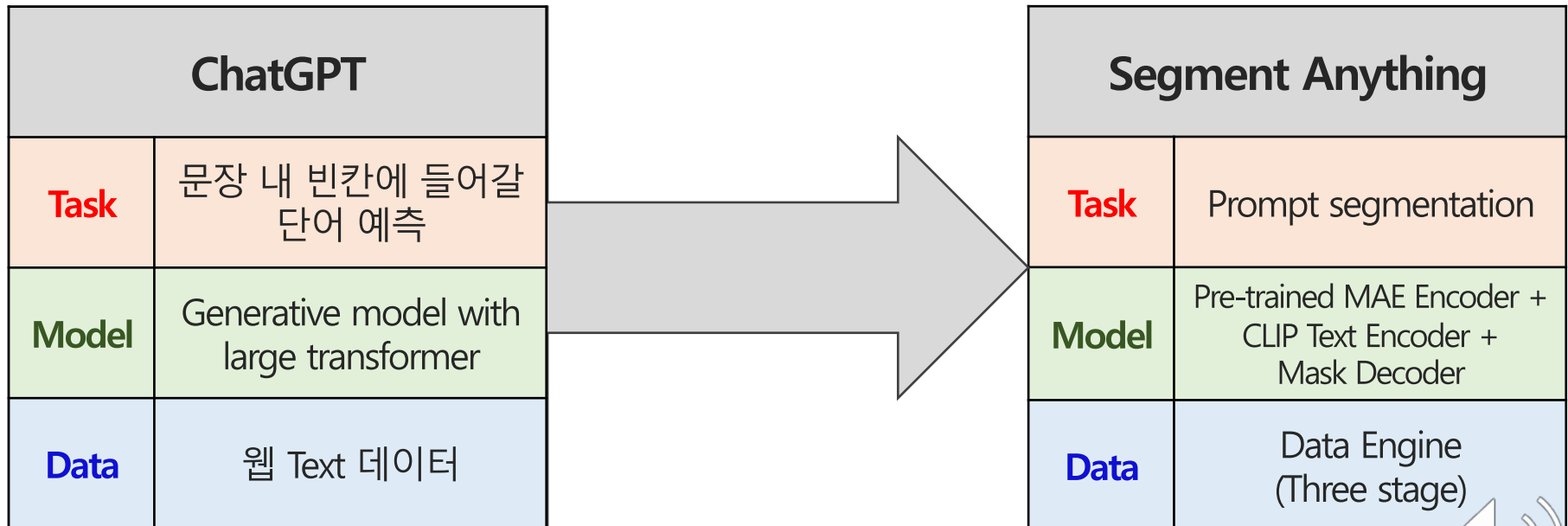


# Segment Anything

- Summary

## ❖ SAM 요약

- 연구 시작 시 가졌던 물음에 대한 답변
  - What **task** will enable zero-shot generalization?
  - What is the corresponding **model** architecture?
  - What **data** can power this task and model?



# Segment Anything

- 데이터에 대한 윤리적 검토

## ❖ Responsible AI

- 여러 학회에서는 개발한 모델이 인종/ 지역에 관계 없이 잘 작동하는 지 여부를 확인
- 이를 위해 SAM 저자들은 인종에 따른 사람 검출 성능에 대해 리포트
- 수집한 입력 이미지들이 어떠한 국가 및 대륙에서 수집되었는지도 리포트

	mIoU at	
	1 point	3 points
<i>perceived gender presentation</i>		
feminine	54.4 ± 1.7	90.4 ± 0.6
masculine	55.7 ± 1.7	90.1 ± 0.6
<i>perceived age group</i>		
older	62.9 ± 6.7	92.6 ± 1.3
middle	54.5 ± 1.3	90.2 ± 0.5
young	54.2 ± 2.2	91.2 ± 0.7

	mIoU at	
	1 point	3 points
<i>perceived skin tone</i>		
1	52.9 ± 2.2	91.0 ± 0.9
2	51.5 ± 1.4	91.1 ± 0.5
3	52.2 ± 1.9	91.4 ± 0.7
4	51.5 ± 2.7	91.7 ± 1.0
5	52.4 ± 4.2	92.5 ± 1.4
6	56.7 ± 6.3	91.2 ± 2.4

