

Tobig's Conference

# Computer Vision

## 오늘의 방

### : 내 방 리뉴얼 프로젝트

Tobig's 22th  
강연주 송진하 장민지 조하늘 진민찬

# 소개 목차

---

01

## INTRODUCTION

- 주제 선정 배경
- 아이디어 소개
- 팀원 소개

02

## PIPELINE

- 파이프라인

03

## MODELING

- Grounded SAM2
- BLIP-2
- MiDaS(DPT)
- Web Crawling

04

## DEMO

- DEMO 시현

05

## CONCLUSION

- 결론
- 마무리 및 피드백

---

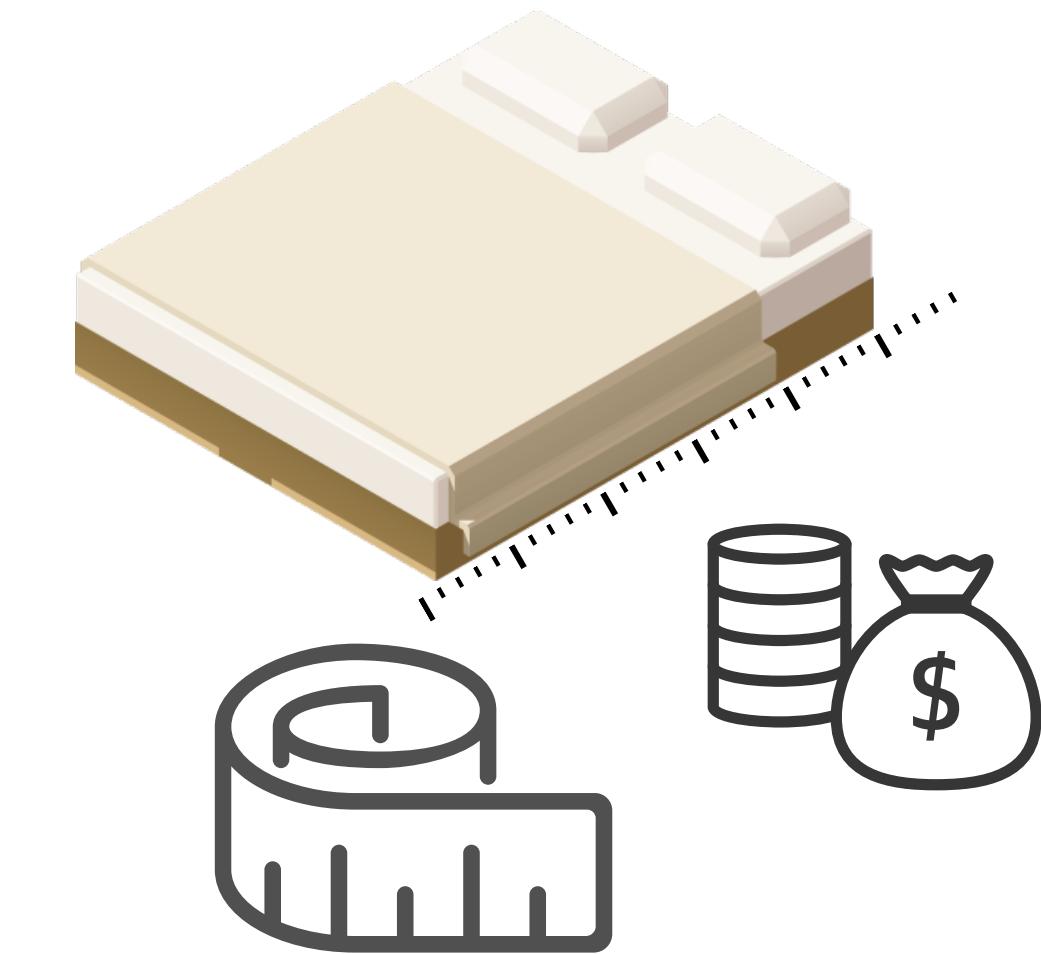
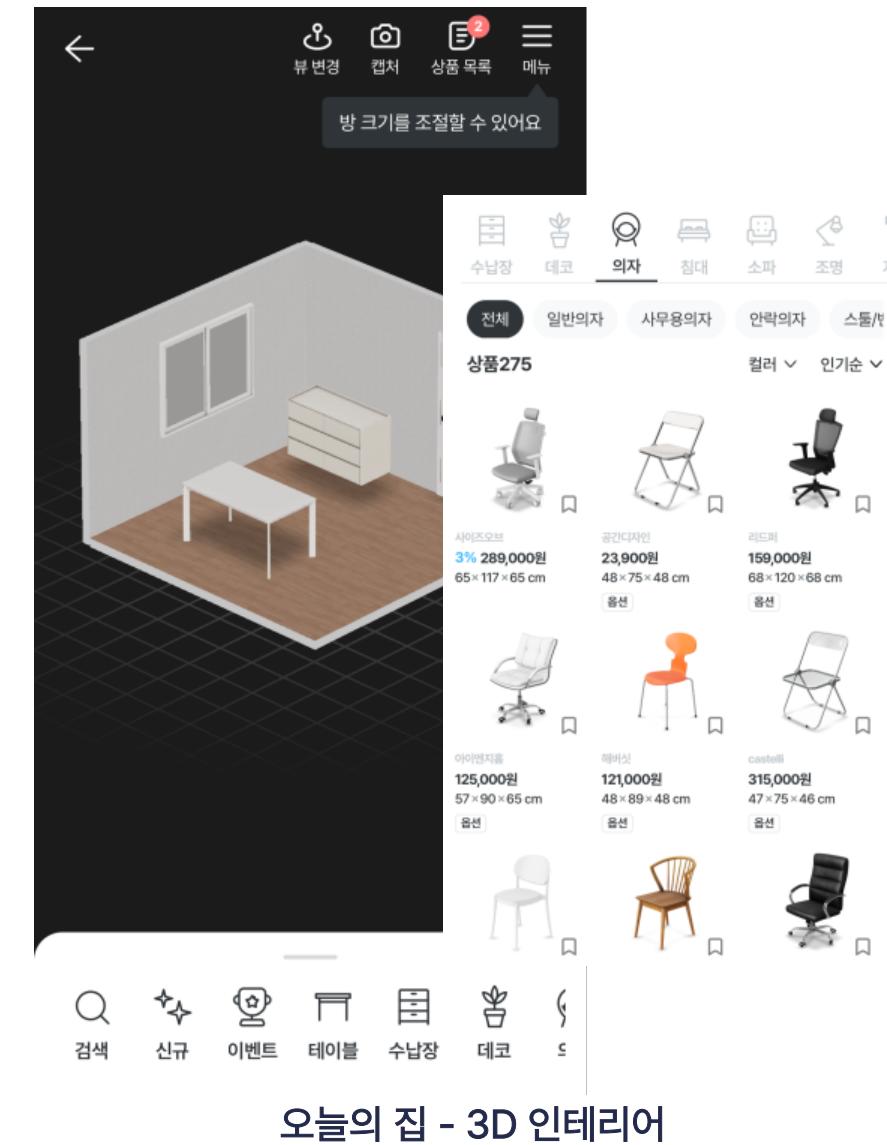
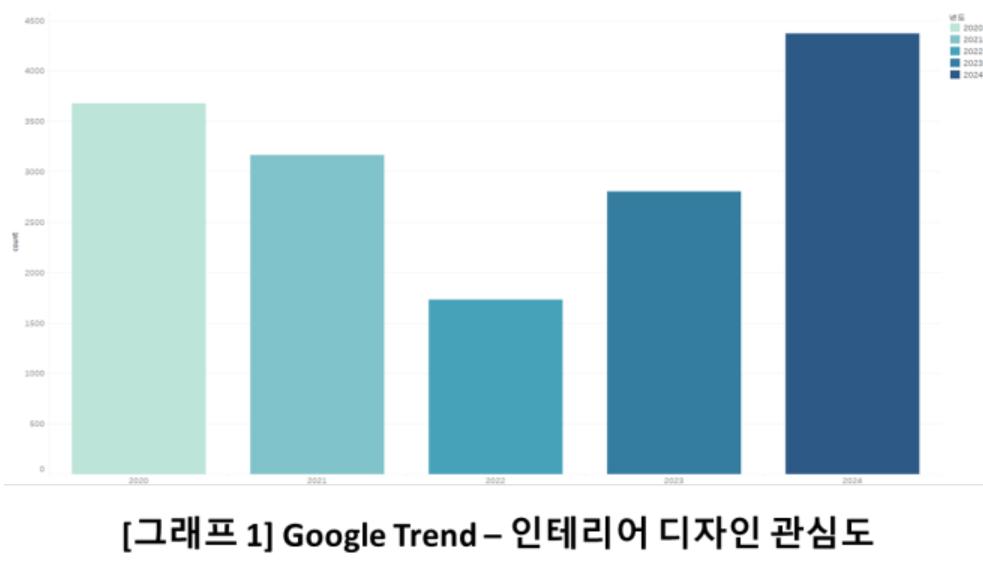
01

