# F-LMM: Grounding Frozen Large Multimodal Models

Size Wu[1] Sheng Jin[2] Wenwei Zhang[3] Lumin Xu[4]
Wentao Liu[2,3] Wei Li[1] Chen Change Loy[1]

[1]S-Lab, Nanyang Technological University [2]Sensetime Research and Tetras.AI
[3]Shanghai AI Laboratory [4]The Chinese University of Hong Kong

# Outline

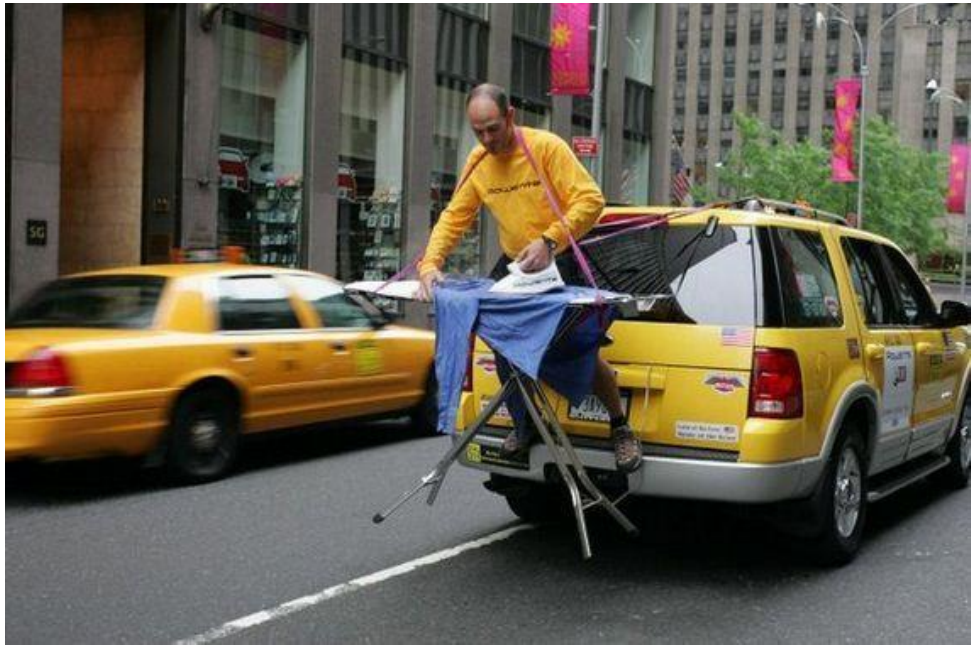
- Background
- Introduction
- Method
- Results

# Outline

- Background
- Introduction
- Method
- Results

# Background

- Large Multimodal Models



`User:` What is unusual about this image?

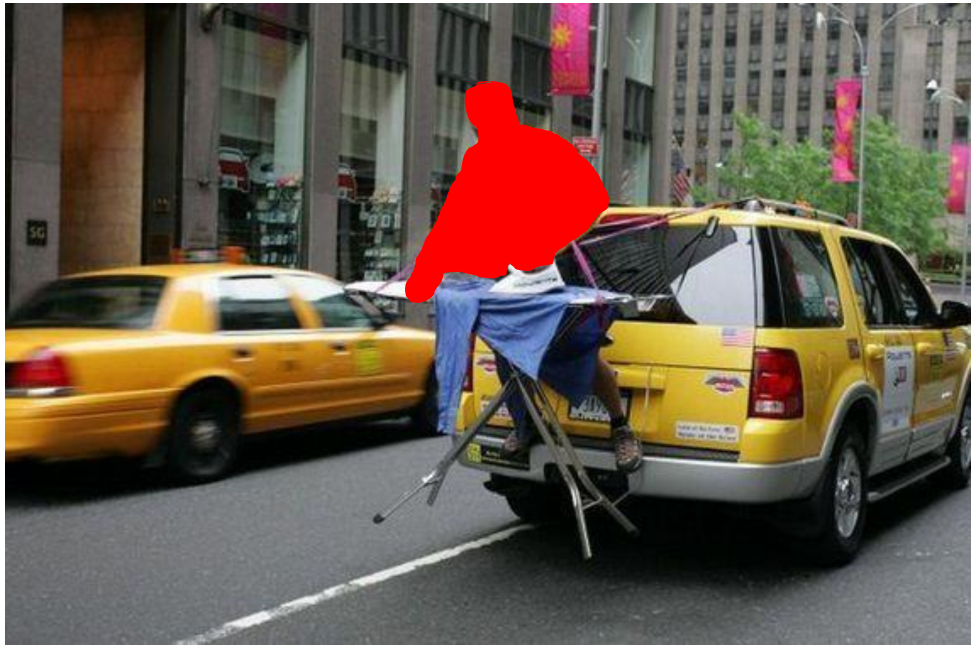
`Model:` The unusual aspect of this image is a man ironing clothes on the back of a minivan or van. This is not a typical place to perform this activity, as one would usually iron clothes in a more stationary and safe location, such as a home, using a regular ironing board. The scene depicted in the image is peculiar as it involves a makeshift ironing setup on a vehicle, which can be both unsafe and unconventional.

Qualitative results of LMM. Example taken from LLaVA[1].

[1] Visual Instruction Tuning, Liu et.al., NeurIPS 2024

# Background

- **Grounding** Large Multimodal Models

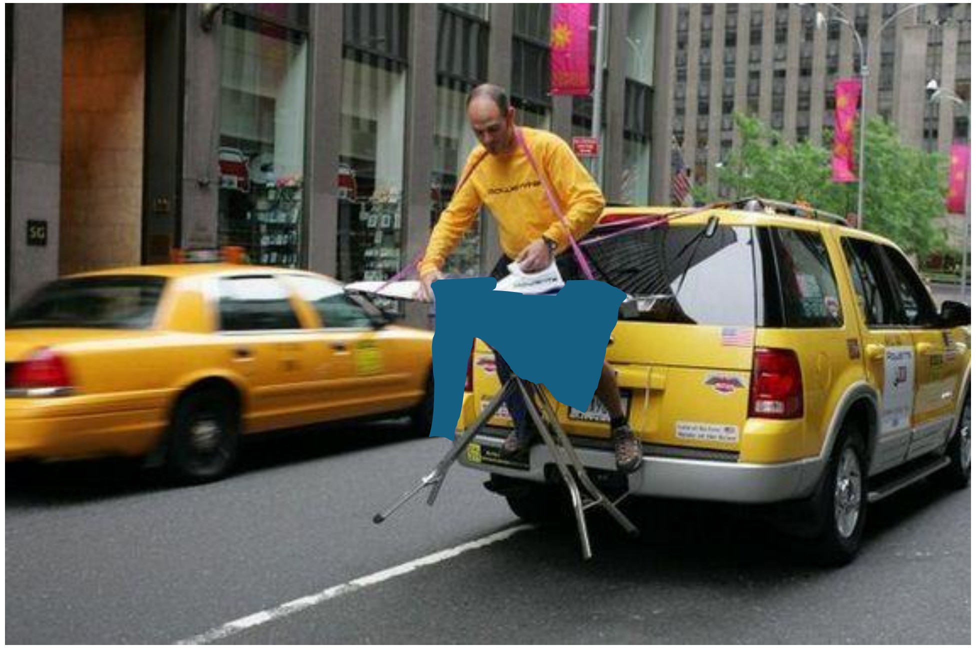


**User:** What is unusual about this image?

**Model:** The unusual aspect of this image is a man ironing clothes on the back of a minivan or van. This is not a typical place to perform this activity, as one would usually iron clothes in a more stationary and safe location, such as a home, using a regular ironing board. The scene depicted in the image is peculiar as it involves a makeshift ironing setup on a vehicle, which can be both unsafe and unconventional.

# Background

- **Grounding** Large Multimodal Models

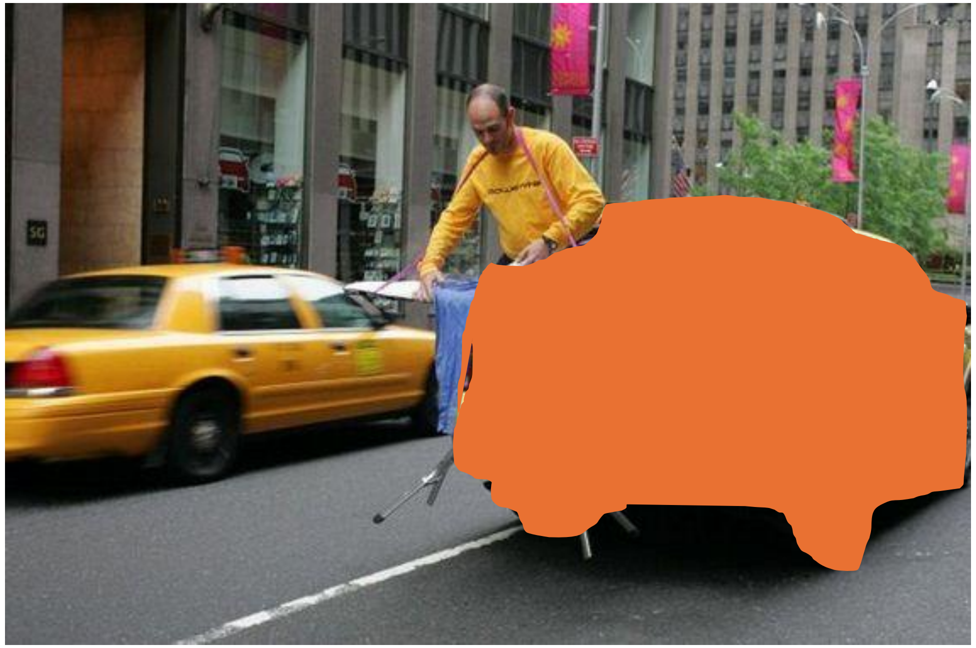


**User:** What is unusual about this image?

**Model:** The unusual aspect of this image is a man ironing **clothes** on the back of a minivan or van. This is not a typical place to perform this activity, as one would usually iron clothes in a more stationary and safe location, such as a home, using a regular ironing board. The scene depicted in the image is peculiar as it involves a makeshift ironing setup on a vehicle, which can be both unsafe and unconventional.