	# countries	SA-1B		% images		
		#imgs	#masks	SA-1B	COCO	O.I.
Africa	54	300k	28M	2.8%	3.0%	1.7%
Asia & Oceania	70	3.9M	423M	36.2%	11.4%	14.3%
Europe	47	5.4M	540M	49.8%	34.2%	36.2%
Latin America & Carib.	42	380k	36M	3.5%	3.1%	5.0%
North America	4	830k	80M	7.7%	48.3%	42.8%
high income countries	81	5.8M	598M	54.0%	89.1%	87.5%
middle income countries	108	4.9M	499M	45.0%	10.5%	12.0%
low income countries	28	100k	9.4M	0.9%	0.4%	0.5%

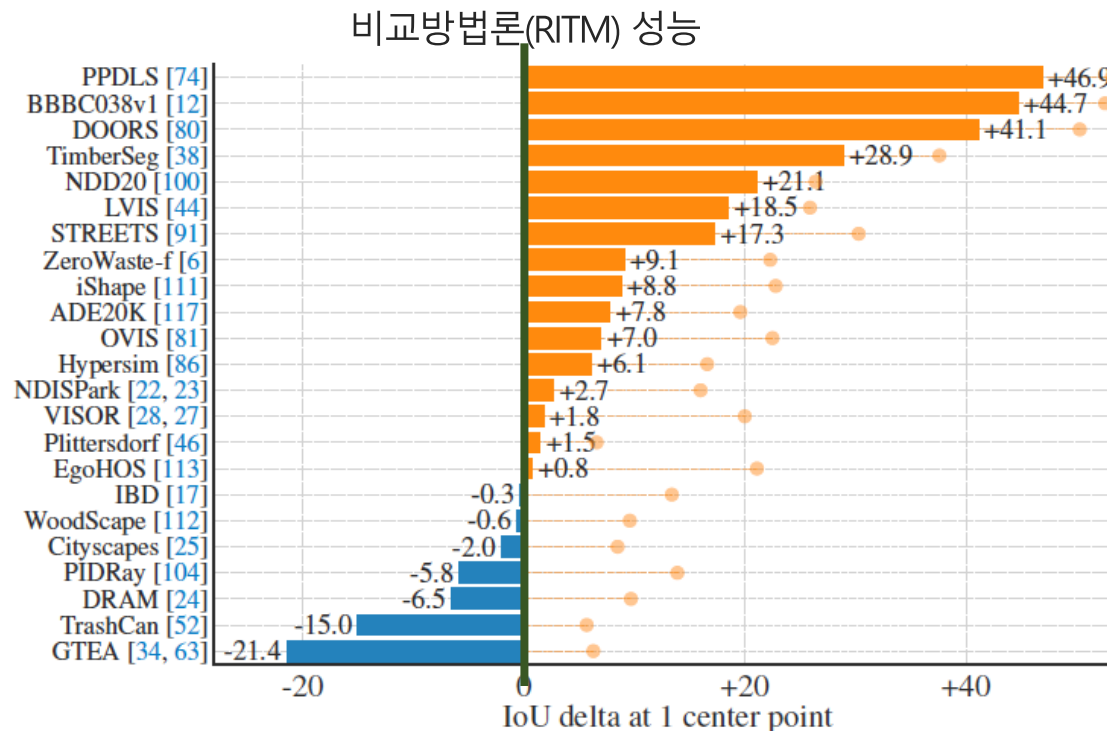


# Segment Anything

- Zero-shot Transfer Learning Experiments

## ❖ Single Point Prompt 를 사용해 생성한 Mask 에 대한 성능

- 해당 성능 도출을 위해 학습에 사용하지 않았던 공개된 Segmentation 데이터 셋 사용
  - Zero-shot Learning in SAM: 학습하지 않은 문제(ex. Semantic segmentation)에 대한 성능
- 특정 이미지에 단일 Point Prompt 만 입력해서 Mask 세 장 생성(포인트는 가장 신뢰하는 것)
- 실제 Ground truth와 비교해서 mean Intersection over Union(mIoU) 산출



23개 데이터 셋 중

16개 데이터 셋 SAM 승리



# Segment Anything

- Zero-shot Transfer Learning Experiments

## ❖ Zero-shot Edge Detection 문제를 수행한 결과

- SAM을 사용해 Mask를 생성한 뒤, Sobel 필터를 사용해 Edge로 변환
- 기존 Edge Detection 데이터 셋 내 Ground truth보다 정확하게 Edge를 인식