CHAPTER

# INTRODUCTION

## 01 Introduction

# 주제 선정 배경



# 인테리어 관심 증가

# 기존 플랫폼의 한계

# 치수 입력의 번거로움과 비용 문제

# 아이디어 소개



핸드폰으로 내 방 사진을  
찍어 사진 업로드



사진 속  
방 분위기 분석  
및  
크기 추정



컴퓨터 비전 & 추천 시스템  
: 내 방 분위기에 어울리는  
가구 대체 시스템 구축



## 01 Introduction

# 오늘의 방 팀원 소개

### 22기 강연주

*Seoul National University  
of Science and Technology*

Crawling

### 22기 송진하

*Seoul National University  
of Science and Technology*

Modeling

### 22기 장민지

*Ewha Womans University*

Crawling

### 22기 조하늘

*Sungshin Women's University*  
Modeling,  
Preparation  
for presentation

### 22기 진민찬

*Seoul National University  
of Science and Technology*  
Pipeline,  
Modeling

---

---

02

CHAPTER

# PIPELINE

# PIPELINE



---

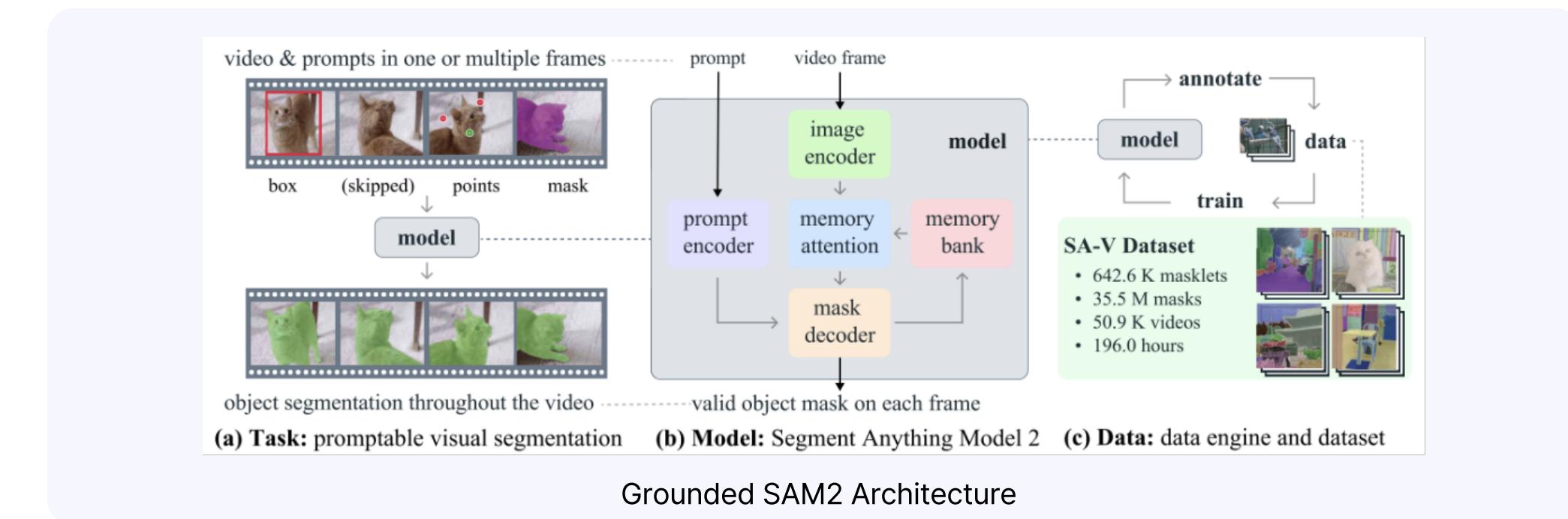
---

03

CHAPTER

# MODELING

# Grounded SAM2



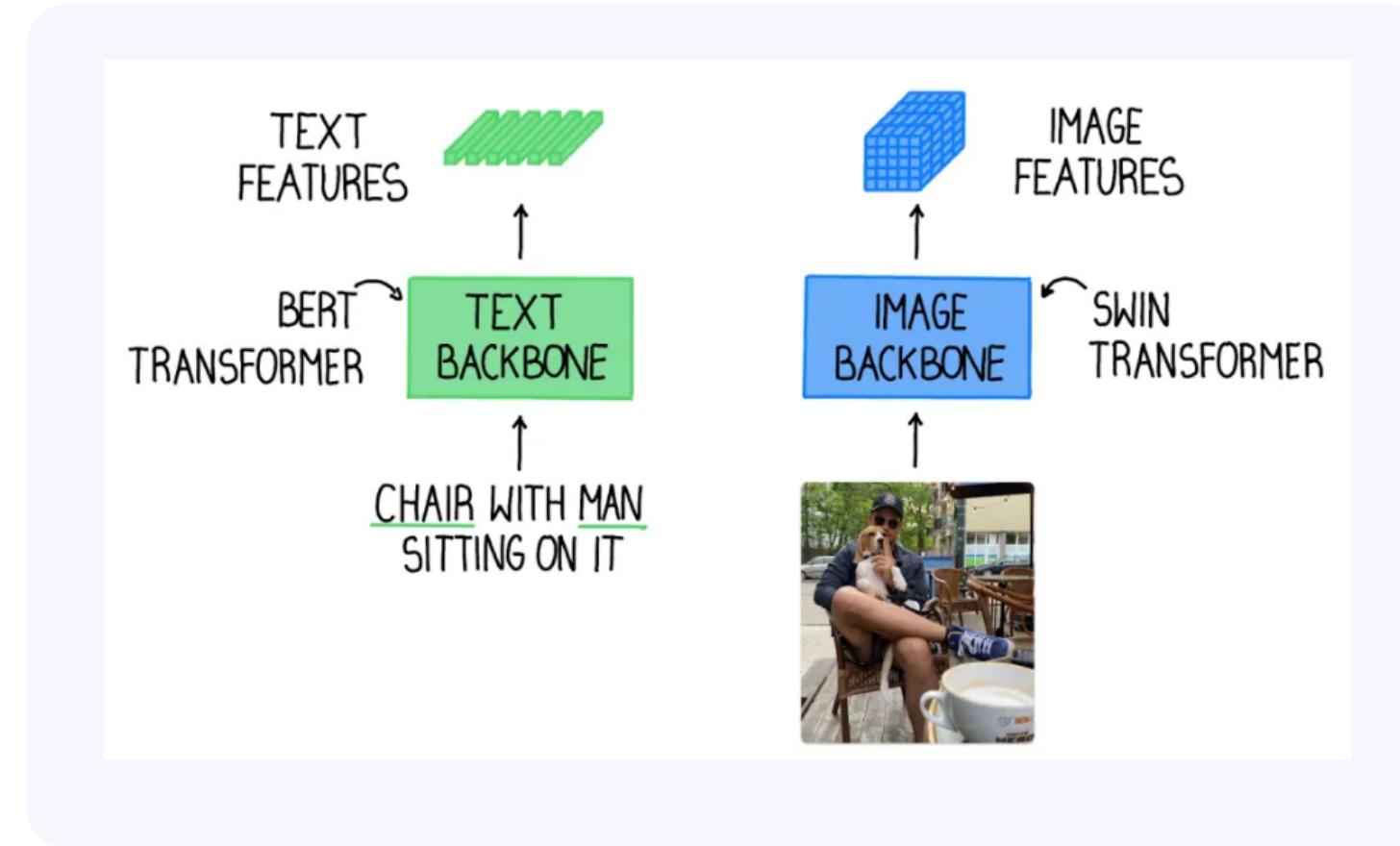
## 텍스트 기반 객체 탐지

- Grounding DINO를 사용하여 입력된 텍스트 프롬프트에 따라 객체 탐지.