# Background

- **Grounding** Large Multimodal Models



**User:** What is unusual about this image?

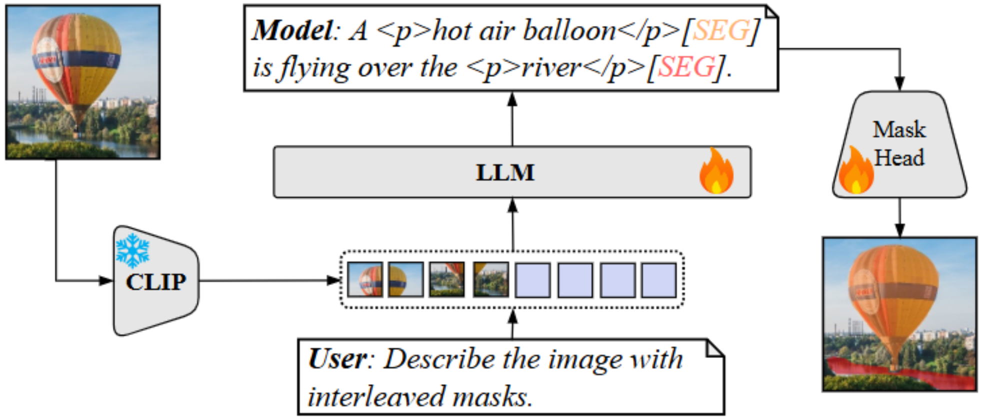
**Model:** The unusual aspect of this image is a man ironing clothes on the back of a minivan or van. This is not a typical place to perform this activity, as one would usually iron clothes in a more stationary and safe location, such as a home, using a regular ironing board. The scene depicted in the image is peculiar as it involves a makeshift ironing setup on a vehicle, which can be both unsafe and unconventional.

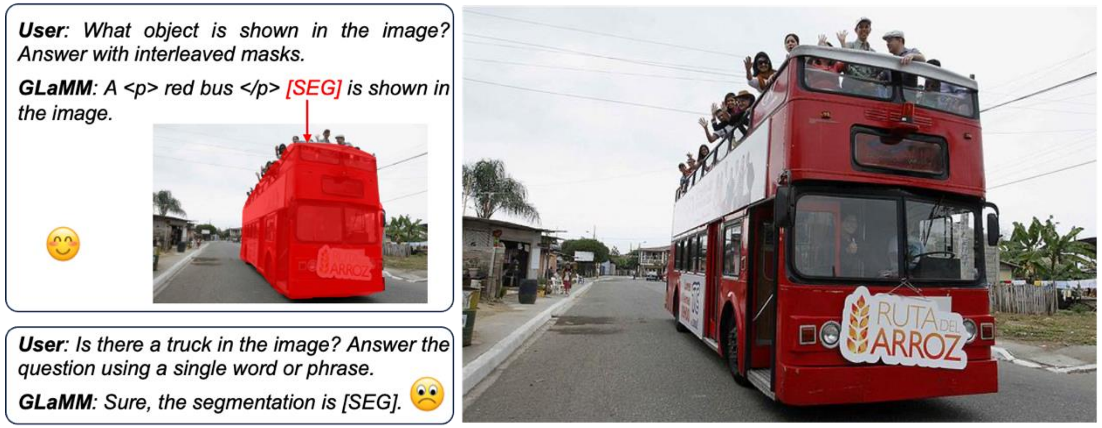
# Outline

- Background
- Introduction
- Method
- Results

# Introduction

- The vanishment of Conversation Ability in Existing Grounding LMMs [2,3,4,5]


*Model*: *A <p>hot air balloon</p>[SEG] is flying over the <p>river</p>[SEG].*
LLM
Mask Head
CLIP
*User*: *Describe the image with interleaved masks.*



*User*: *What object is shown in the image? Answer with interleaved masks.*

*GLaMM*: *A <p> red bus </p> [SEG] is shown in the image.* 😊

*User*: *Is there a truck in the image? Answer the question using a single word or phrase.*

*GLaMM*: *Sure, the segmentation is [SEG].* 😟

Analysis:

- Simple and fixed patterns in segmentation instructions

  `“Segmentation xxx in the image.”`

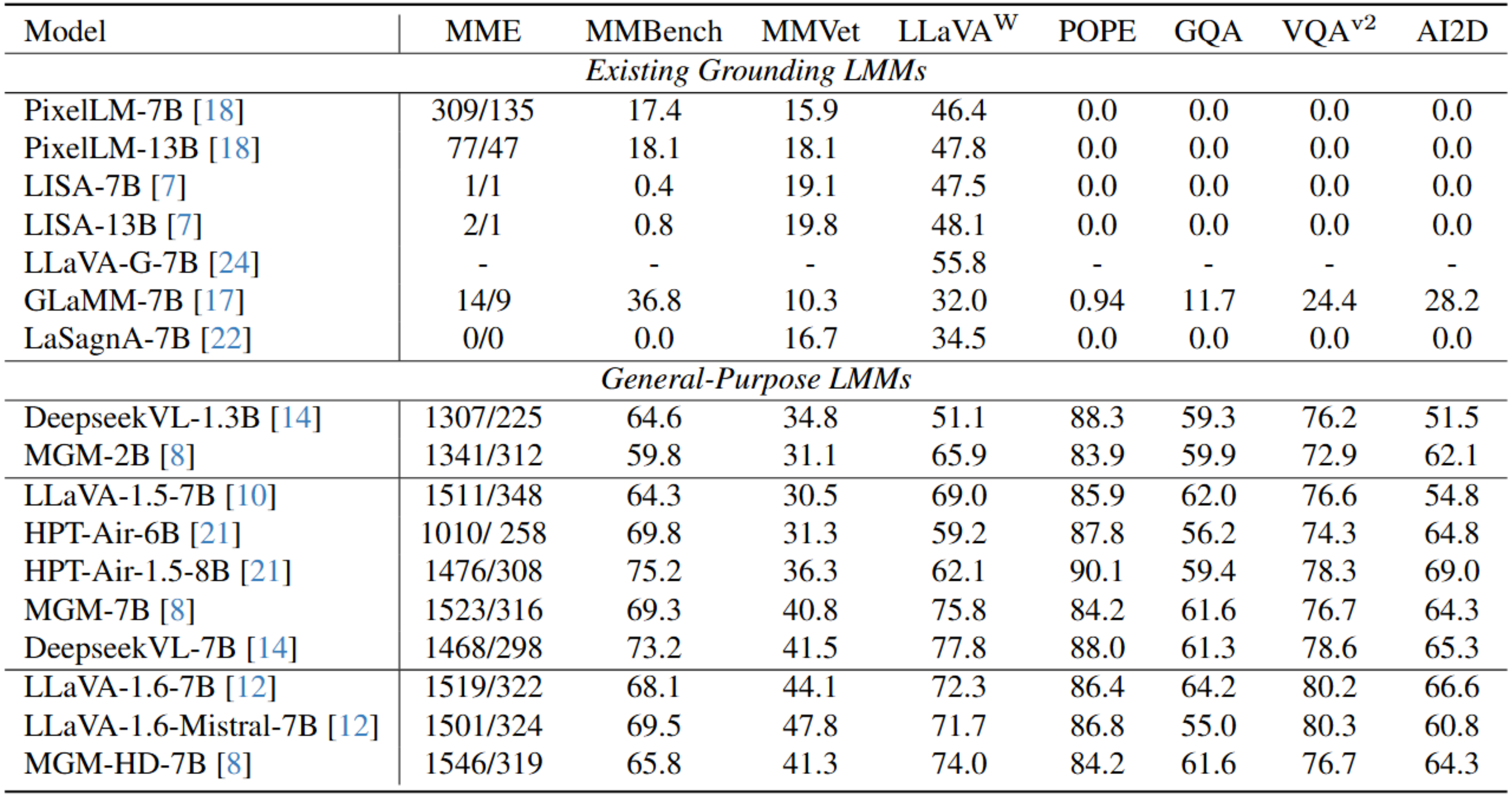
  `“Please describe the image in detail. Answer with interleaved masks”`

- Overfitting to segmentation datasets

*[2] LISA: Reasoning Segmentation via Large Language Model, Lai et.al., CVPR 2024*
*[3] GLaMM: Pixel Grounding Large Multimodal Model, Rasheed et.al., CVPR 2024*
*[4] PixelLM: Pixel Reasoning with Large Multimodal Model, Ren et.al., CVPR 2024*
*[5] GROUNDHOG: Grounding Large Language Models to Holistic Segmentation, Zhang et.al., CVPR 2024*