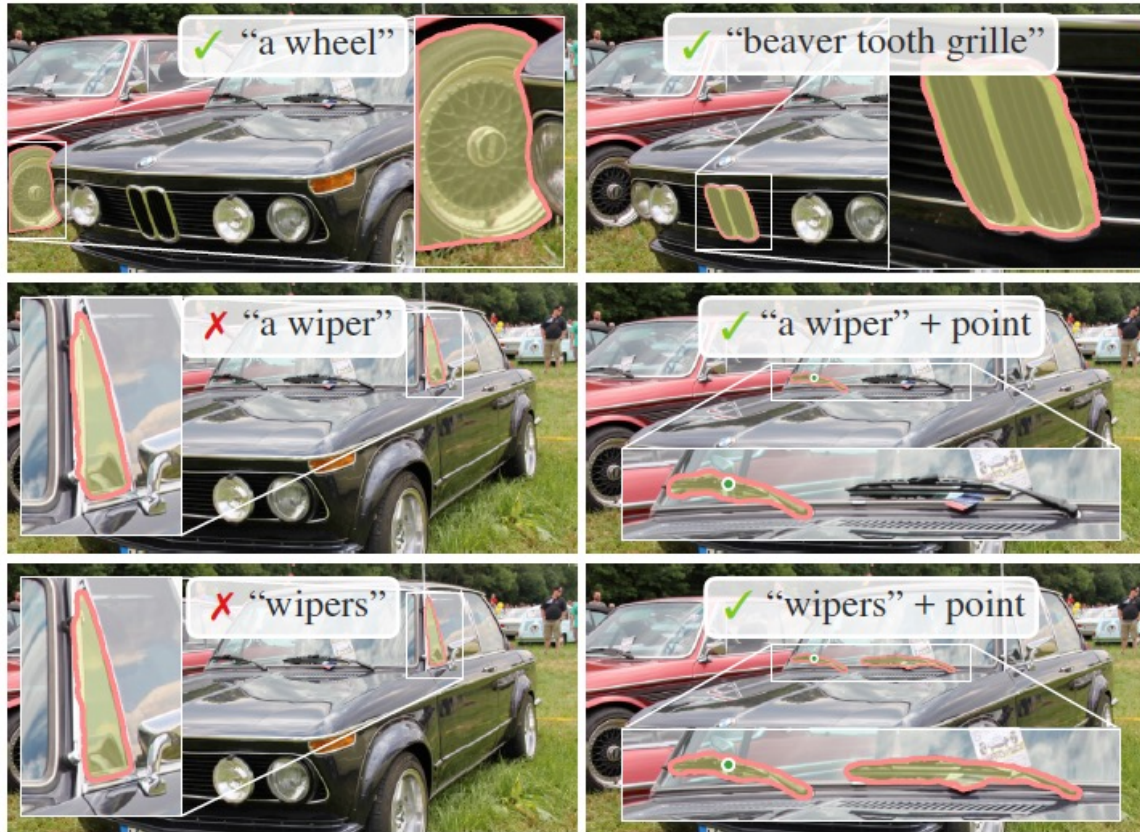


# Segment Anything

- Prompt engineering

## ❖ Prompt 변화에 따른 Mask 생성 결과

- Prompt를 정확하게 입력해준다면 원하는 객체에 대해 정확히 인식할 수 있음
- ChatGPT, SAM을 통해 최근 각광 받고 있는 연구 분야 중 하나인 Prompt engineering 설명