- "소파. 책상. 의자."와 같은 간단한 텍스트 입력을 지원하며, 다양한 가구 및 물체를 식별

## 정확한 이미지 세분화

- SAM2 (Segment Anything Model)을 사용하여 탐지된 객체를 픽셀 단위로 세분화.
- Grounding DINO에서 생성된 경계 상자를 바탕으로 세분화 결과를 더욱 정교하게 함.

# Grounded SAM2



Grounding DINO Architecture

## DINO(Differentiable Image Noise Optimisation)란?

이미지와 텍스트 간의 관계를 학습하여 이미지 생성 및 텍스트를 설명할 수 있는 모델

### - Image Backbone:

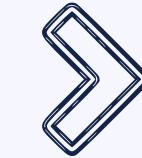
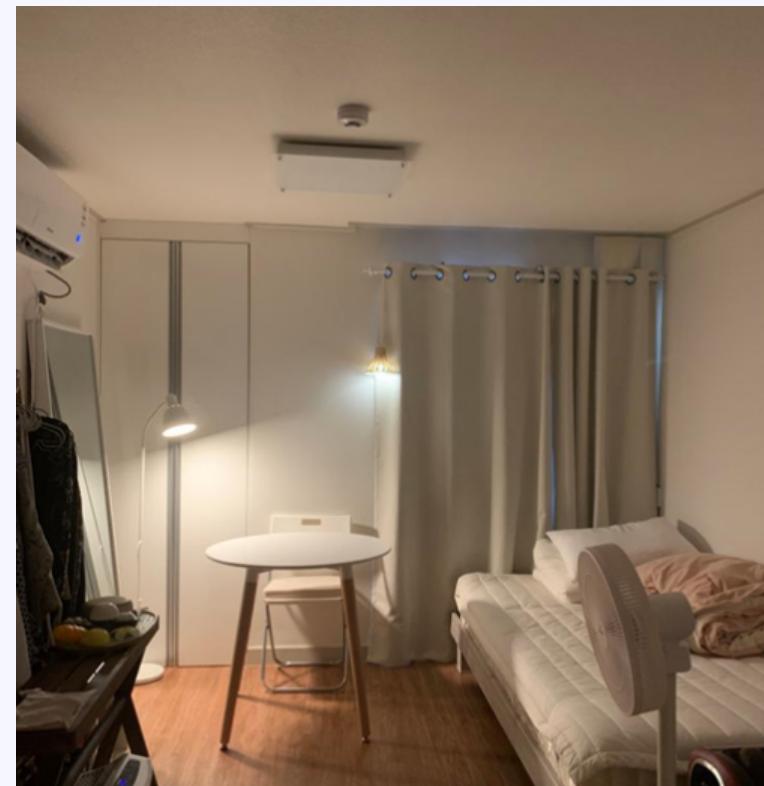
Swin Transformer(Multi-scale 이미지 feature extractor)로  
더욱 풍부한 이미지 feature 생성

### - Text Backbone:

BERT로 텍스트 feature 생성  
⇒ text는 text backbone에서 추출하고, image는 image backbone에서 각자 추출

# Grounded SAM2

이미지로부터 가구를 탐지(Detection) 및 세분화(Segmentation) 후 라벨링

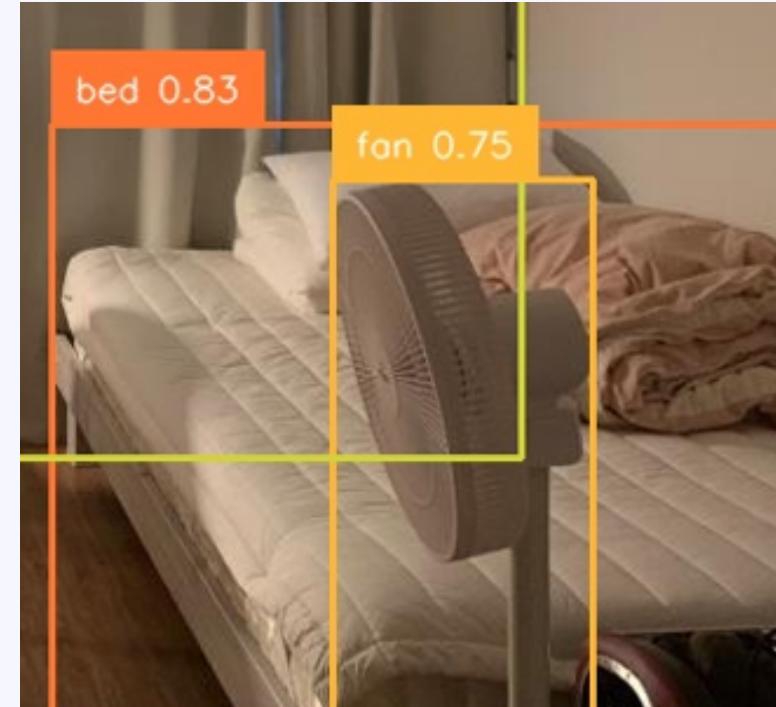


TEXT PROMPT를 기반으로  
객체의 Bounding Box 및 Segmentation을 예측하여 Json 형식으로 저장

TEXT\_PROMPT = "bed. chair. table. curtains. mirror. fan."