# Introduction

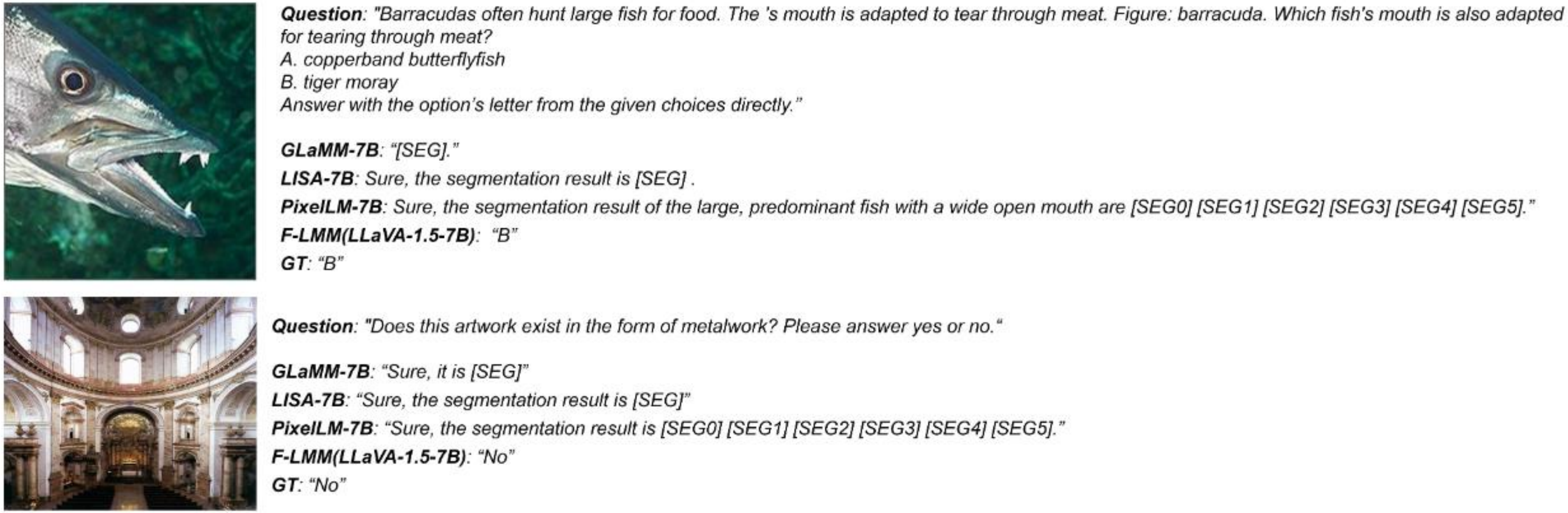
- The vanishment of Conversation Ability in Existing Grounding LMMs



| Model | MME | MMBench | MMVet | LLaVA$^{W}$ | POPE | GQA | VQA$^{v2}$ | AI2D |
|---|---|---|---|---|---|---|---|---|
| *Existing Grounding LMMs* | | | | | | | | |
| PixelLM-7B [18] | 309/135 | 17.4 | 15.9 | 46.4 | 0.0 | 0.0 | 0.0 | 0.0 |
| PixelLM-13B [18] | 77/47 | 18.1 | 18.1 | 47.8 | 0.0 | 0.0 | 0.0 | 0.0 |
| LISA-7B [7] | 1/1 | 0.4 | 19.1 | 47.5 | 0.0 | 0.0 | 0.0 | 0.0 |
| LISA-13B [7] | 2/1 | 0.8 | 19.8 | 48.1 | 0.0 | 0.0 | 0.0 | 0.0 |
| LLaVA-G-7B [24] | - | - | - | 55.8 | - | - | - | - |
| GLaMM-7B [17] | 14/9 | 36.8 | 10.3 | 32.0 | 0.94 | 11.7 | 24.4 | 28.2 |
| LaSagnA-7B [22] | 0/0 | 0.0 | 16.7 | 34.5 | 0.0 | 0.0 | 0.0 | 0.0 |
| *General-Purpose LMMs* | | | | | | | | |
| DeepseekVL-1.3B [14] | 1307/225 | 64.6 | 34.8 | 51.1 | 88.3 | 59.3 | 76.2 | 51.5 |
| MGM-2B [8] | 1341/312 | 59.8 | 31.1 | 65.9 | 83.9 | 59.9 | 72.9 | 62.1 |
| LLaVA-1.5-7B [10] | 1511/348 | 64.3 | 30.5 | 69.0 | 85.9 | 62.0 | 76.6 | 54.8 |
| HPT-Air-6B [21] | 1010/ 258 | 69.8 | 31.3 | 59.2 | 87.8 | 56.2 | 74.3 | 64.8 |
| HPT-Air-1.5-8B [21] | 1476/308 | 75.2 | 36.3 | 62.1 | 90.1 | 59.4 | 78.3 | 69.0 |
| MGM-7B [8] | 1523/316 | 69.3 | 40.8 | 75.8 | 84.2 | 61.6 | 76.7 | 64.3 |
| DeepseekVL-7B [14] | 1468/298 | 73.2 | 41.5 | 77.8 | 88.0 | 61.3 | 78.6 | 65.3 |
| LLaVA-1.6-7B [12] | 1519/322 | 68.1 | 44.1 | 72.3 | 86.4 | 64.2 | 80.2 | 66.6 |
| LLaVA-1.6-Mistral-7B [12] | 1501/324 | 69.5 | 47.8 | 71.7 | 86.8 | 55.0 | 80.3 | 60.8 |
| MGM-HD-7B [8] | 1546/319 | 65.8 | 41.3 | 74.0 | 84.2 | 61.6 | 76.7 | 64.3 |

# Introduction

- The vanishment of Conversation Ability in Existing Grounding LMMs



***Question***: *"Barracudas often hunt large fish for food. The 's mouth is adapted to tear through meat. Figure: barracuda. Which fish's mouth is also adapted for tearing through meat?*
*A. copperband butterflyfish*
*B. tiger moray*
*Answer with the option's letter from the given choices directly."*

***GLaMM-7B***: *"[SEG]."*

***LISA-7B***: *Sure, the segmentation result is [SEG] .*

***PixelLM-7B***: *Sure, the segmentation result of the large, predominant fish with a wide open mouth are [SEG0] [SEG1] [SEG2] [SEG3] [SEG4] [SEG5]."*

***F-LMM(LLaVA-1.5-7B)***: *"B"*

***GT***: *"B"*

***Question***: *"Does this artwork exist in the form of metalwork? Please answer yes or no."*

***GLaMM-7B***: *"Sure, it is [SEG]"*

***LISA-7B***: *"Sure, the segmentation result is [SEG]"*

***PixelLM-7B***: *"Sure, the segmentation result is [SEG0] [SEG1] [SEG2] [SEG3] [SEG4] [SEG5]."*

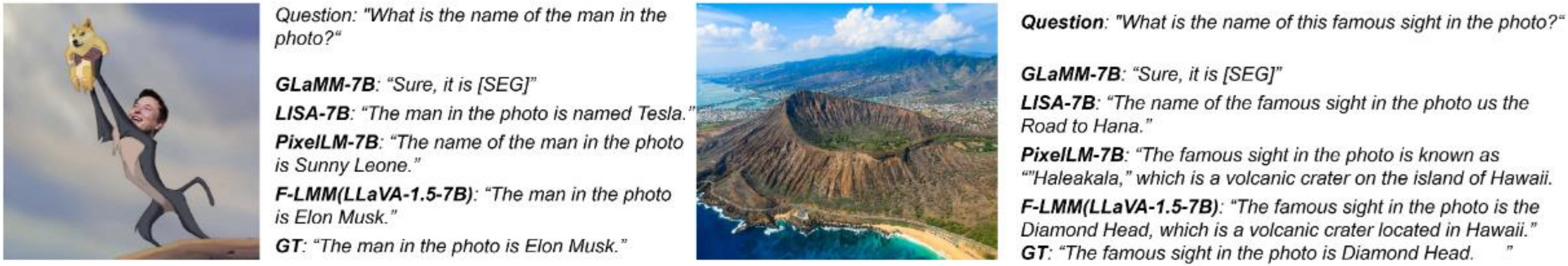
***F-LMM(LLaVA-1.5-7B)***: *"No"*

***GT***: *"No"*

**Qualitative results**. Examples from MMBench and MME.

# Introduction

- The vanishment of Conversation Ability in Existing Grounding LMMs



*Question: "What is the name of the man in the photo?“*

***GLaMM-7B**: “Sure, it is [SEG]”*

***LISA-7B**: “The man in the photo is named Tesla.”*

***PixelLM-7B**: “The name of the man in the photo is Sunny Leone.”*

***F-LMM(LLaVA-1.5-7B)**: “The man in the photo is Elon Musk.”*

***GT**: “The man in the photo is Elon Musk.”*

***Question**: "What is the name of this famous sight in the photo?“*

***GLaMM-7B**: “Sure, it is [SEG]”*

***LISA-7B**: “The name of the famous sight in the photo us the Road to Hana.”*

***PixelLM-7B**: “The famous sight in the photo is known as “”Haleakala,” which is a volcanic crater on the island of Hawaii.*

***F-LMM(LLaVA-1.5-7B)**: “The famous sight in the photo is the Diamond Head, which is a volcanic crater located in Hawaii.”*

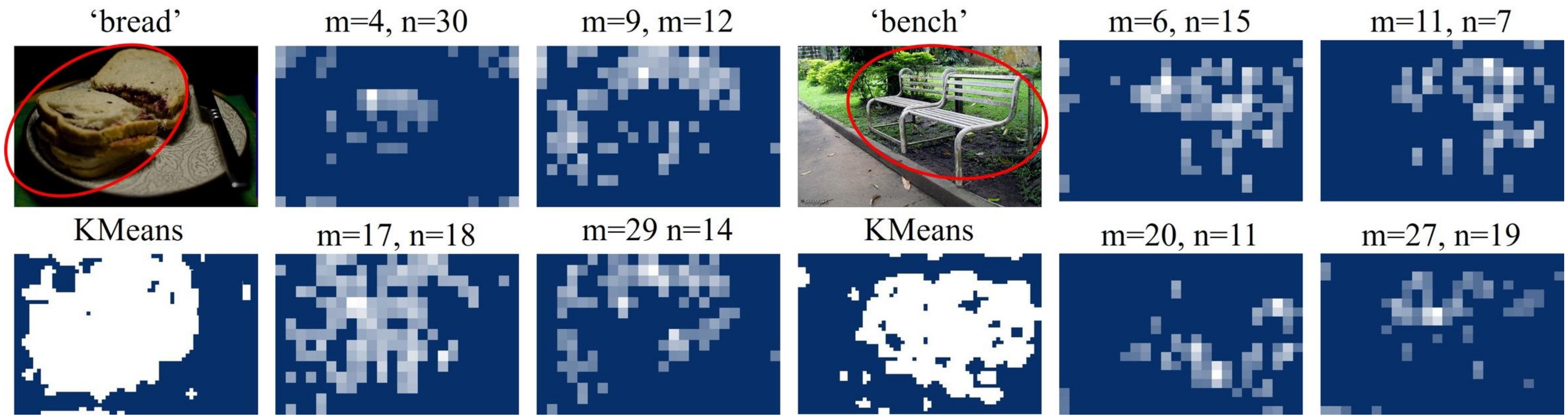
***GT**: “The famous sight in the photo is Diamond Head. ”*