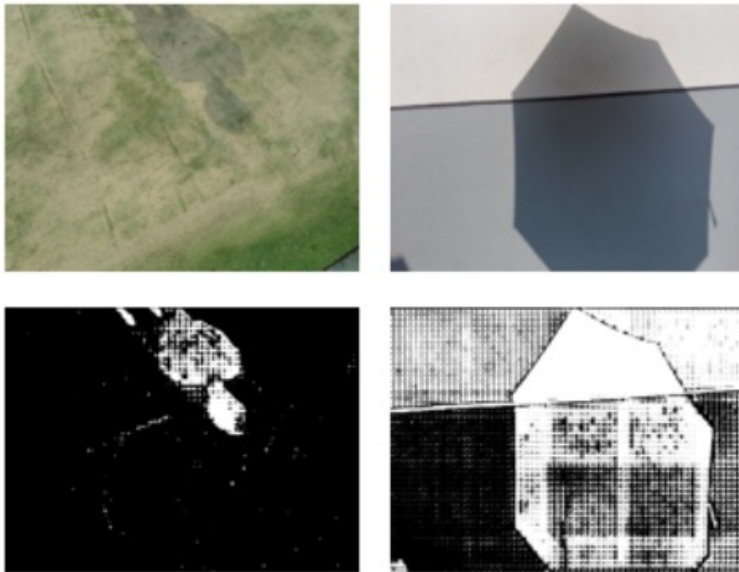
# Segment Anything

- Shadow Detection and Camouflaged Object Detection

## ❖ 탐지하기 어려운 객체에 대한 SAM 예측 결과

- 배경과 관심 객체 사이 구별이 어려운 두 분야에 대한 SAM 예측 결과
  - Shadow Detection: 그림자 위치를 정확하게 인식
  - Camouflaged Object Detection: 배경과 관심 객체 사이 구분이 어려운 상황에 객체 인식
- SAM은 정확하게 인식을 하지 못하고 있는 상황

### Shadow Detection



### Camouflaged Object Detection



# 목차

---

- Introduction
- Segment Anything
- **SAM-Adapter**
- Conclusion



# SAM-Adapter

- Segment Anything(2023, arXiv)

## ❖ SAM-Adapter(2023 ICCV – 3회 인용, arXiv – 34회 인용)

- KOKONI 3D 라는 회사를 중심으로 진행된 SAM 기반 전이학습 연구 내용 제안
- 전이 학습 시 성능 효과를 위한 Adapter를 추가해 성능 향상 도모
  - Shadow Detection, Camouflaged Object Detection을 위해 미세 조정 진행

## SAM-Adapter: Adapting Segment Anything in Underperformed Scenes

Tianrun Chen<sup>1, 2, +</sup> Lanyun Zhu<sup>4, +</sup> Chaotao Ding<sup>3, +</sup> Runlong Cao<sup>3+</sup> Yan Wang<sup>5</sup>  
Shangzhan Zhang<sup>1</sup> Zejian Li<sup>1</sup> Lingyun Sun<sup>1</sup> Ying Zang<sup>3, \*</sup> Papa Mao<sup>2</sup>

Zhejiang University<sup>1</sup> KOKONI, Moxin (Huzhou) Technology<sup>2</sup>  
Huzhou University<sup>3</sup> Singapore University of Technology and Design<sup>4</sup> Beihang University<sup>5</sup>  
{tianrun.chen, zhang3z, zejianlee, sunly}@zju.edu.cn lanyun\_zhu@mymail.sutd.edu.sg  
2021388117@stu.zjhu.edu.cn crl1657@163.com  
wangyan9509@gmail.com info@kokoni3d.com 02750@zjhu.edu.cn