# Grounded SAM2

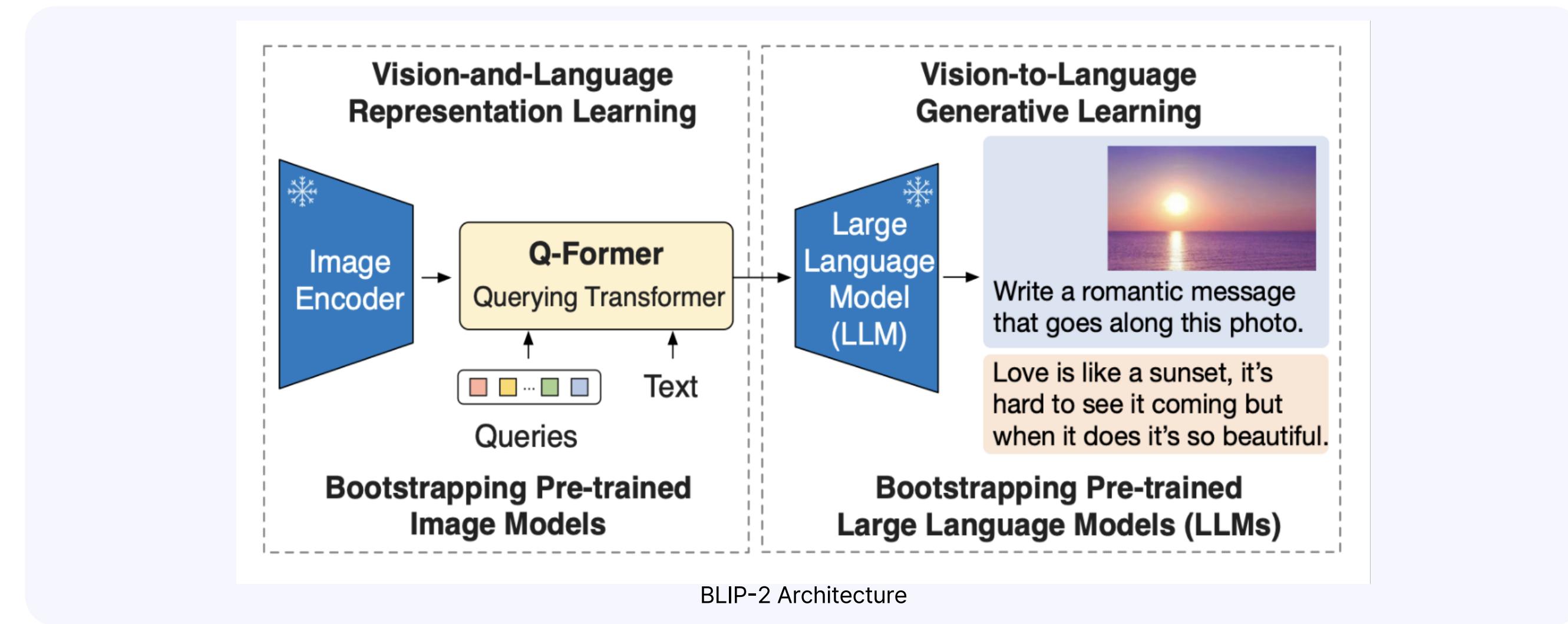
EX) BED



```
{  
    "class_name": "bed",  
    "bbox": [  
        415.91546630859375,  
        488.1807556152344,  
        749.6010131835938,  
        749.5123291015625  
    ],  
    "segmentation": {  
        "size": [  
            750,  
            750  
        ],  
        "score": [  
            0.84765625  
        ]  
    },  
    "box_format": "xyxy",  
    "img_width": 750,  
    "img_height": 750
```

- 객체의 Bounding Box를 예측.
- 해당 영역의 Segmentation Mask를 생성하여 Json 형식으로 저장.