**Qualitative results.** Examples from LLaVA-In-The-Wild.

# Introduction

- Our straightforward and resource-efficient solution:

***F-LMM**: Grounding Frozen Large Multimodal Models*

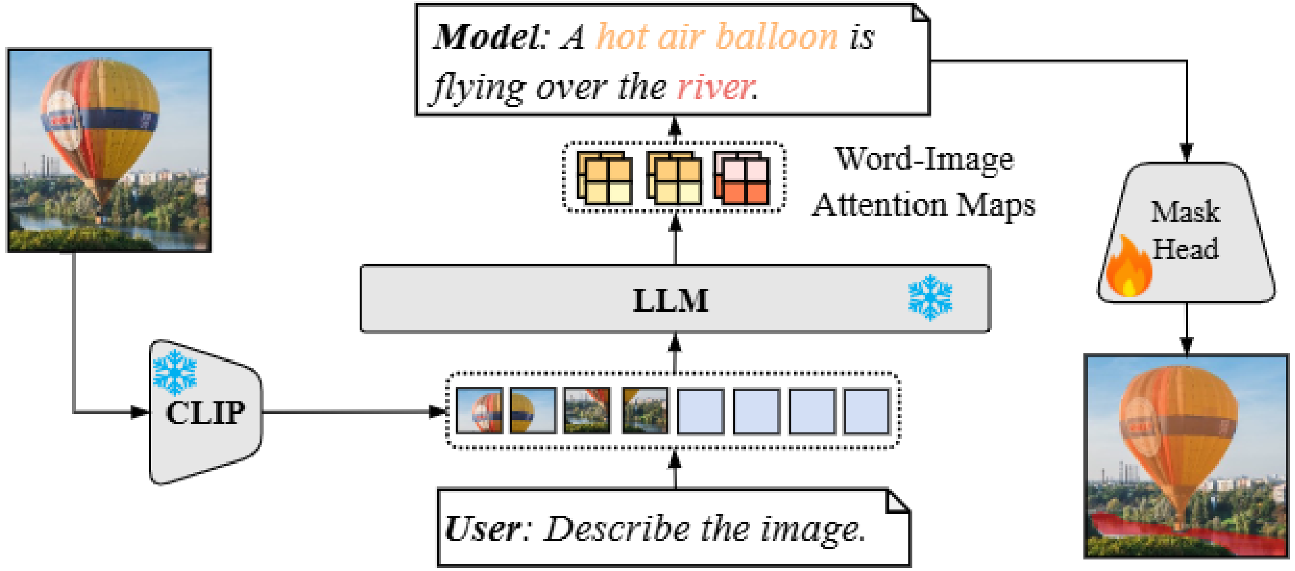


Segmentation priors exist in word-image attention maps.

# Introduction

- Our straightforward and resource-efficient solution:

***F-LMM**: Grounding Frozen Large Multimodal Models*



Model: A hot air balloon is flying over the river.
Word-Image Attention Maps
Mask Head
LLM
CLIP
User: Describe the image.

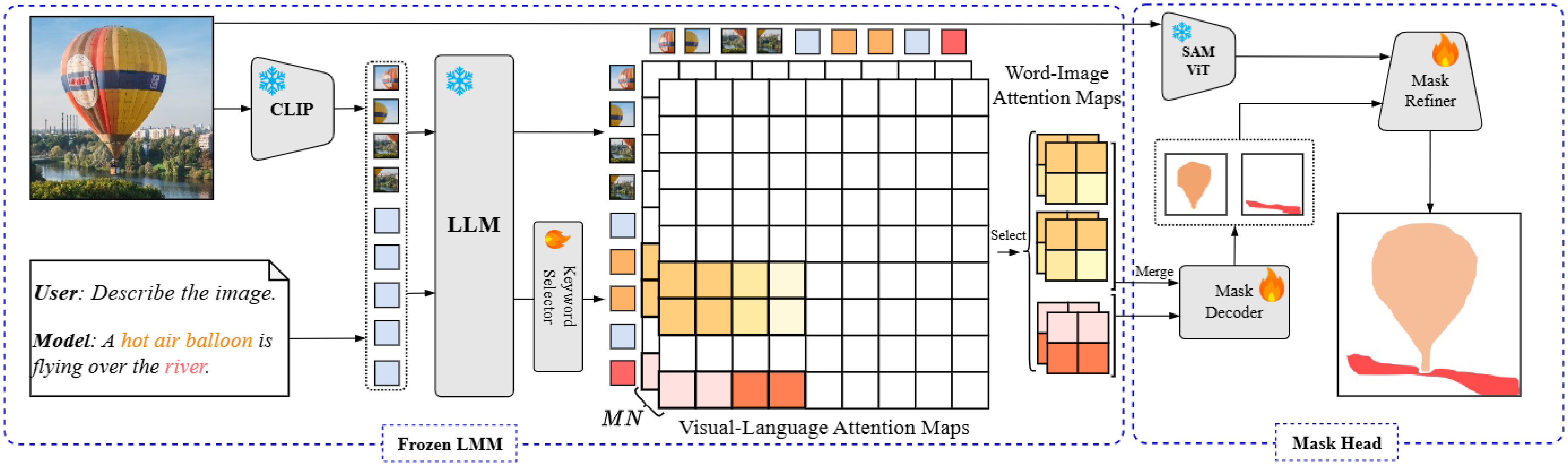
F-LMM: Exploring segmentation priors in word-attention maps

# Outline

- Background
- Introduction
- Method
- Results

# Method

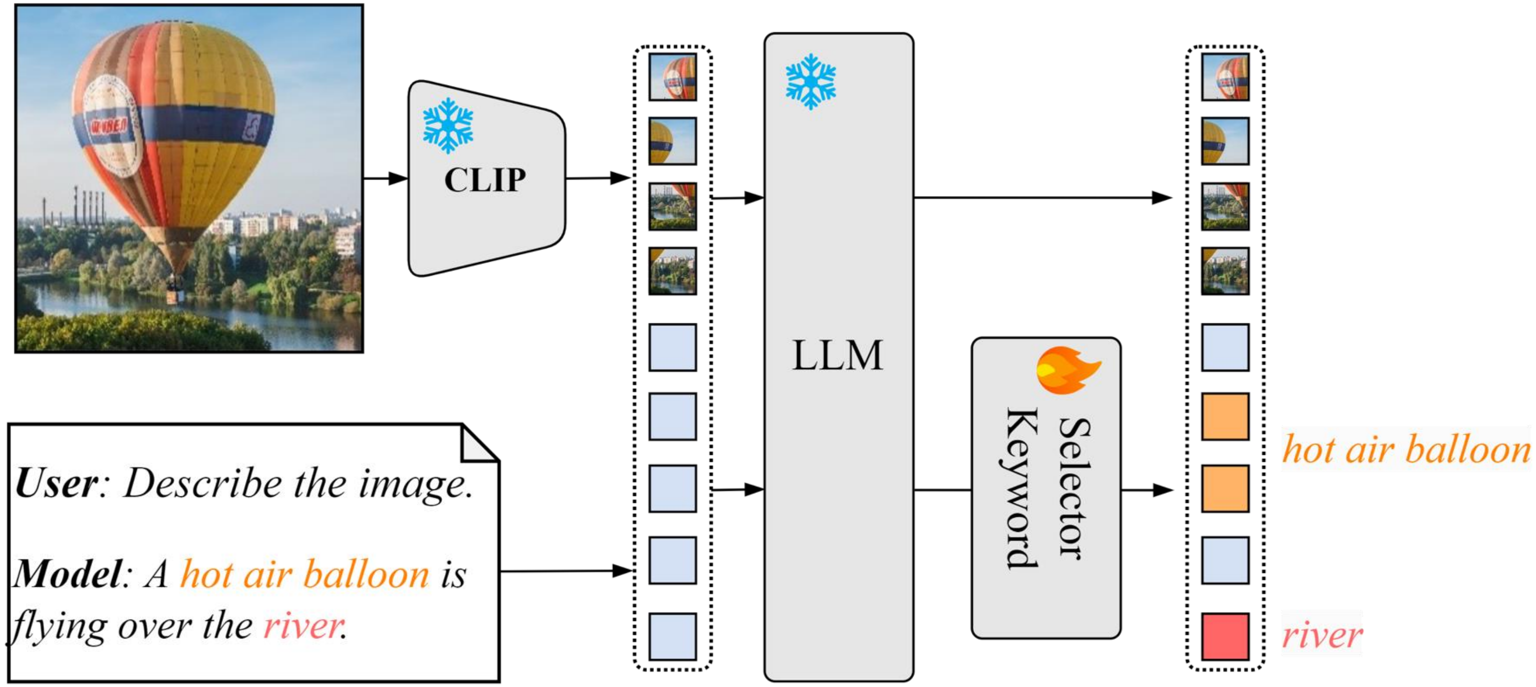
- **F-LMM:** Grounding Frozen Large Multimodal Models


CLIP
LLM
Keyword Selector
Word-Image Attention Maps
Select
Merge
SAM ViT
Mask Refiner
Mask Decoder
MN
Visual-Language Attention Maps
Frozen LMM
Mask Head
User: Describe the image.
Model: A hot air balloon is flying over the river.

Overall pipeline of F-LMM. The main modules are a **mask decoder**, a **mask refiner** and a **keyword selector**.