Project Page: <http://research.kokoni3d.com/sam-adapter>

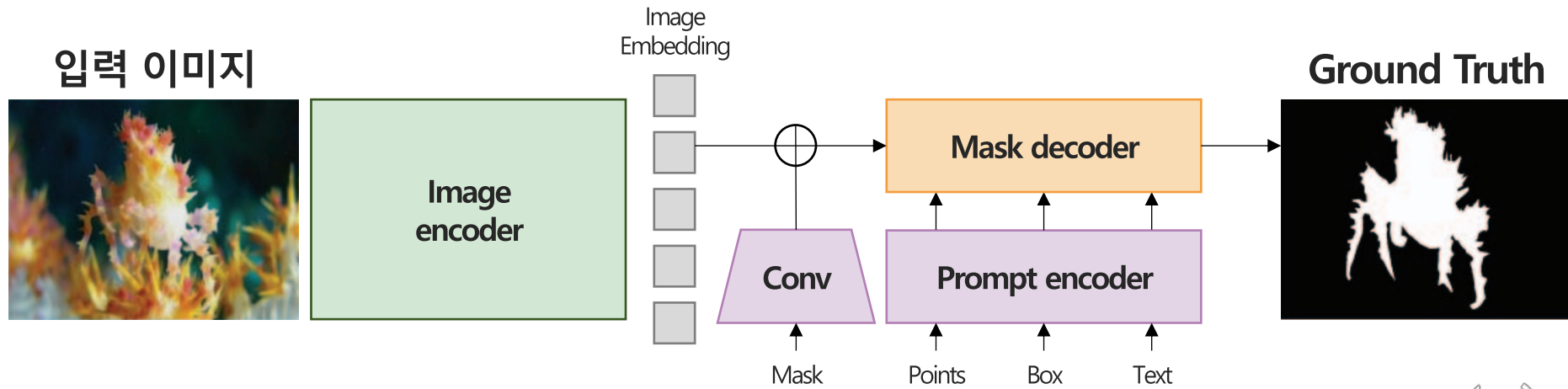


# SAM-Adapter

- 기존 SAM 모델 구조

## ❖ 다른 문제를 위한 사전 학습된 SAM 가중치 전이학습

- SAM은 Image encoder, Prompt encoder, Mask decoder로 구성
- SAM-Adapter에서는 Prompt를 입력 받는 상황을 고려 X
- 기존에 있던 Prompt encoder는 제거하고 전이 학습 진행

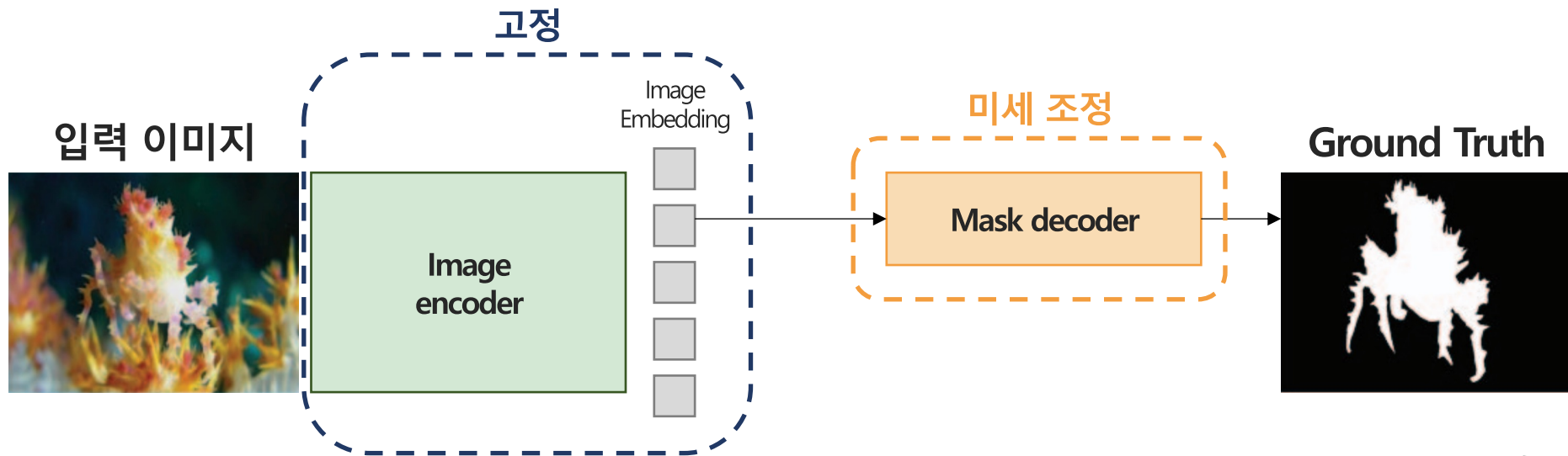


# SAM-Adapter

- Fine-tuning for Mask decoder

## ❖ Encoder 부분은 고정하고 Mask decoder만 재학습

- SAM 내 Encoder 부분은 엄청나게 많은 데이터를 사용해 학습된 Encoder
- 따라서 Encoder 내 가중치는 **고정(Frozen)**한 후 Mask decoder 만 **미세 조정**
  - 배경과 유사하지만 관심 객체 내 일부분인 영역을 정확히 인식하기 위한 미세 조정
- '입력 이미지 내 미세한 패턴에 대한 특징 어떻게 추출하지?' 라는 의문 발생

