

2023 / 12

Advanced AI Developments

OGQ Industrial and Academic Achievements

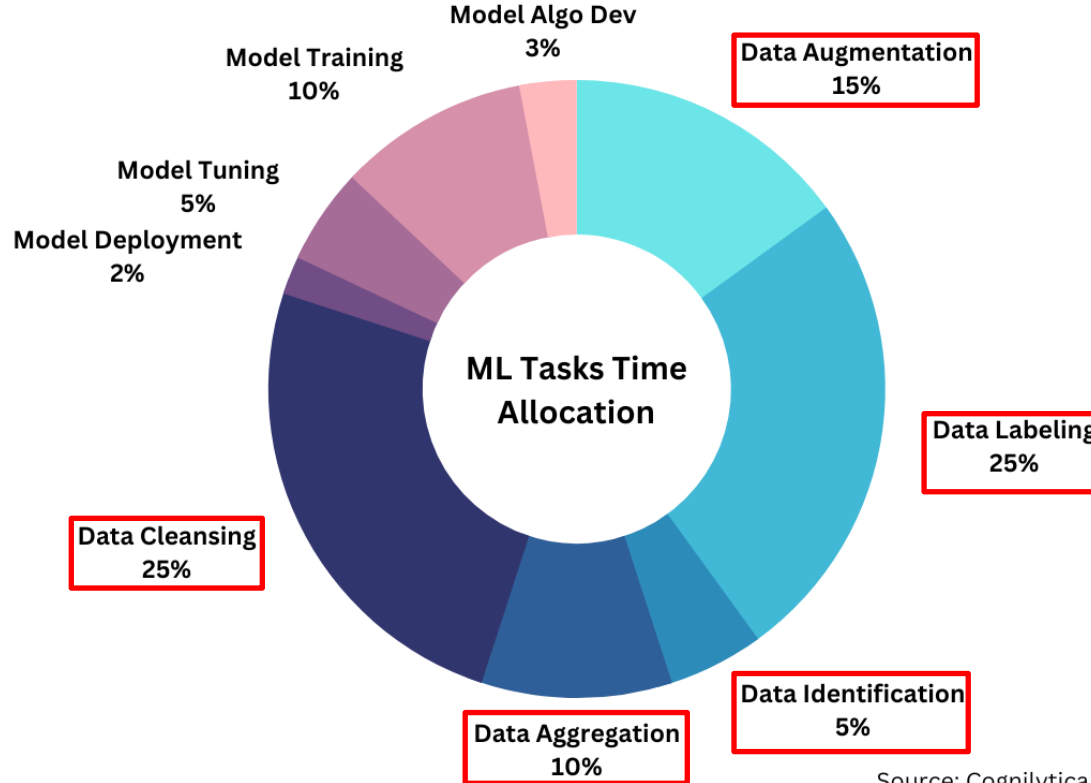
Future Trajectory at OGQ

OGQ / Sanghyun Jo

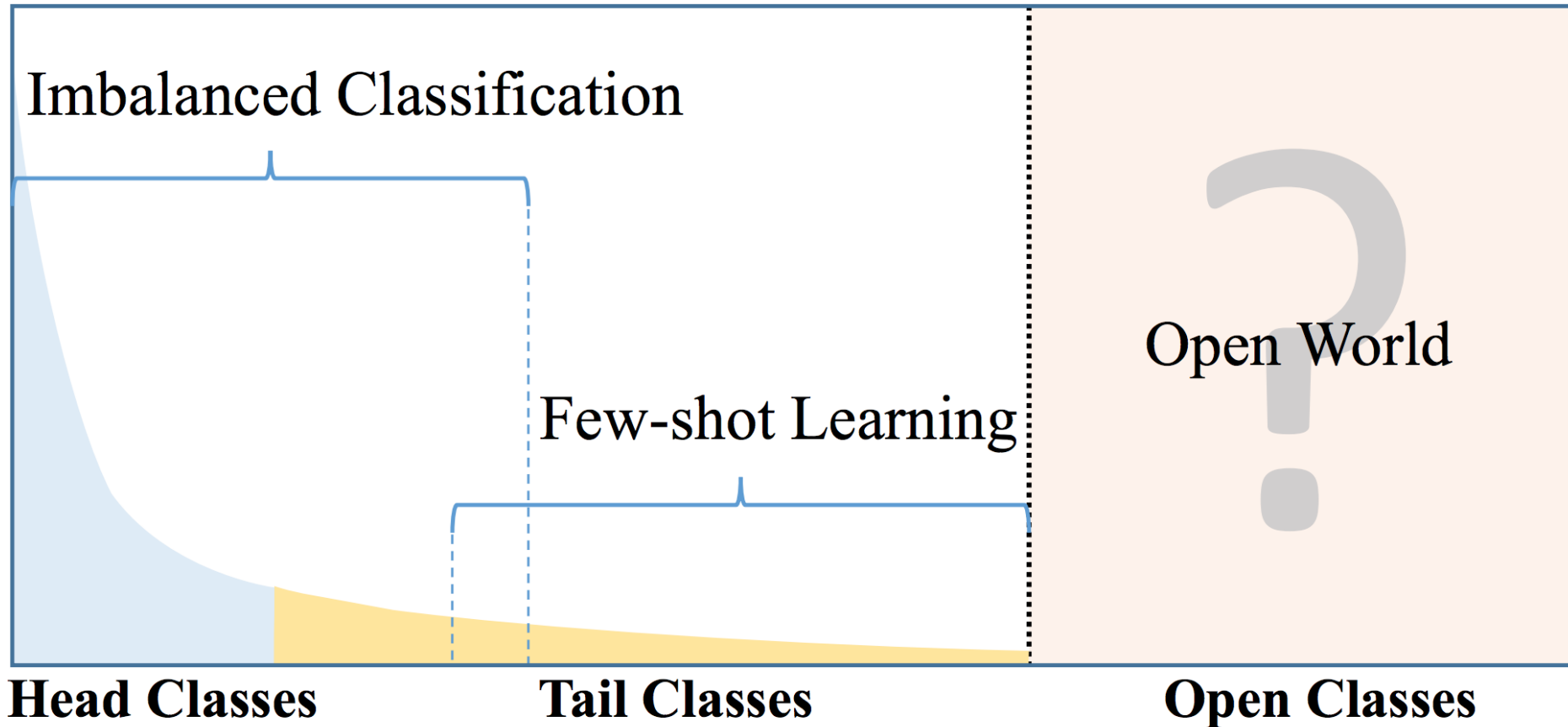


Overcoming Challenges in AI Application for Industry

- ✓ Data preparation, including cleansing and labeling, is critical, accounting for 80% of the machine learning workflow.



Limitations of Open-Source Data: Class Imbalance



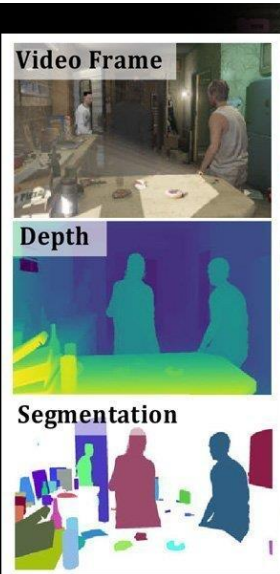
Limitations of Open-Source Data: Domain Adaptation

- ✓ Employing synthetic datasets to circumvent dataset constraints. However, existing models still exhibit overfitting to synthetic datasets, degrading real-world performance.

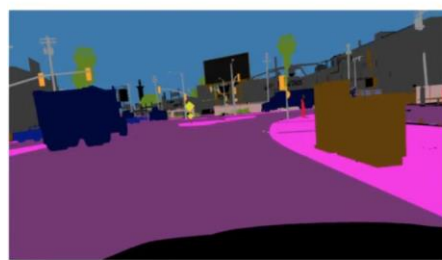


Limitations of Open-Source Data: Domain Adaptation

- ✓ Employing synthetic datasets to circumvent dataset constraints. However, existing models still exhibit overfitting to synthetic datasets, degrading real-world performance.