# Method

- **F-LMM:** Grounding Frozen Large Multimodal Models
  - Keyword Selector: A linear layer + sigmoid



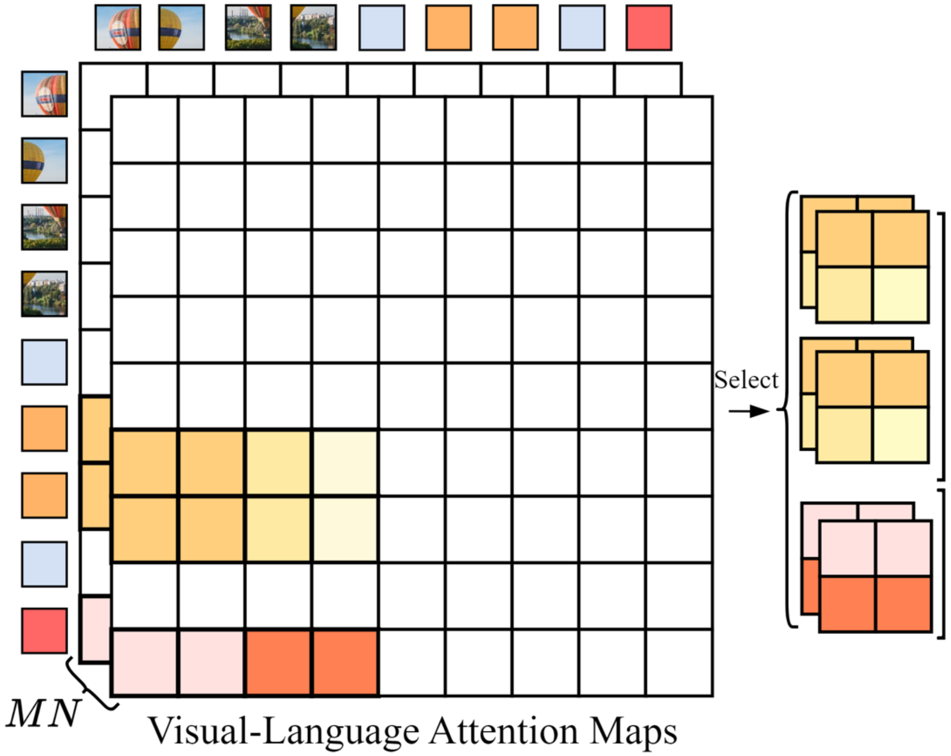
*User*: *Describe the image.*

*Model*: *A hot air balloon is flying over the river.*

We use a linear layer to discover visual objects in the text sequence

# Method

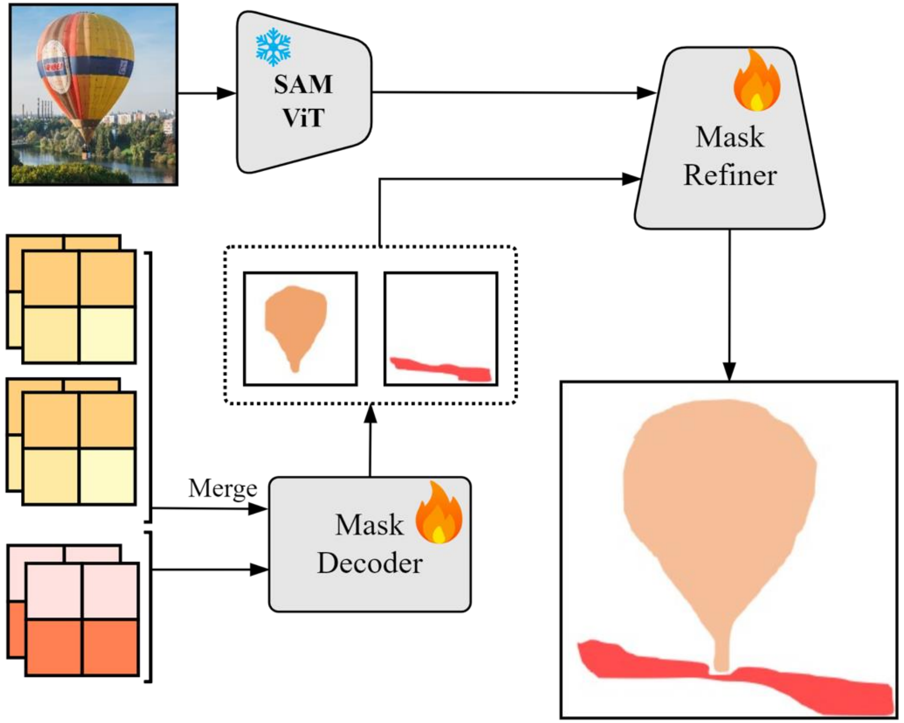
- **F-LMM:** Grounding Frozen Large Multimodal Models
  - Collecting word-image attention maps



Extracting word-image correspondences in vision-language attention

# Method

- **F-LMM:** Grounding Frozen Large Multimodal Models
  - Mask head

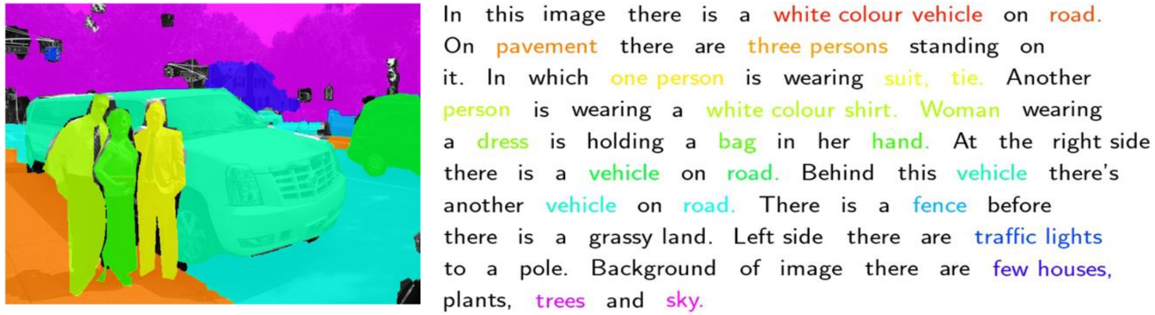


Perform visual grounding with the mask decoder and the mask refiner.

# Method

- **F-LMM:** Grounding Frozen Large Multimodal Models
  - Training

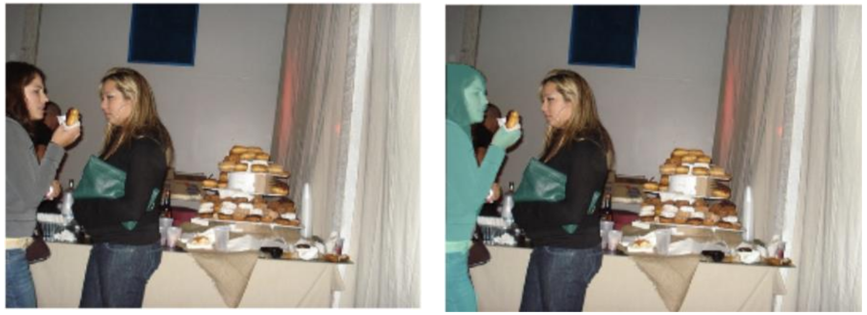
PNG + RES datasets ---- 190K data samples



In this image there is a white colour vehicle on road. On pavement there are three persons standing on it. In which one person is wearing suit, tie. Another person is wearing a white colour shirt. Woman wearing a dress is holding a bag in her hand. At the right side there is a vehicle on road. Behind this vehicle there's another vehicle on road. There is a fence before there is a grassy land. Left side there are traffic lights to a pole. Background of image there are few houses, plants, trees and sky.

Panoptic Narrative Grounding (PNG) [6]

“A woman with a donut”



Referring Expression Segmentation (RES) [7]

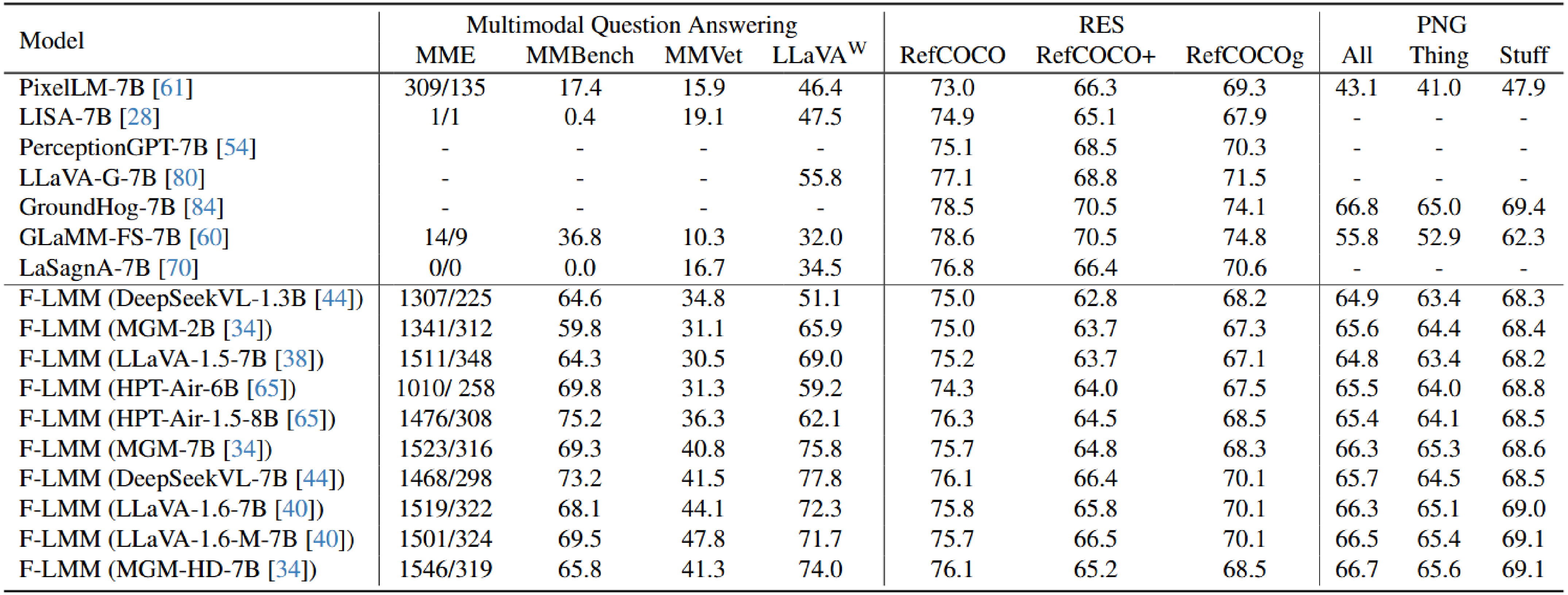
*[6] Panoptic Narrative Grounding, González et.al., ICCV 2021*
*[7] Referitgame: Referring to objects in photographs of natural scenes, Kazemzadeh et.al., EMNLP 2014*