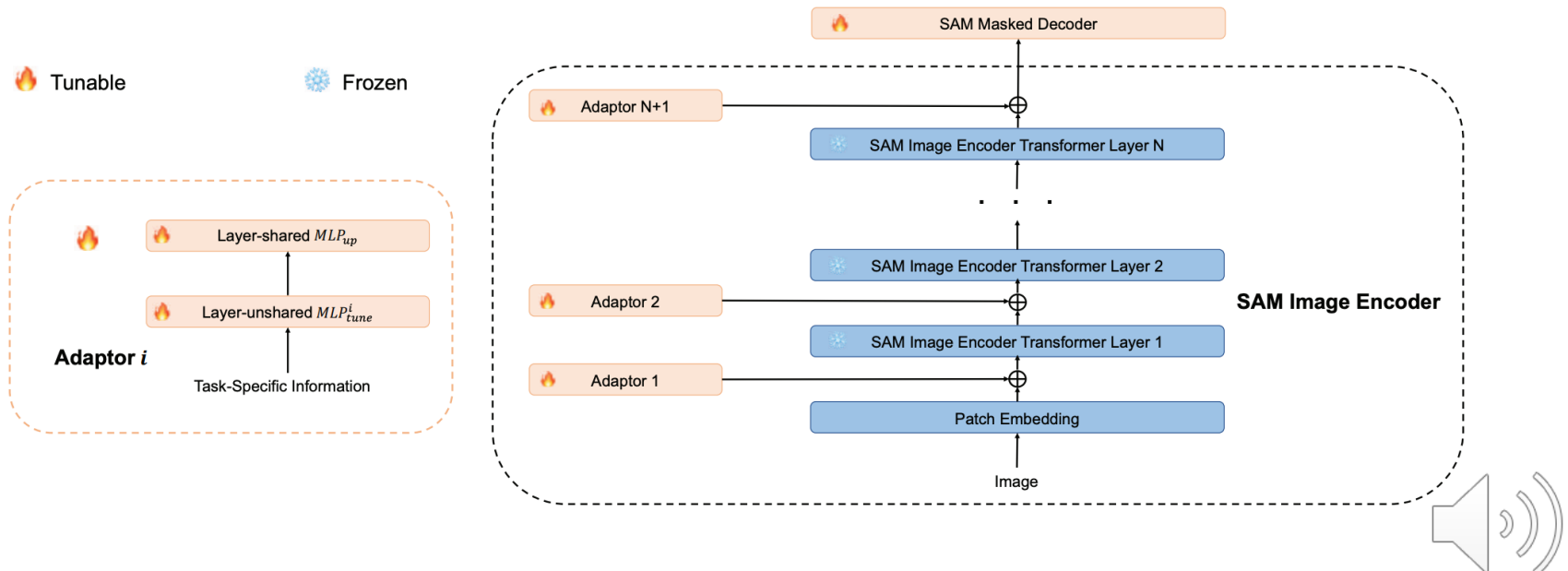


# SAM-Adapter

- Task-Specific Information Injection

## ❖ 특정 데이터 셋에 알맞은 특징(Task-Specific)을 추출하기 위한 Adaptor

- Transformer 블록 연산 결과와 Task-Specific 정보를 결합하여 미세 조정 진행
- 기존 Transformer 내 가중치는 고정되고 Adaptor내부에 대해서만 학습 진행
  - Adaptor 는 Task 정보를 학습하는  $MLP_{tune}^i$ 와 Upsampling을 위한  $MLP_{up}$ 로 구성

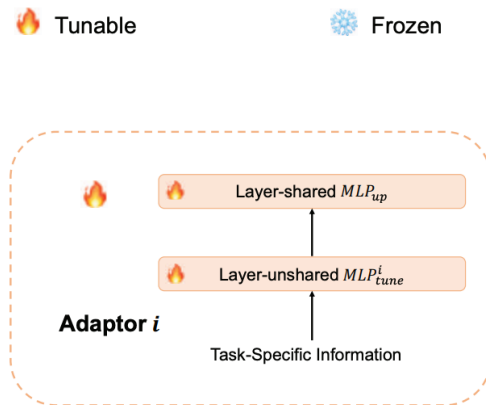


# SAM-Adapter

- Task-Specific Information Injection

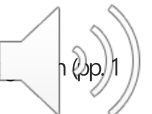
## ❖ Adaptor에 대한 입력 데이터

- 서로 다른 두 MLP와 GELU 활성화 함수로 특정 Adaptor 를 구성
- Adaptor 사용을 통해 배경과 배경과 유사한 관심 객체를 분리할 수 있을 것이라는 기대



$$P^i = MLP_{up}(GELU(MLP_{tune}^i(F_i)))$$

- $P^i$ :  $i^{th}$  Adaptor /  $F_i = F_{pe} + F_{hfc}$
- $F_{pe}$ : Patch embedding /  $F_{hfc}$ : high frequency components from FFT
- $F \rightarrow [F \text{ (FFT)}]$