Source Image



Source Label



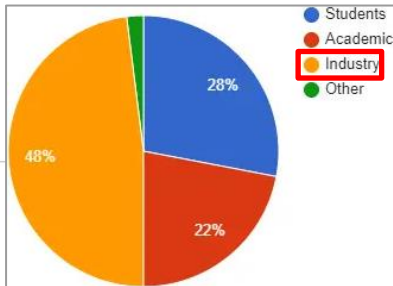
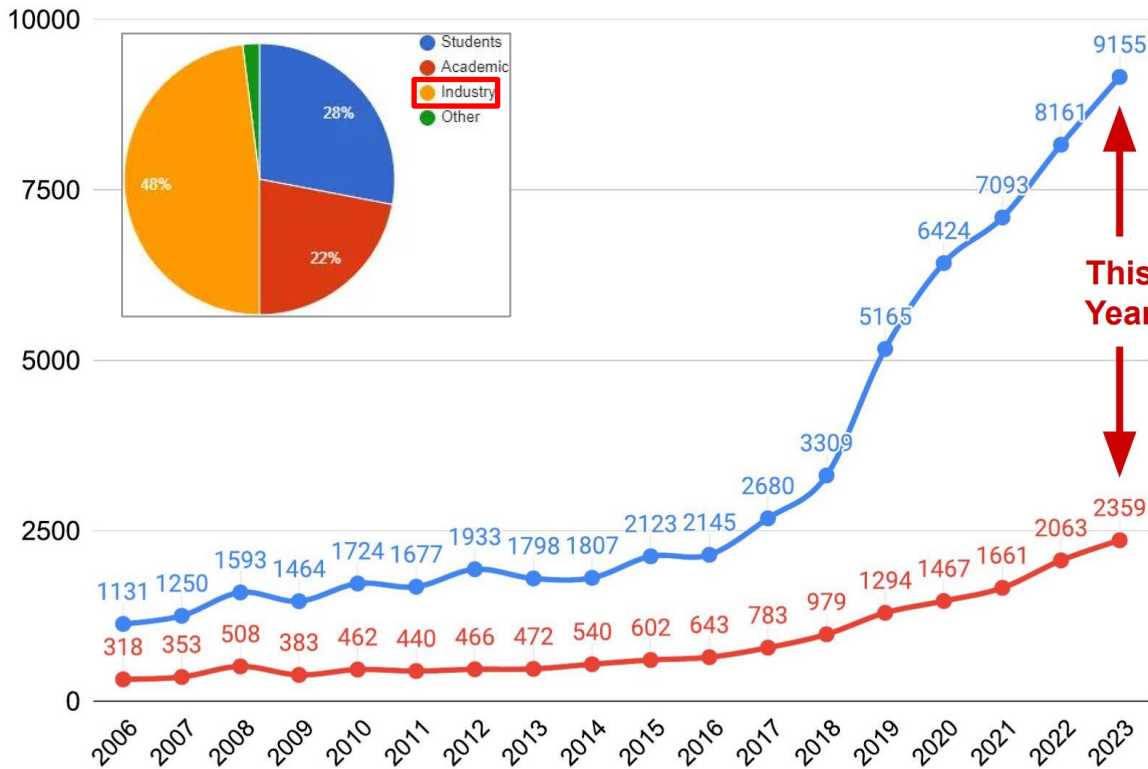
Target Image



Target Label

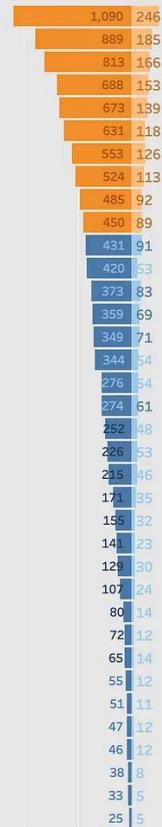
AI Report (CVPR 2023)

Submitted Papers Accepted Papers



- 1 3D from multi-view and sensors
- 2 Image and video synthesis and generation
- 3 Humans: Face, body, pose, gesture, movement
- 4 Transfer, meta, low-shot, continual, or long-tail learning
- 5 Recognition: Categorization, detection, retrieval
- 6 Vision, language, and reasoning
- 7 Low-level vision
- 8 Segmentation, grouping and shape analysis
- 9 Deep learning architectures and techniques
- 10 Multi-modal learning

- 11 3D from single images
- 12 Medical and biological vision, cell microscopy
- 13 Video: Action and event understanding
- 14 Autonomous driving
- 15 Self-supervised or unsupervised representation learning
- 16 Datasets and evaluation
- 17 Scene analysis and understanding
- 18 Adversarial attack and defense
- 19 Efficient and scalable vision
- 20 Computational imaging
- 21 Video: Low-level analysis, motion, and tracking
- 22 Vision applications and systems
- 23 Vision + graphics
- 24 Robotics
- 25 Transparency, fairness, accountability, privacy, ethics in vision
- 26 Explainable computer vision
- 27 Embodied vision: Active agents, simulation
- 28 Document analysis and understanding
- 29 Machine learning (other than deep learning)
- 30 Physics-based vision and shape-from-X
- 31 Biometrics
- 32 Others
- 33 Optimization methods (other than deep learning)
- 34 Photogrammetry and remote sensing
- 35 Computer vision theory
- 36 Computer vision for social good



OpenAI DevDay

- ✓ Highlight the importance of building a technical moat for business expansion.