# Outline

- Background
- Introduction
- Method
- Results

# Results

- Main evaluation
  - Measuring chat ability: question-answering benchmarks
  - Measuring visual grounding ability: Referring Segmentation (RES) & Panoptic Narrative Grounding (PNG)



| Model | Multimodal Question Answering | | | | RES | | | PNG | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | MME | MMBench | MMVet | LLaVA$^W$ | RefCOCO | RefCOCO+ | RefCOCOg | All | Thing | Stuff |
| PixelLM-7B [61] | 309/135 | 17.4 | 15.9 | 46.4 | 73.0 | 66.3 | 69.3 | 43.1 | 41.0 | 47.9 |
| LISA-7B [28] | 1/1 | 0.4 | 19.1 | 47.5 | 74.9 | 65.1 | 67.9 | - | - | - |
| PerceptionGPT-7B [54] | - | - | - | - | 75.1 | 68.5 | 70.3 | - | - | - |
| LLaVA-G-7B [80] | - | - | - | 55.8 | 77.1 | 68.8 | 71.5 | - | - | - |
| GroundHog-7B [84] | - | - | - | - | 78.5 | 70.5 | 74.1 | 66.8 | 65.0 | 69.4 |
| GLaMM-FS-7B [60] | 14/9 | 36.8 | 10.3 | 32.0 | 78.6 | 70.5 | 74.8 | 55.8 | 52.9 | 62.3 |
| LaSagnA-7B [70] | 0/0 | 0.0 | 16.7 | 34.5 | 76.8 | 66.4 | 70.6 | - | - | - |
| F-LMM (DeepSeekVL-1.3B [44]) | 1307/225 | 64.6 | 34.8 | 51.1 | 75.0 | 62.8 | 68.2 | 64.9 | 63.4 | 68.3 |
| F-LMM (MGM-2B [34]) | 1341/312 | 59.8 | 31.1 | 65.9 | 75.0 | 63.7 | 67.3 | 65.6 | 64.4 | 68.4 |
| F-LMM (LLaVA-1.5-7B [38]) | 1511/348 | 64.3 | 30.5 | 69.0 | 75.2 | 63.7 | 67.1 | 64.8 | 63.4 | 68.2 |
| F-LMM (HPT-Air-6B [65]) | 1010/ 258 | 69.8 | 31.3 | 59.2 | 74.3 | 64.0 | 67.5 | 65.5 | 64.0 | 68.8 |
| F-LMM (HPT-Air-1.5-8B [65]) | 1476/308 | 75.2 | 36.3 | 62.1 | 76.3 | 64.5 | 68.5 | 65.4 | 64.1 | 68.5 |
| F-LMM (MGM-7B [34]) | 1523/316 | 69.3 | 40.8 | 75.8 | 75.7 | 64.8 | 68.3 | 66.3 | 65.3 | 68.6 |
| F-LMM (DeepSeekVL-7B [44]) | 1468/298 | 73.2 | 41.5 | 77.8 | 76.1 | 66.4 | 70.1 | 65.7 | 64.5 | 68.5 |
| F-LMM (LLaVA-1.6-7B [40]) | 1519/322 | 68.1 | 44.1 | 72.3 | 75.8 | 65.8 | 70.1 | 66.3 | 65.1 | 69.0 |
| F-LMM (LLaVA-1.6-M-7B [40]) | 1501/324 | 69.5 | 47.8 | 71.7 | 75.7 | 66.5 | 70.1 | 66.5 | 65.4 | 69.1 |
| F-LMM (MGM-HD-7B [34]) | 1546/319 | 65.8 | 41.3 | 74.0 | 76.1 | 65.2 | 68.5 | 66.7 | 65.6 | 69.1 |

# Results

- More complex scenarios
    - Reasoning Segmentation [2]

Quantitative results on Reasoning Segmentation

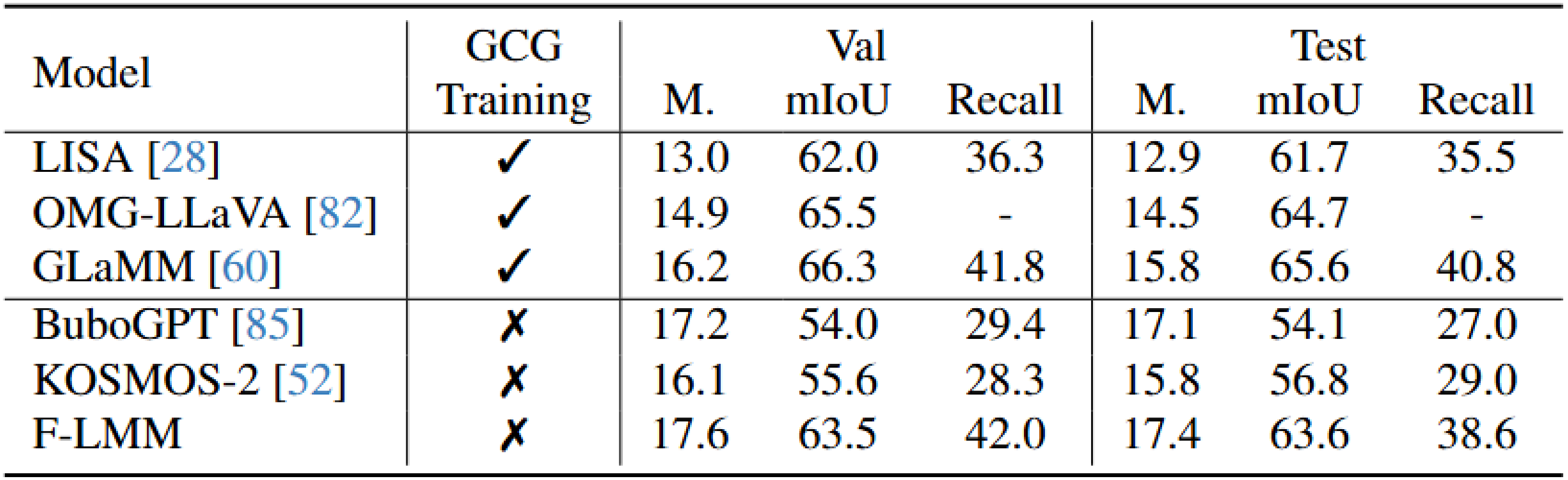
| Model | Val | Test | | |
|---|---|---|---|---|
| | - | Short | Long | All |
| X-Decoder [87] | 22.6 | 20.4 | 22.2 | 21.7 |
| SEEM [88] | 25.5 | 20.1 | 25.6 | 24.3 |
| GroundingSAM [41] | 26.0 | 17.8 | 22.4 | 21.3 |
| OVSeg [35] | 28.5 | 18.0 | 28.7 | 26.1 |
| LISA [28] | 44.4 | 37.6 | 36.6 | 36.8 |
| F-LMM | 46.7 | 36.9 | 49.1 | 46.2 |

*[2] LISA: Reasoning Segmentation via Large Language Model, Lai et.al., CVPR 2024*

# Results

- More complex scenarios
  - Grounded Conversation Generation (GCG) [3]

**GCG Evaluation**. M.: METEOR, measuring caption quality.



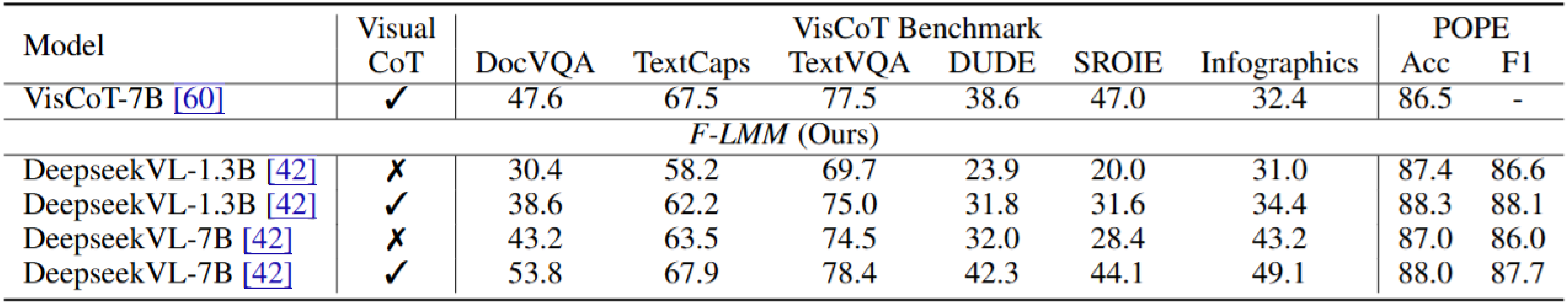
| Model | GCG Training | Val |  |  | Test |  |  |
|---|---|---|---|---|---|---|---|
|  |  | M. | mIoU | Recall | M. | mIoU | Recall |
| LISA [28] | ✓ | 13.0 | 62.0 | 36.3 | 12.9 | 61.7 | 35.5 |
| OMG-LLaVA [82] | ✓ | 14.9 | 65.5 | - | 14.5 | 64.7 | - |
| GLaMM [60] | ✓ | 16.2 | 66.3 | 41.8 | 15.8 | 65.6 | 40.8 |
| BuboGPT [85] | ✗ | 17.2 | 54.0 | 29.4 | 17.1 | 54.1 | 27.0 |
| KOSMOS-2 [52] | ✗ | 16.1 | 55.6 | 28.3 | 15.8 | 56.8 | 29.0 |
| F-LMM | ✗ | 17.6 | 63.5 | 42.0 | 17.4 | 63.6 | 38.6 |

*[3] GLaMM: Pixel Grounding Large Multimodal Model, Rasheed et.al., CVPR 2024*

# Results

- More complex scenarios
  - Visual Chain-of-Thought Reasoning (VisCoT)[8]

Unleashing visual chain-of-thought reasoning with both excellent grounding and instruction-following ability.