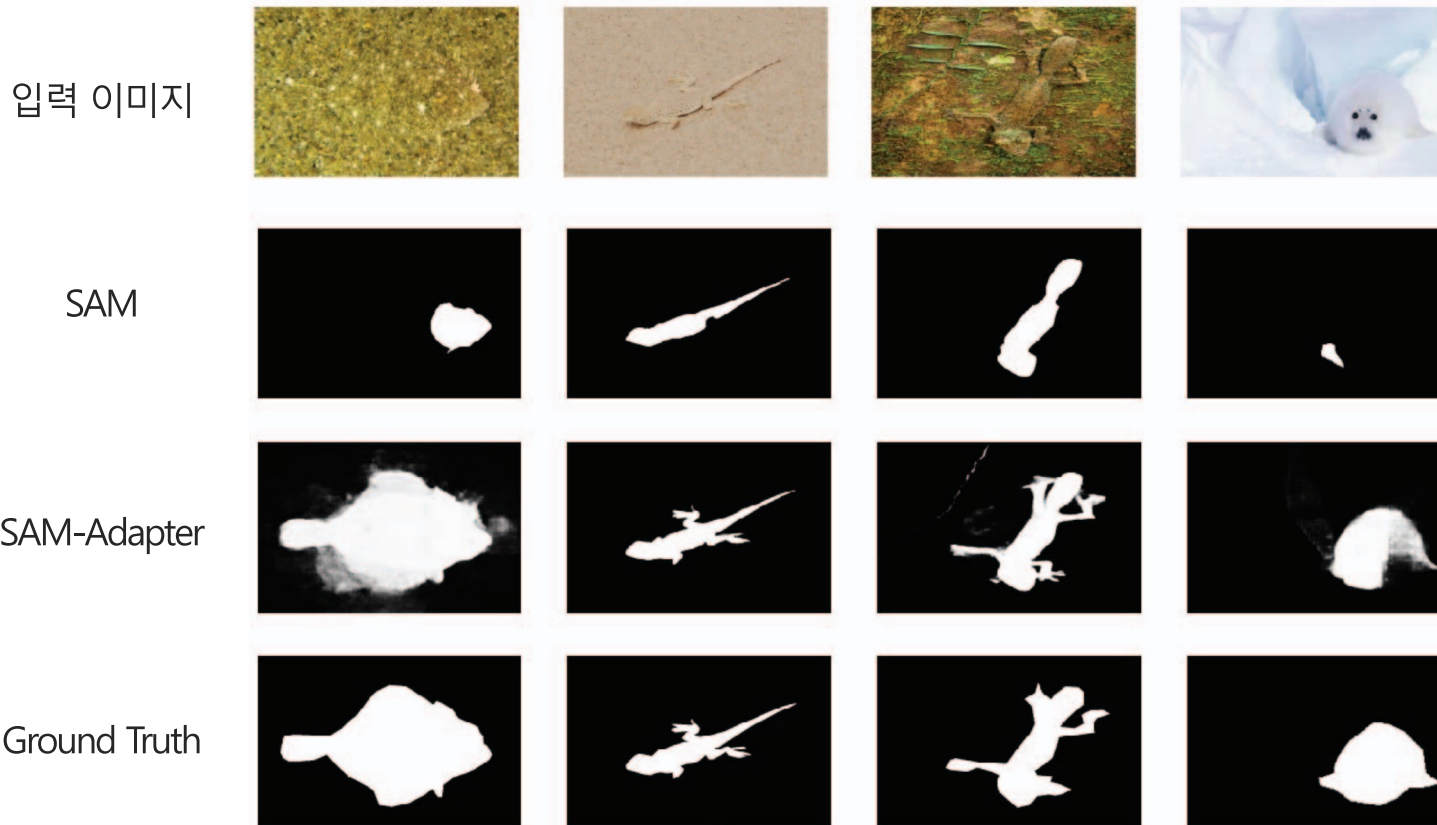


# SAM-Adapter

- Camouflaged Image Segmentation

## ❖ SAM과 SAM-Adapter 사이 예측 결과 비교

- SAM는 정확하게 객체와 배경을 분리하지 못하는 상황
- Adapter 추가 및 Mask decoder 미세 조정을 통해 정확히 객체를 인식 가능

