

2024 SAIL Seminar

## **SAM 2: Segment Anything in Images and Videos**

---

Nikhila Ravi et al.(Meta AI Research), arXiv 2024 (# of citation: 18)



순천향대학교 미래융합기술학과

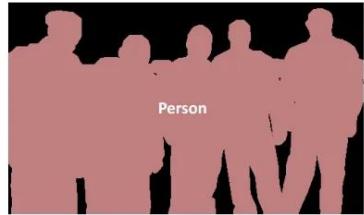
Senseable AI Lab

석사과정 김병훈

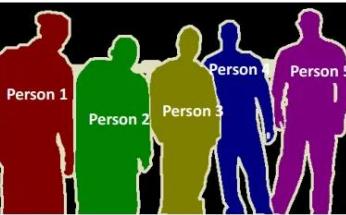
# Background & Intro

## Segmentation Task

- 배경과 다른 사물을 구분해 특정한 객체의 영역을 찾아내는 것
- 동일한 이미지도 때에 따라 **분할의 목적이 다를 수 있음**
- 사용자의 **의도를 파악**하기 위해 **Prompt**가 필요

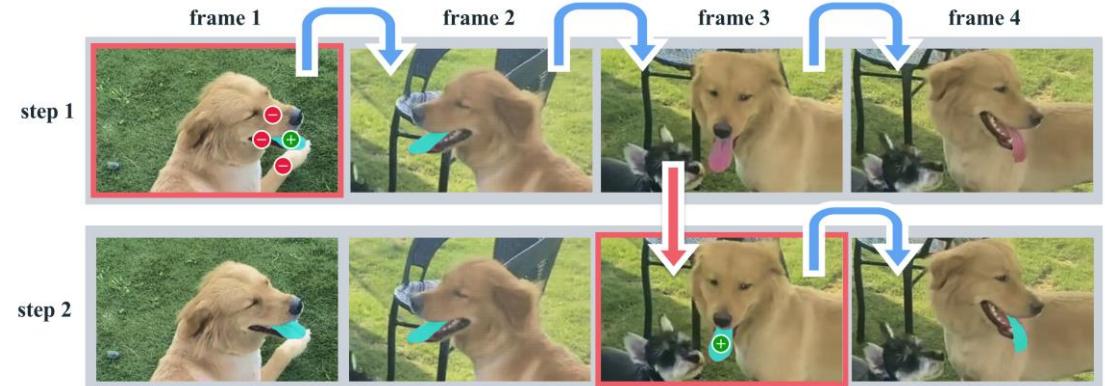


Semantic Segmentation



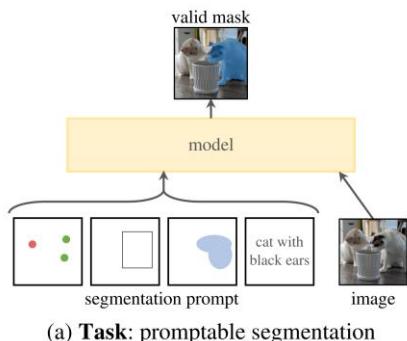
Instance Segmentation

- **Promptable Visual Segmentation (PVS) task** ※ masklet: 시공간 마스크
- 사용자와 **Interaction** 하여 prompt를 토대로 **\*masklet**을 예측
  - 비디오에서 어느 프레임에 대한 prompt를 제공하더라도 좋은 segmentation 결과를 보여주는 것을 목표