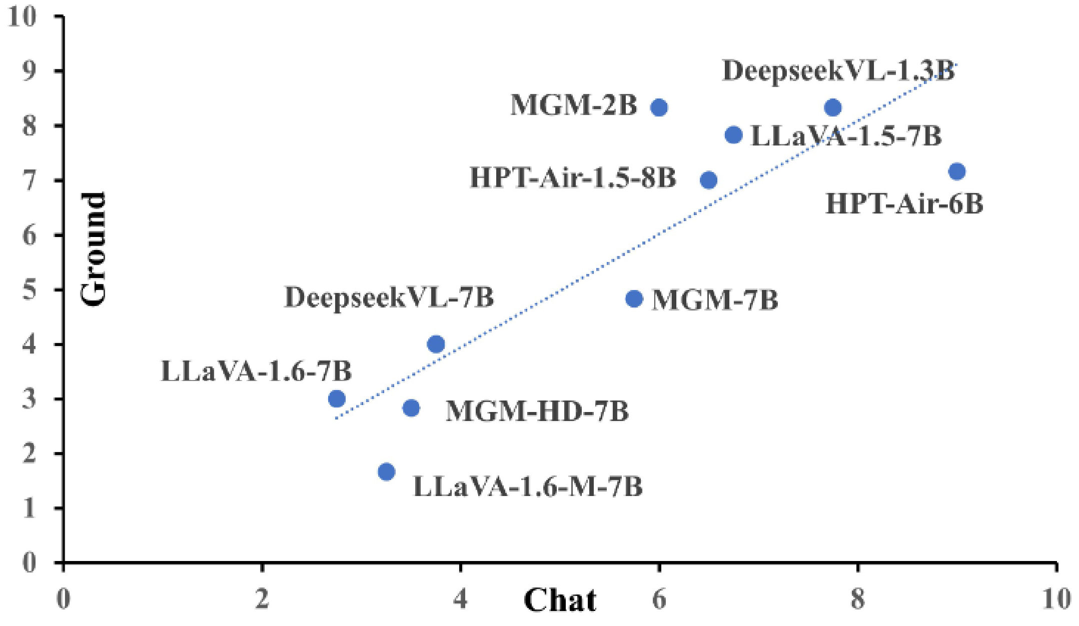
| Model | Visual CoT | VisCoT Benchmark<br>DocVQA | TextCaps | TextVQA | DUDE | SROIE | Infographics | POPE<br>Acc | F1 |
|---|---|---|---|---|---|---|---|---|---|
| VisCoT-7B [60] | ✓ | 47.6 | 67.5 | 77.5 | 38.6 | 47.0 | 32.4 | 86.5 | - |
| *F-LMM* (Ours) | | | | | | | | | |
| DeepseekVL-1.3B [42] | ✗ | 30.4 | 58.2 | 69.7 | 23.9 | 20.0 | 31.0 | 87.4 | 86.6 |
| DeepseekVL-1.3B [42] | ✓ | 38.6 | 62.2 | 75.0 | 31.8 | 31.6 | 34.4 | 88.3 | 88.1 |
| DeepseekVL-7B [42] | ✗ | 43.2 | 63.5 | 74.5 | 32.0 | 28.4 | 43.2 | 87.0 | 86.0 |
| DeepseekVL-7B [42] | ✓ | 53.8 | 67.9 | 78.4 | 42.3 | 44.1 | 49.1 | 88.0 | 87.7 |

*[8] Visual CoT: Advancing Multi-Modal Language Models with a Comprehensive Dataset and Benchmark for Chain-of-Thought Reasoning, Shao et.al., Arxiv 2024*

# Results

- Analysis
  - Positive Correlation between Conversational & Grounding Ability

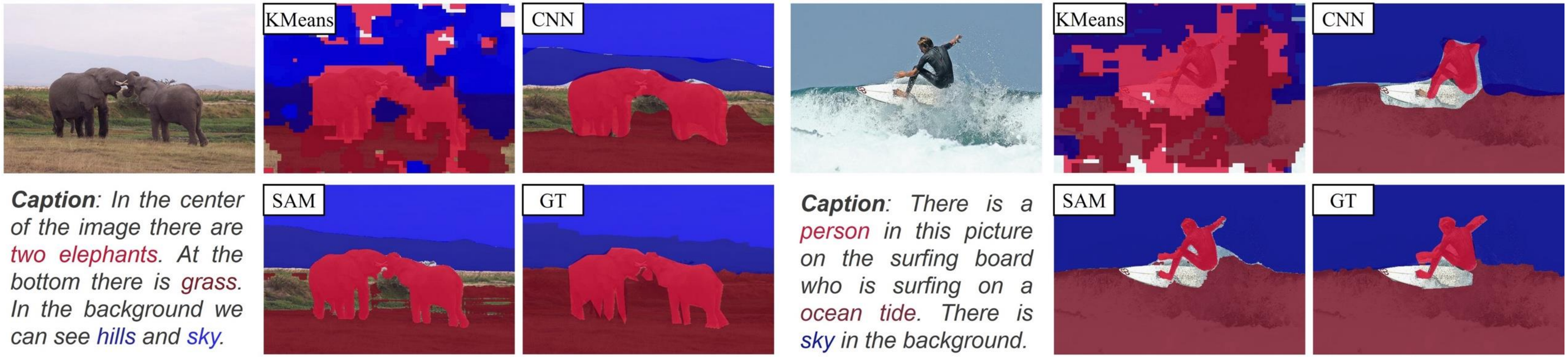
| Model | Chat ↓ | Ground ↓ |
|---|---|---|
| DeepseekVL-1.3B [44] | 7.75 | 8.33 |
| MGM-2B [34] | 6.00 | 8.33 |
| LLaVA-1.5-7B [38] | 6.75 | 7.83 |
| HPT-Air-6B [65] | 9.00 | 7.16 |
| HPT-Air-1.5-8B [65] | 6.50 | 7.00 |
| MGM-7B [34] | 5.75 | 4.83 |
| DeepseekVL-7B [44] | 3.75 | 4.00 |
| LLaVA-1.6-7B [40] | 2.75 | 3.00 |
| LLaVA-1.6-M-7B [40] | 3.25 | 1.66 |
| MGM-HD-7B [34] | 3.50 | 2.83 |



Average ranks of models on question-answering (chat) and grounding benchmarks.

# Results

- Visualization
  - From attention maps to segmentation masks



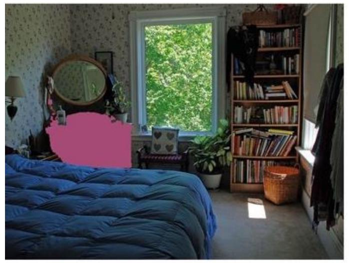
***Caption**: In the center of the image there are two elephants. At the bottom there is grass. In the background we can see hills and sky.*

***Caption**: There is a person in this picture on the surfing board who is surfing on a ocean tide. There is sky in the background.*

**Top-left**: KMeans visualsation of attention maps; **Top-right**: Masks predicted by the CNN-based mask decoder; **Bottom-left**: Masks refined by the SAM-based mask refiner; **Bottom-right**: GT masks

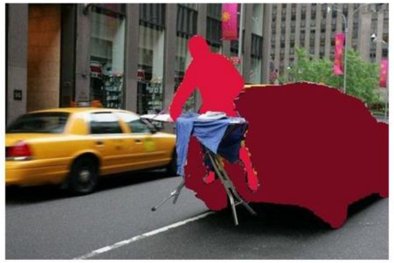
# Results

- Visualization
  - Grounded conversation



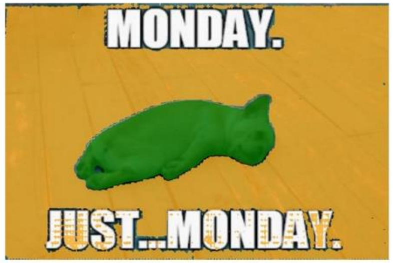
***User***: *Where is the shampoo?*

***Answer***: *On the dresser.*



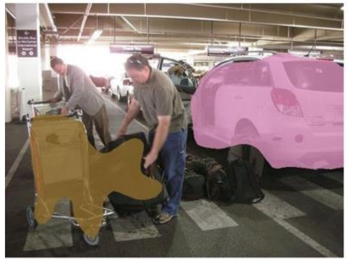
***User:*** *What is unusual about this image?*

***Model:*** *The unusual aspect of this image is the presence of a man in a yellow shirt ironing clothes on the back of a taxi.*



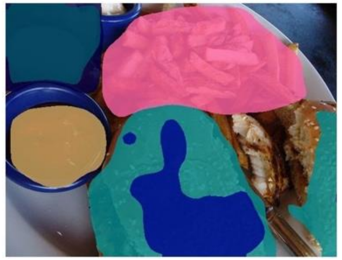
***User:*** *What is the funny aspect of this image?*

***Model:*** *The funny aspect of this image is the juxtaposition of a dog lying on a wooden floor with the caption "MONDAY... JUST... MONDAY."*