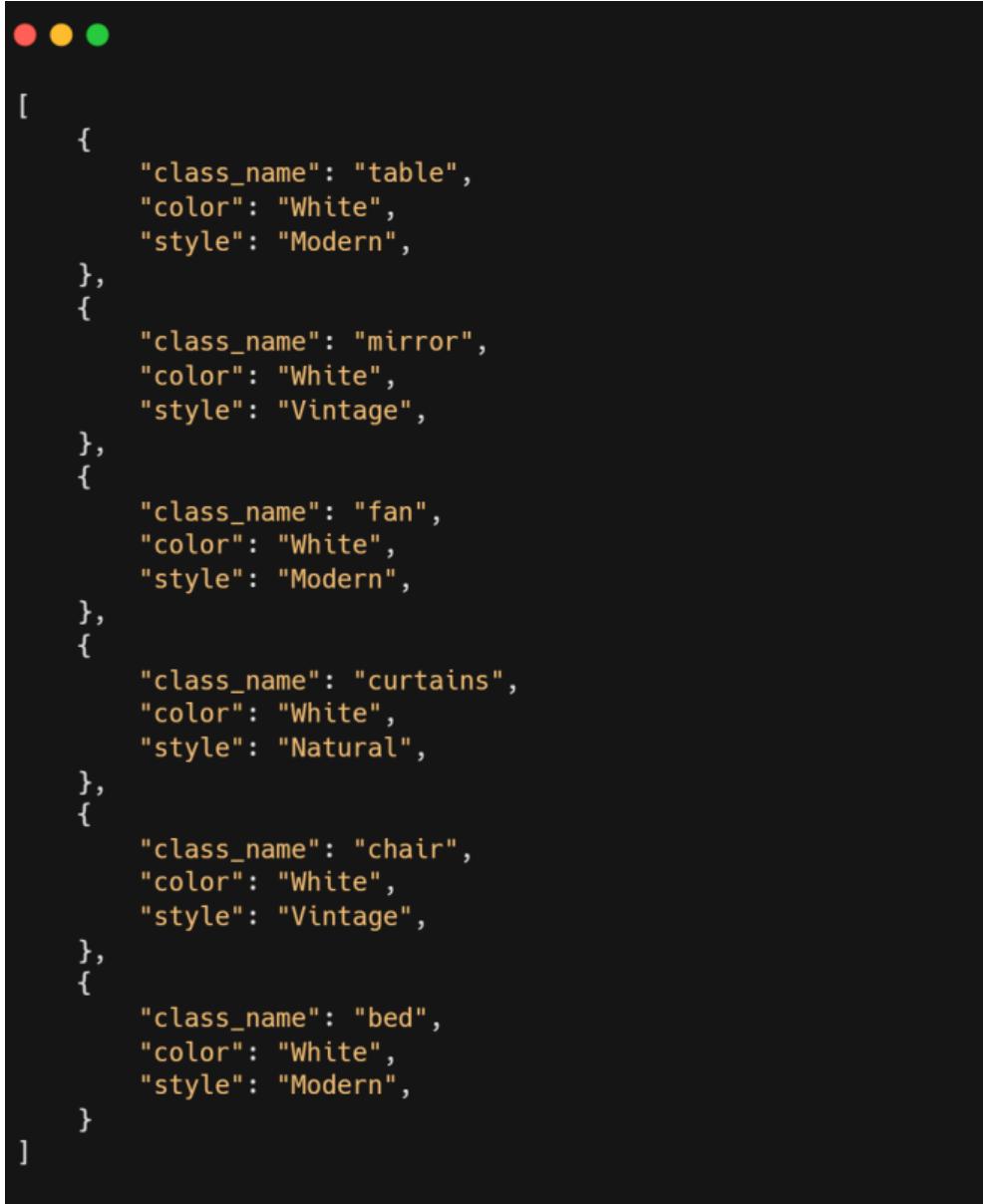
# BLIP-2



## Image-to-text generation with BLIP-2

: 고정된 이미지 인코더와 LLM을 효과적으로 결합하여 이미지와 텍스트 간의 격차를 해소하는 비전-언어 사전 학습 방법

# BLIP-2



```
[  
  {  
    "class_name": "table",  
    "color": "White",  
    "style": "Modern",  
  },  
  {  
    "class_name": "mirror",  
    "color": "White",  
    "style": "Vintage",  
  },  
  {  
    "class_name": "fan",  
    "color": "White",  
    "style": "Modern",  
  },  
  {  
    "class_name": "curtains",  
    "color": "White",  
    "style": "Natural",  
  },  
  {  
    "class_name": "chair",  
    "color": "White",  
    "style": "Vintage",  
  },  
  {  
    "class_name": "bed",  
    "color": "White",  
    "style": "Modern",  
  }]
```

## ➡ STYLE 정보 사용

- 가구의 이미지로부터 Style 추출.
  - 각 객체의 Bounding Box 영역을 잘라내어 가구 이미지 패치를 생성 후 BLIP-2 모델을 사용해 Image Captioning을 수행하여 객체의 스타일을 포함한 설명을 생성.
    - 입력 이미지: 가구의 Crop된 이미지.
    - 입력 프롬프트: Describe the style of the {class\_name}.
- 출력된 가구 설명으로부터 Style에 맞는 단어만 추출하여 저장.

# MiDaS(DPT)

## 1 MiDaS를 활용한 상대 깊이맵 추출

입력 이미지에서 픽셀 간 상대적인 깊이 정보를 생성.

## 2 핀홀 카메라 모델을 활용한 물체까지의 거리 측정

상대 깊이맵과 거리 정보를 결합하여  
실제 스케일의 깊이맵을 생성.

## 3 이미지 전체의 실제 깊이 맵 생성

카메라의 내부 파라미터를 사용해  
특정 물체까지의 절대 거리를 계산.

## 4 가구 크기 추정

실제 깊이맵을 기반으로  
가구의 크기와 거리 정보를 계산하여 추정.

# MiDaS(DPT)

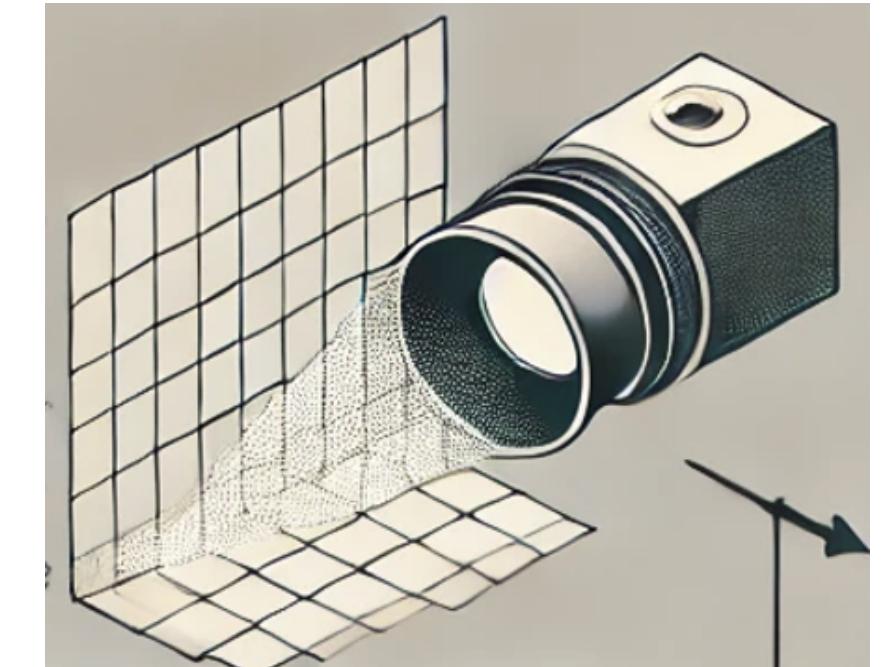
## (1) MiDaS를 활용한 상대 깊이맵 추출

### MiDaS?

깊이 추정에 사용되는 모델로 다양한 이미지에서 Depth MAP을 추론.

### [상대 깊이 맵 원리]

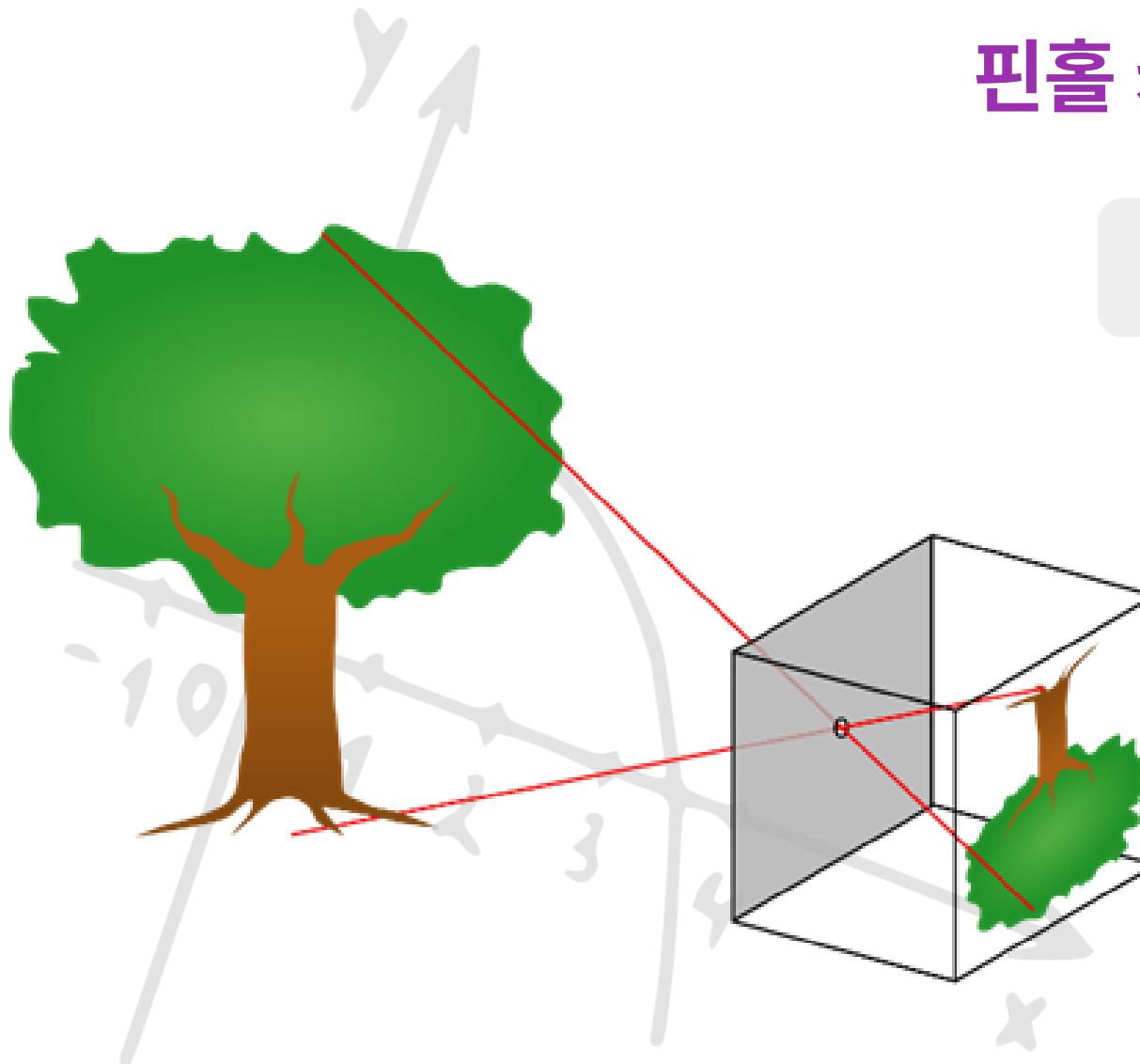
- 이미지 입력: 단일 RGB 이미지.
- 특징 추출: Transformer 기반의 모델(DPT)을 사용해 이미지에서 다양한 수준의 특징을 추출.
  - 저수준 특징(엣지, 텍스처 등)
  - 고수준 특징(구조적, 물체 간 관계 등)
- 깊이 예측: 픽셀 간의 상대적인 깊이를 계산하여 깊이 맵을 출력.
  - 따라서, 화면에서 더 가까운 물체는 더 밝게, 더 먼 물체는 더 어둡게 표현.