Understanding Data Labeling in AI

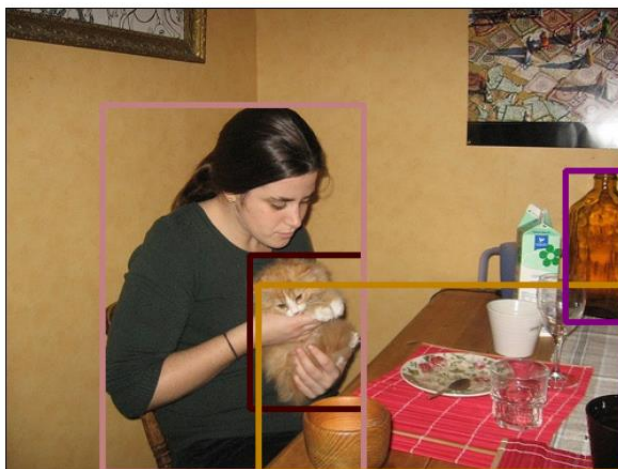
- ✓ Data labeling involves assigning descriptive tags to raw data, such as images.
- ✓ Collecting dense (pixel-wise) annotations demands extensive domain expertise, especially in specialized fields like medicine.

Image Classification

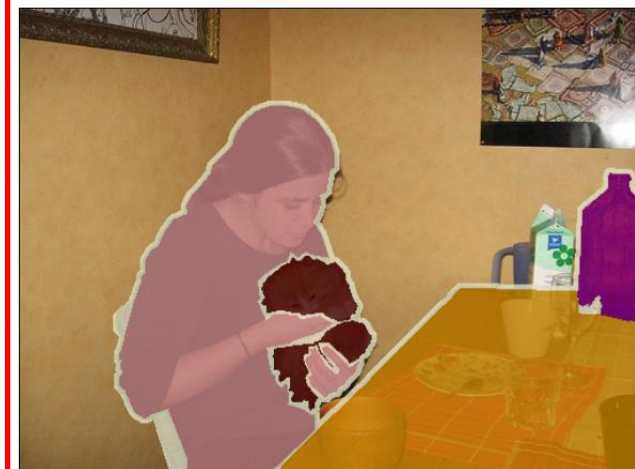


Labeling time: 20s

Object Detection



Semantic Segmentation



Labeling time: 239.7s

Innovations at OGQ: Vision AI

- ✓ Development of an automated labeling system utilizing image-level tags.
- ✓ Our technology produces highly dense annotations within 2 seconds per image (x123).



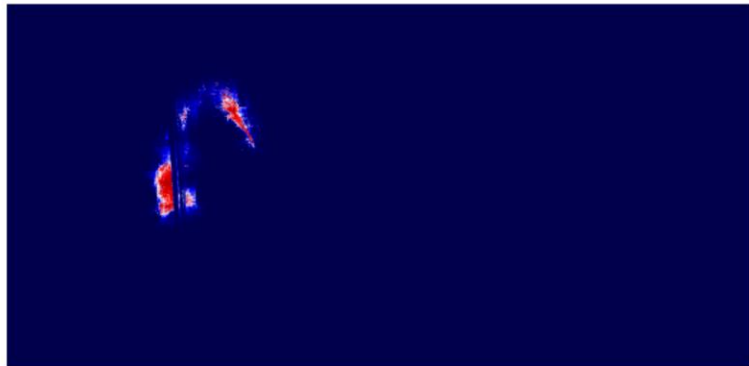
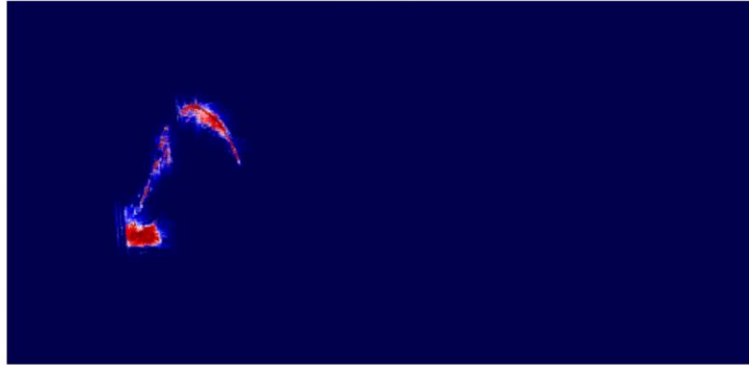
Innovations at OGQ: Vision AI