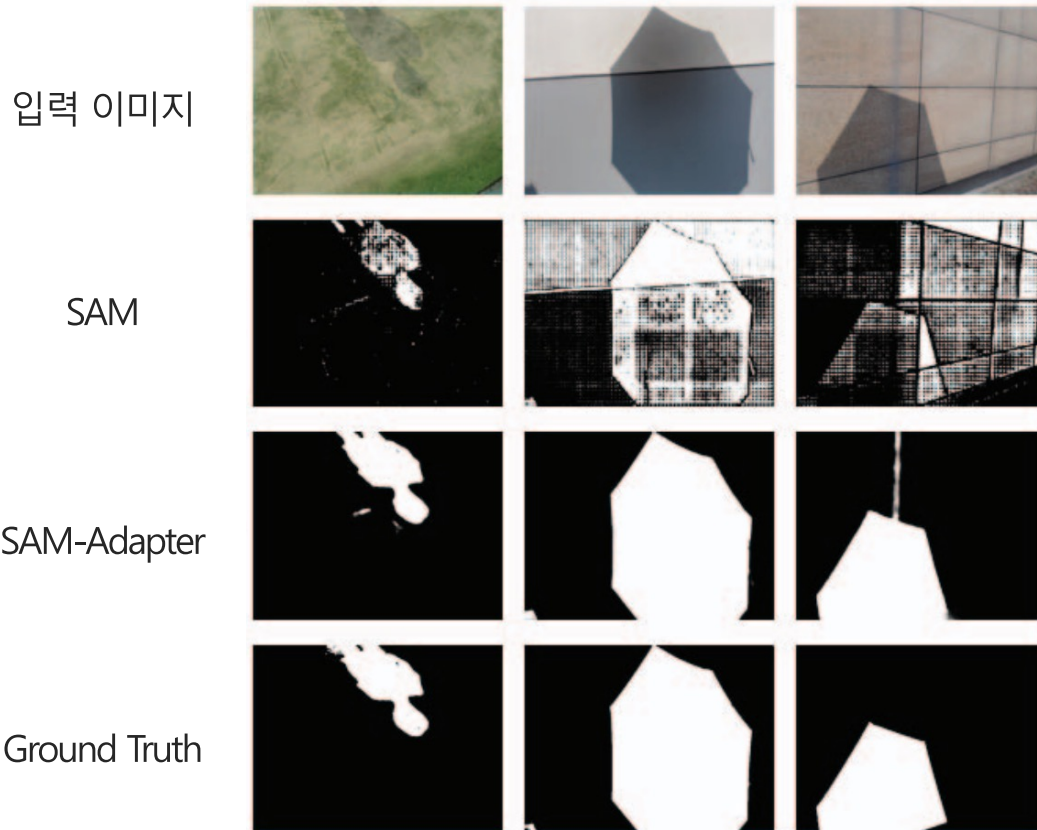


# SAM-Adapter

- Shadow Detection

## ❖ SAM과 SAM-Adapter 사이 예측 결과 비교

- SAM는 정확하게 객체와 배경을 분리하지 못하는 상황
- Adapter 추가 및 Mask decoder 미세 조정을 통해 정확히 객체를 인식 가능



# 목차

---

- Introduction
- Segment Anything
- SAM-Adapter
- Conclusion



# Conclusion

---

## ❖ SAM 요약

- ChatGPT에 영감을 받아 이미지 데이터를 위한 SAM 모델에 대한 소개
- Prompt Segmentation 문제를 정의하고 SAM 학습 진행
- 사전학습된 Image encoder와 여러 Prompt에 대한 encoder에서 특징을 추출
- 추출된 특징들을 결합해 원하는 객체에 대한 Mask 생성
- 학습을 위한 Mask 생성을 위한 세 가지 단계가 존재하며 대부분 Fully-automatic 방식으로 생성

## ❖ SAM-Adapter 요약

- 특정 문제를 위한 사전 학습된 SAM 모델 가중치에 대한 미세조정 연구
- Image encoder 내 Adaptor를 다수 추가하고 Adaptor와 Mask decoder 학습 진행
  - SAM 방식으로 학습된 Image encoder 내부 가중치는 학습 대상에서 제외



---

**고맙습니다.**