# MiDaS(DPT)

## (2) 핀홀 카메라 모델을 활용한 물체까지의 거리 측정

### 핀홀 카메라 모델(Pinhole Camera Model)



$$Z = (\text{focal\_length} * \text{실제 크기} * \text{이미지 해상도}) / (\text{픽셀 크기} * \text{센서 크기})$$

- Z: 물체까지의 거리(깊이).
- focal\_length: 카메라의 초점 거리(센서 기준, mm 단위).
- 실제 크기: 물체의 실제 크기(현실 세계에서, mm 단위).
- 픽셀 크기: 이미지에서 측정된 물체의 크기(픽셀 단위).
- 센서 크기: 카메라 센서 크기(가로 또는 세로, mm 단위).
- 이미지 해상도: 이미지의 가로 또는 세로 픽셀 해상도.

# MiDaS(DPT)

## (2) 핀홀 카메라 모델을 활용한 물체까지의 거리 측정

### EXIF 데이터란?

스마트폰으로 찍은 사진 안에 들어있는 스마트폰 카메라에 대한 정보(초점거리, 해상도 등)

### EXIF 데이터가 필요한 이유

#### EXIF 정보 추출

스마트폰 기종, 카메라의 종류가 다르므로 초점 거리, 이미지 해상도, 센서 크기가 달라짐.  
따라서 고정된 값이 아닌 기기마다 다른 값이 필요.

### EXIF 데이터의 사용

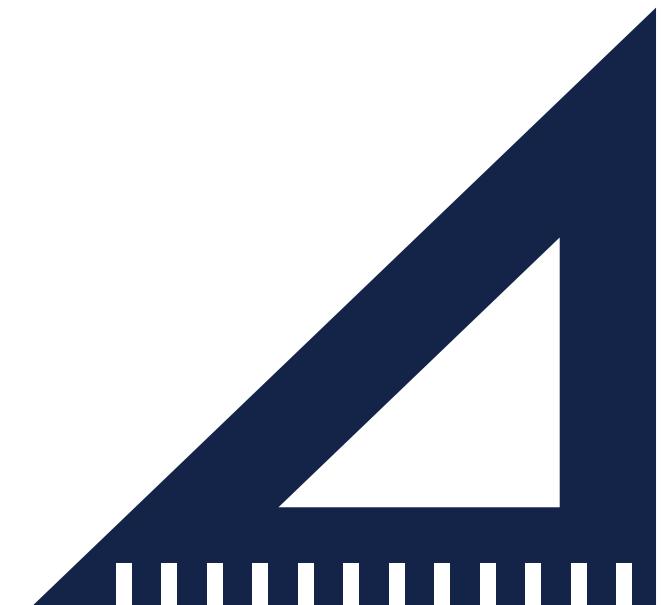
핀홀 카메라 공식에서 물체 거리 계산할 때 초점거리, 이미지 해상도, 센서 크기가 필요.

# MiDaS(DPT)

## (3) 이미지 전체의 실제 깊이 맵 생성

- 상대 깊이맵에서 계산된 절대 거리 값을 기준으로 이미지를 스케일링하여 전체 이미지의 절대 깊이맵(실제 환경의 깊이 정보를 반영한 맵)을 생성.