## Segment Anything Model 1(SAM 1)

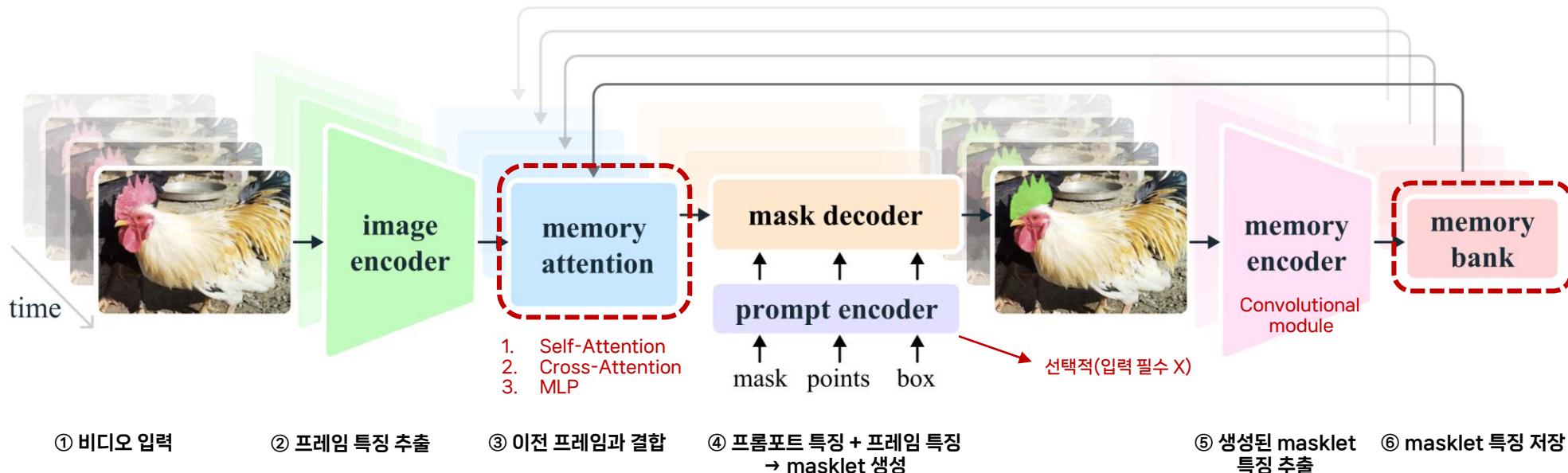
- 하나의 입력 값에 대해서 그 **입력을 포함하는 여러 대상**(전체(Whole), 부분(Part), 부분 내 영역(Subpart))을 예측하도록 함(Foundation Model 구축 목표)



→ **Video에 적용  
(SAM 2)**

시간의 흐름을 고려해야 함  
→ 어려운 Task

# Architecture

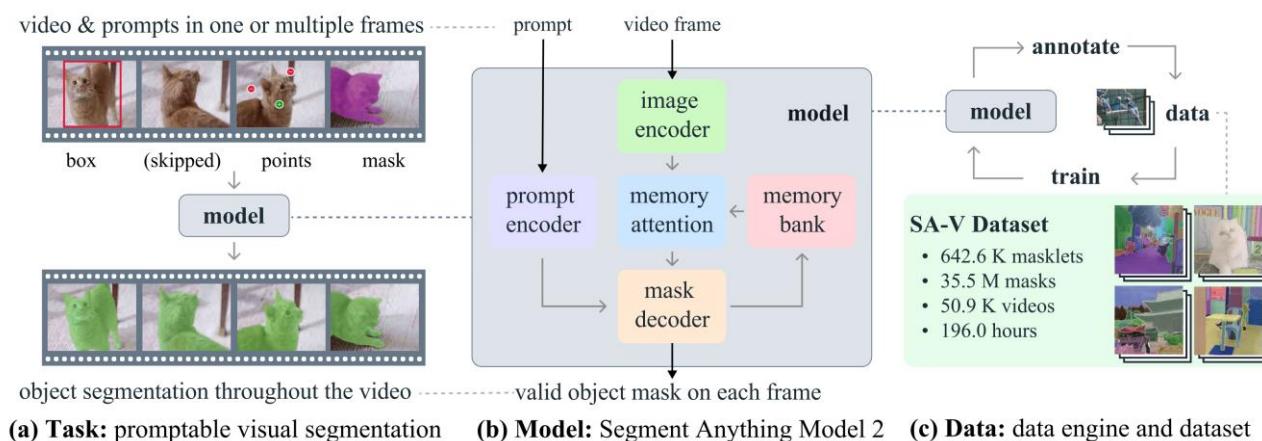


## Memory Attention

- 현재 프레임의 특징을 이전 프레임의 특징 및 예측 값과 결합

## Memory Bank

- 비디오에서 타깃 객체의 과거 예측 정보를 저장
- 저장 내용
  - Masklet feature map
  - Object pointer



# Application



- Demo: [Link](#)
- Paper: [Link](#)
- Github: [Link](#)
- Fine Tuning Tutorial: [Link](#)



 Meta