***User***: *Where might the picture be taken?*

***Answer***: *The photo appears to be taken in a parking garage or a similar indoor setting. This is suggested by the presence of a car with its doors open, and the presence of a luggage cart with suitcases and bags.*

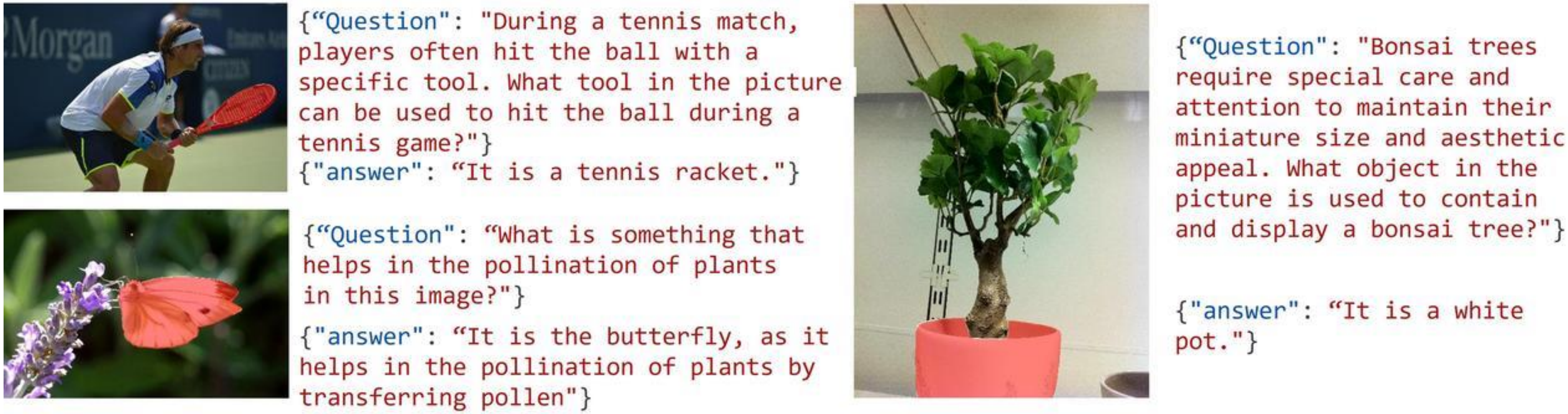


***User:*** *Can you name the foods in this picture?*

***Model:*** *The image shows a plate with various food items. There's a sandwich with sesame seeds on it, a side of French Fries, and a small bowl containing a red sauce, possibly ketchup or a similar condiment. Additionally, there's a small bowl of vegetables, which appears to be sliced onions.*

# Results

- Visualization
    - Reasoning Segmentation



{“Question": "During a tennis match, players often hit the ball with a specific tool. What tool in the picture can be used to hit the ball during a tennis game?"}
{"answer": “It is a tennis racket."}

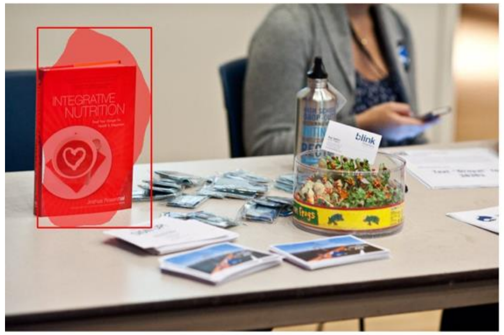
{“Question": “What is something that helps in the pollination of plants in this image?"}
{"answer": “It is the butterfly, as it helps in the pollination of plants by transferring pollen"}

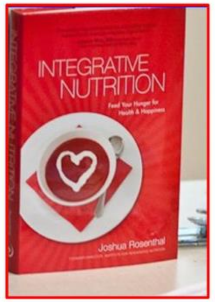
{“Question": "Bonsai trees require special care and attention to maintain their miniature size and aesthetic appeal. What object in the picture is used to contain and display a bonsai tree?"}

{"answer": “It is a white pot."}

# Results

- Visualization
  - Visual Chain-of-Thought Reasoning





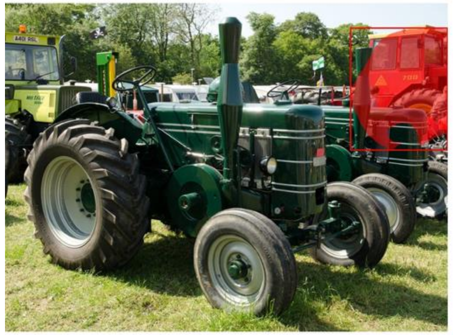
***Question**: What is the title of the book on the table?*

***Thought**: First think which object in this image is most relevant to the question.*

***Answer**: The object most relevant to the question is the book on the table.*

***Question**: What is the title of the book on the table?*

***Answer**: Integrative Nutrition.*



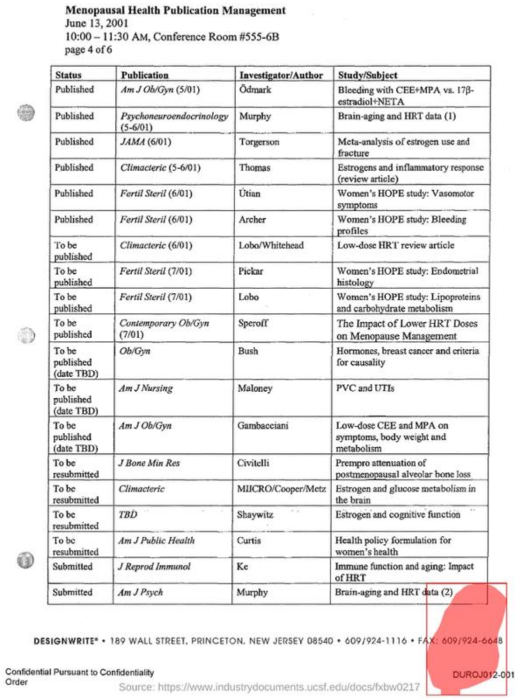
***Question**: What number is prominently displayed on the red tractor?*

***Thought**: First think which object in this image is most relevant to the question.*

***Answer**: The object most relevant to the question is the red tractor.*

***Question**: what number is prominently displayed on the red tractor?*

***Answer**: 700*



**Menopausal Health Publication Management**
June 13, 2001
10:00 – 11:30 AM, Conference Room #555-6B
page 4 of 6

| Status | Publication | Investigator/Author | Study/Subject |
|---|---|---|---|
| Published | *Am J Ob/Gyn* (5/01) | Ödmark | Bleeding with CEE+MPA vs. 17β-estradiol+NETA |
| Published | *Psychoneuroendocrinology* (5-6/01) | Murphy | Brain-aging and HRT data (1) |
| Published | *JAMA* (6/01) | Torgerson | Meta-analysis of estrogen use and fracture |
| Published | *Climacteric* (5-6/01) | Thomas | Estrogens and inflammatory response (review article) |
| Published | *Fertil Steril* (6/01) | Útian | Women's HOPE study: Vasomotor symptoms |
| Published | *Fertil Steril* (6/01) | Archer | Women's HOPE study: Bleeding profiles |
| To be published | *Climacteric* (6/01) | Lobo/Whitehead | Low-dose HRT review article |
| To be published | *Fertil Steril* (7/01) | Pickar | Women's HOPE study: Endometrial histology |
| To be published | *Fertil Steril* (7/01) | Lobo | Women's HOPE study: Lipoproteins and carbohydrate metabolism |
| To be published | *Contemporary Ob/Gyn* (7/01) | Speroff | The Impact of Lower HRT Doses on Menopause Management |
| To be published (date TBD) | *Ob/Gyn* | Bush | Hormones, breast cancer and criteria for causality |
| To be published (date TBD) | *Am J Nursing* | Maloney | PVC and UTIs |
| To be published (date TBD) | *Am J Ob/Gyn* | Gambacciani | Low-dose CEE and MPA on symptoms, body weight and metabolism |
| To be resubmitted | *J Bone Min Res* | Civitelli | Prempro attenuation of postmenopausal alveolar bone loss |
| To be resubmitted | *Climacteric* | MIICRO/Cooper/Metz | Estrogen and glucose metabolism in the brain |
| To be resubmitted | *TBD* | Shaywitz | Estrogen and cognitive function |
| To be resubmitted | *Am J Public Health* | Curtis | Health policy formulation for women's health |
| Submitted | *J Reprod Immunol* | Ke | Immune function and aging: Impact of HRT |
| Submitted | *Am J Psych* | Murphy | Brain-aging and HRT data (2) |

DESIGNWRITE® • 189 WALL STREET, PRINCETON, NEW JERSEY 08540 • 609/924-1116 • FAX: 609/924-6648

Confidential Pursuant to Confidentiality Order

DUROJ012-001293

Source: https://www.industrydocuments.ucsf.edu/docs/fxbw0217

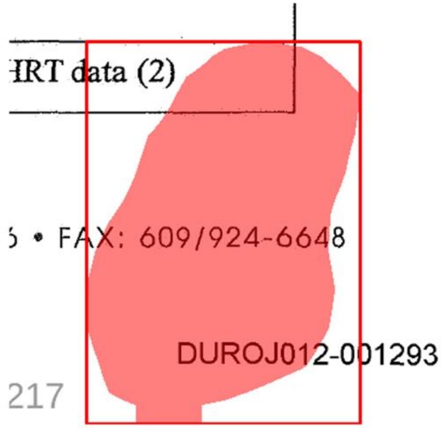
***Question**: What is the fax number present in the document?*

***Thought**: First think which object in this image is most relevant to the question.*

***Answer**: The object most relevant to the question is the bottom of the image, where the fax number is mentioned.*

***Question**: What is the fax number present in the document?*

***Answer**: 609/924-6648*


HRT data (2)
• FAX: 609/924-6648
DUROJ012-001293
217

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

S-LAB FOR ADVANCED INTELLIGENCE

商汤 sensetime

# Thanks!

CVPR Nashville JUNE 11-15, 2025