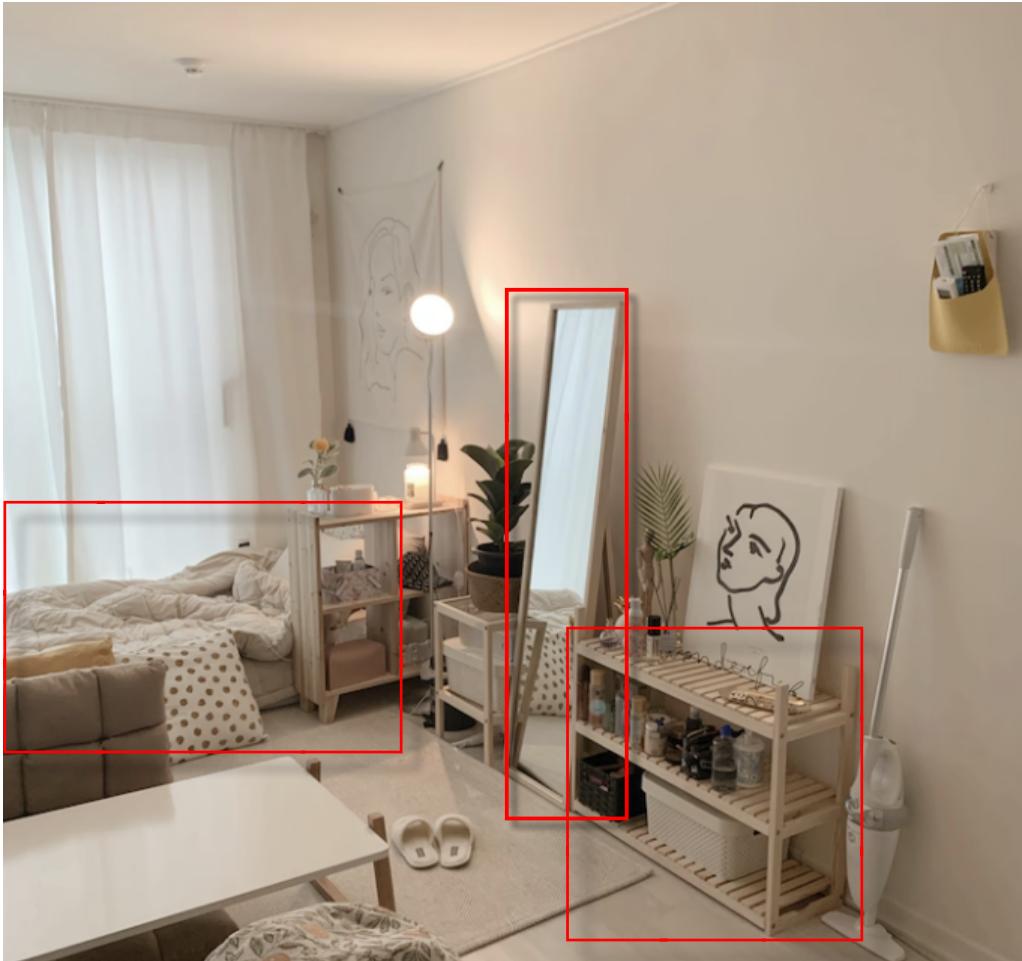
=> 이를 통해 이미지의 모든 픽셀에 대해 실제 거리를 추정



# MiDaS(DPT)

## (4) 가구 크기 추정

기존



현 이미지에 대해 기울임없이 bbox 생성

RotatedREct



Contour(윤곽선)와 RotatedRect(기울어진 사각형)을 사용하여  
기울어진 물체에 대해서도 비교적 정확한 측정이 가능

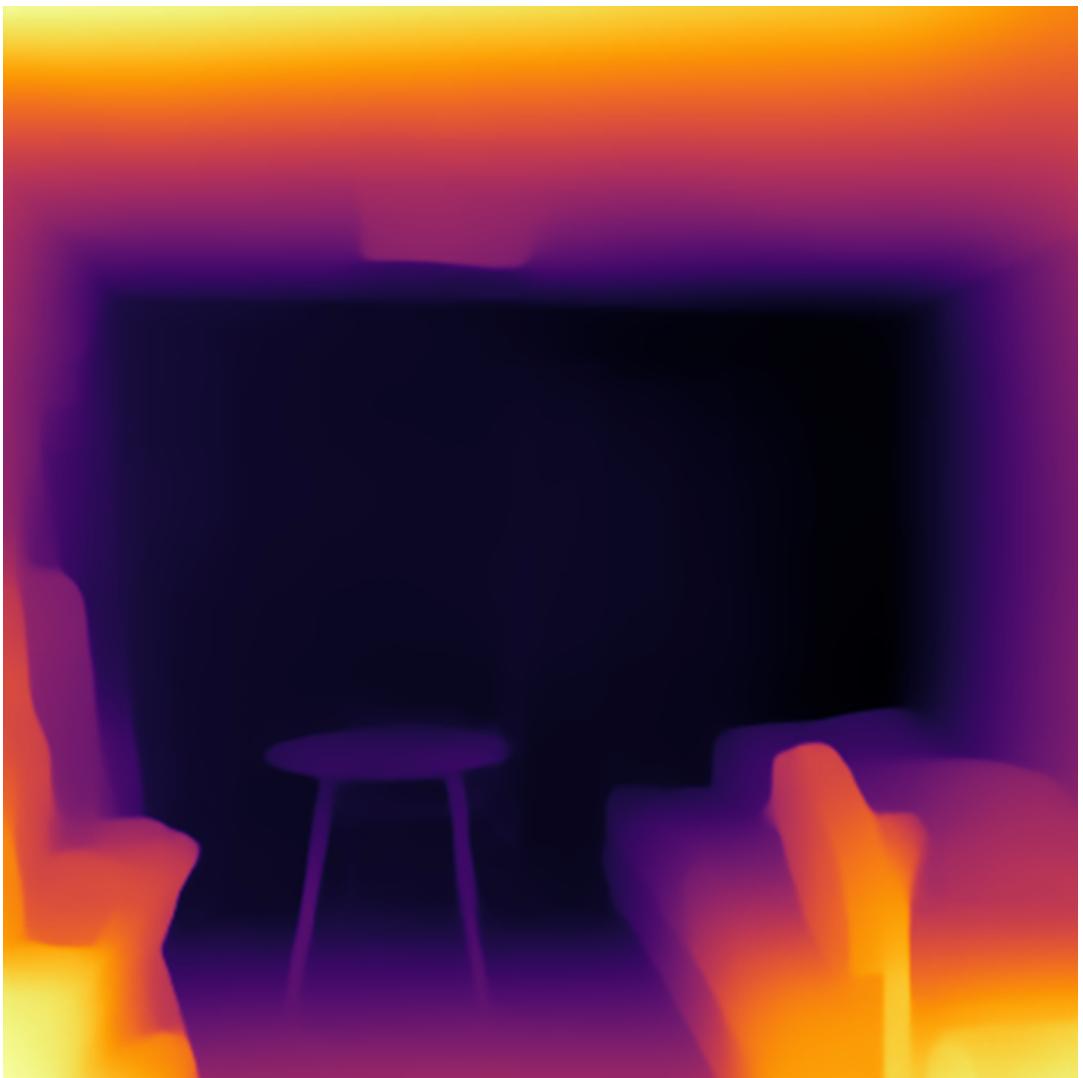
### 03 Modeling

# MiDaS(DPT)

## 실제 결과



원본 이미지



깊이 추정 이미지

```
[  
  {  
    "class_name": "table",  
    "color": "White",  
    "style": "Modern",  
    "width_mm": 538.27,  
    "depth_mm": 1614.26  
  },  
  {  
    "class_name": "mirror",  
    "color": "White",  
    "style": "Vintage",  
    "width_mm": 273.24,  
    "depth_mm": 2088.15  
  },  
  {  
    "class_name": "fan",  
    "color": "White",  
    "style": "Modern",  
    "width_mm": 340.38,  
    "depth_mm": 1812.81  
  },  
  {  
    "class_name": "curtains",  
    "color": "White",  
    "style": "Natural",  
    "width_mm": 851.58,  
    "depth_mm": 2984.41  
  },  
  {  
    "class_name": "chair",  
    "color": "White",  
    "style": "Vintage",  
    "width_mm": 263.05,  
    "depth_mm": 1319.13  
  },  
  {  
    "class_name": "bed",  
    "color": "White",  
    "style": "Modern",  
    "width_mm": 1000,  
    "depth_mm": 2000  
  }]
```

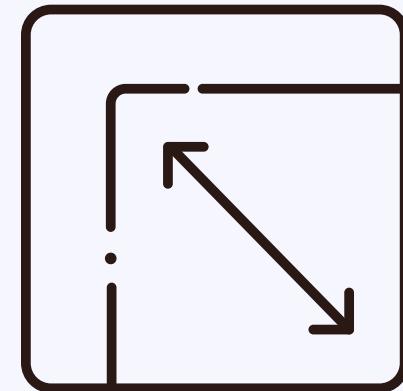
Json 파일

# WEB CRAWLING



## Selenium

자바스크립트로 동적으로 렌더링 되는 웹페이지에서도  
데이터를 크롤링할 수 있음



## 사이즈 별 카테고리 선택

데이터 세부 정부 크롤링에 사이즈 정보를 크롤링 하지  
않아도 되는 편리성



## 매 step마다 실행완료/ 에러 print 삽입

코드 에러 발생 지점을 탐색하기 수월

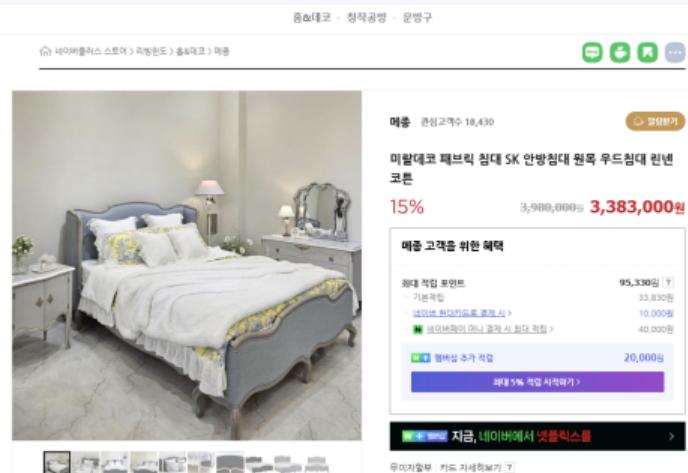
⇒ 데이터 크롤링 후 엑셀 파일로 저장

## 03 Modeling

# WEB CRAWLING

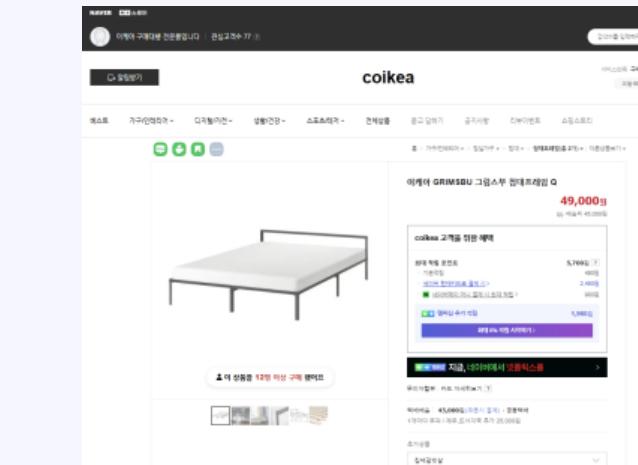
### 데이터 수집 기준

#### 1단계: 분위기 + 침대(+프레임)



ex. 내추럴 침대, 모던 침대 프레임

#### 2단계(1단계 검색 결과가 일관성이 없을 때) :



분위기 추출 모델에서 결과로 도출된 색상 + 침대(+ 프레임)

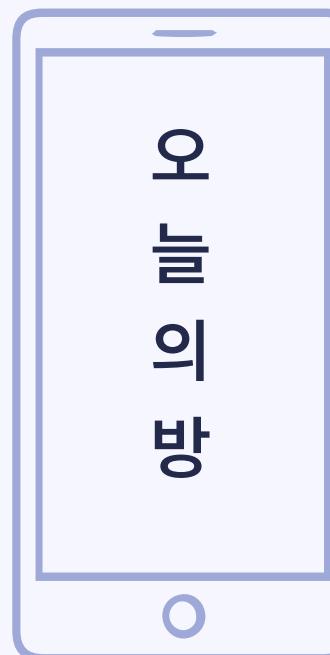
ex. 모던 침대 프레임 → 그레이 침대 프레임  
러블리 침대 프레임 → 핑크 침대 프레임

# WEB CRAWLING

## 가구 추천

원하는 스타일 입력

"내 방을 내추럴로 바꿔줘"



기존 방의 가구 정보 분석



bed | modern | 2010.12mm | 1058.15mm



원하는 스타일의 링크와 가격을 추천

BED

- Title: 라인드  
상품 링크 - Cost: 759000

비슷한 사이즈 추천

사이즈

SS W1220 x D2220 x H1210  
깔판 W1110 x D2050 x H270

내추럴 스타일 추천



---

---

04

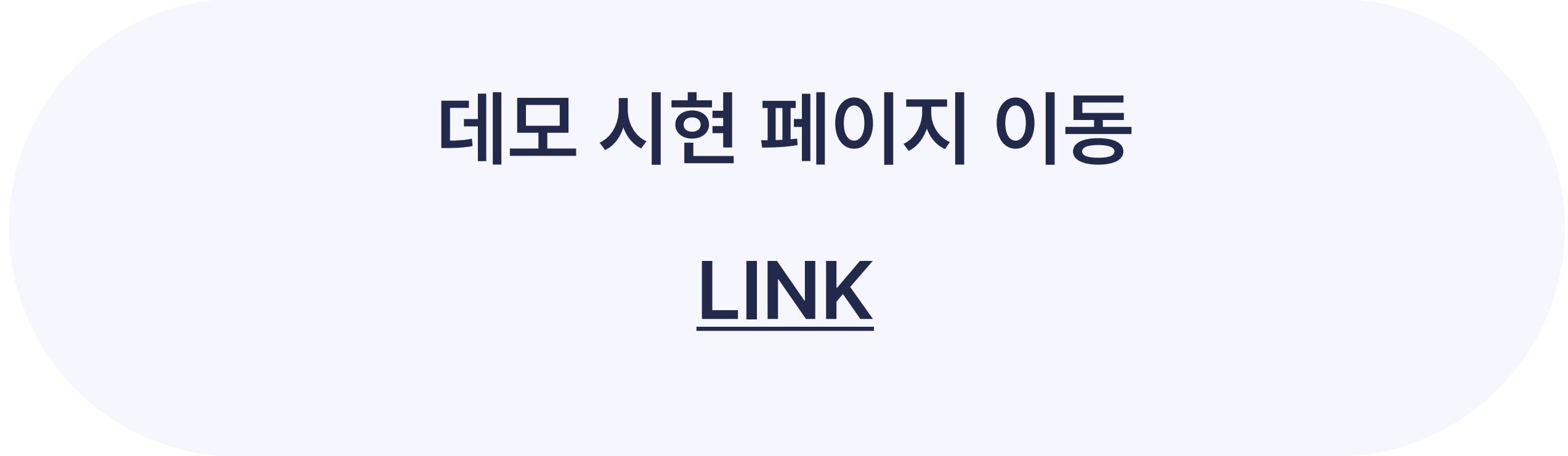
CHAPTER

# DEMO

04 DEMO

## DEMO 시현

---



데모 시현 페이지 이동

LINK

---

---

05

CHAPTER

# CONCLUSION

# 결론

사용자의 개인화된 선호도에 기반하여 전문 지식 없이도  
손쉽게 방을 꾸미고 인테리어 디자인의 접근이 쉬워지도록 함

2D 이미지에서 가구의 크기를 추정하는 기술에 도전하여  
새로운 기술적 가능성을 제시

# 결론

## 앞으로 더 나아갈 점

### 1. 3D 기울어짐 보정의 한계

단일 이미지는 2D 투영 결과만 제공하므로, 3D 공간에서의 실제 크기와 차이가 있을 수 있음.

### 2. Segmentation 품질 의존

세그멘테이션이 정확하지 않으면 크기·깊이 계산의 정확도도 떨어짐.

### 3. MiDaS 깊이 추정 오차

MiDaS 모델 자체의 한계로 인해 반사, 투명체, 복잡한 배경에서 깊이 추정 오차 발생 가능.

추후 더 많은 가구와 공간으로 확장하여 더 편리하고 많은 제품을 적용할 수 있음을 기대함.

# 마무리 및 피드백

## Github

## 피드백

### 참고 자료

- Grounded sam2 github
- Grounding dino blog
- BLIP-2 PAPAR
- Selenium github
- Midas github
- 오늘의 집

---

THANK YOU

발표를  
들어주셔서 감사합니다 :)