- ✓ Development of an automated labeling system utilizing image-level tags.
- ✓ Our technology produces highly dense annotations within 2 seconds per image (x123).

Image



MARS



Innovations at OGQ: Vision AI

- ✓ Robust performance across diverse weather conditions, validated by our real-world data.
- ✓ Achievement of KISA CCTV AI certification in Korea for collapse and fire detection.

Night



Rain



Fog

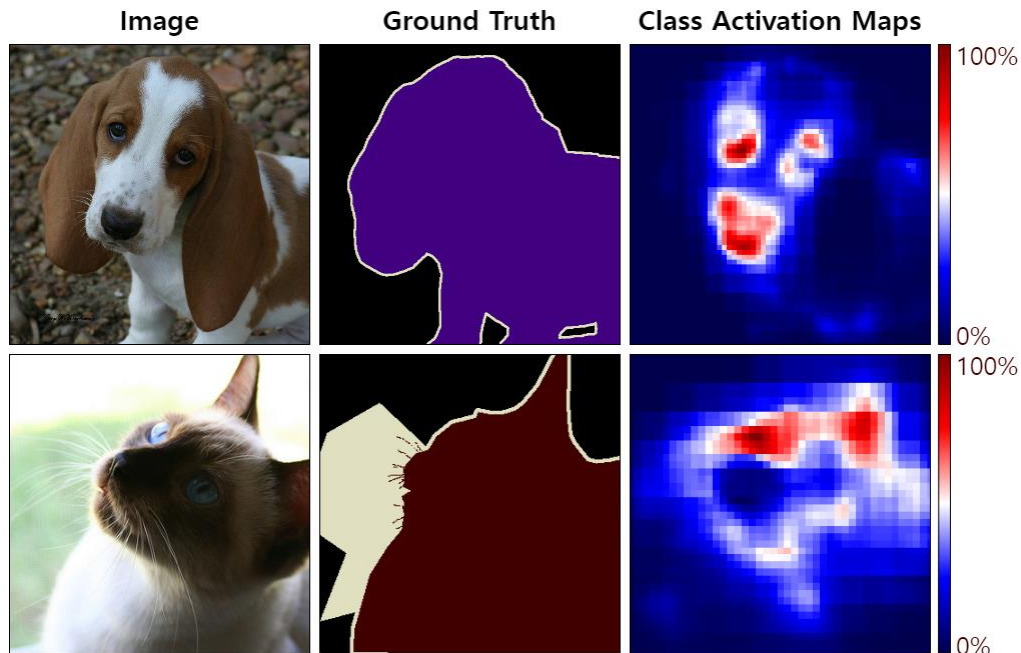
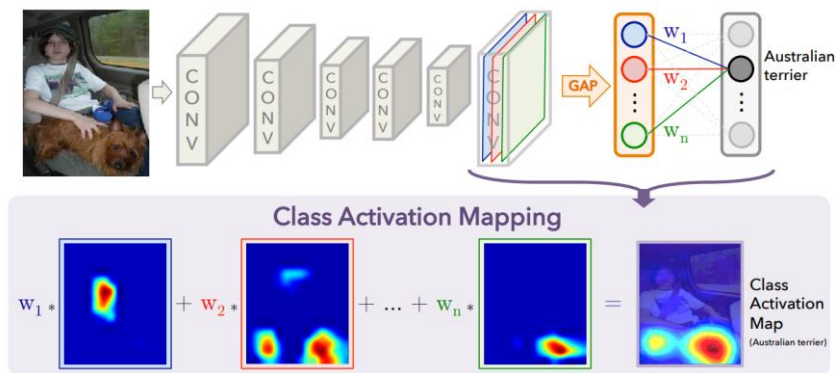


Snow



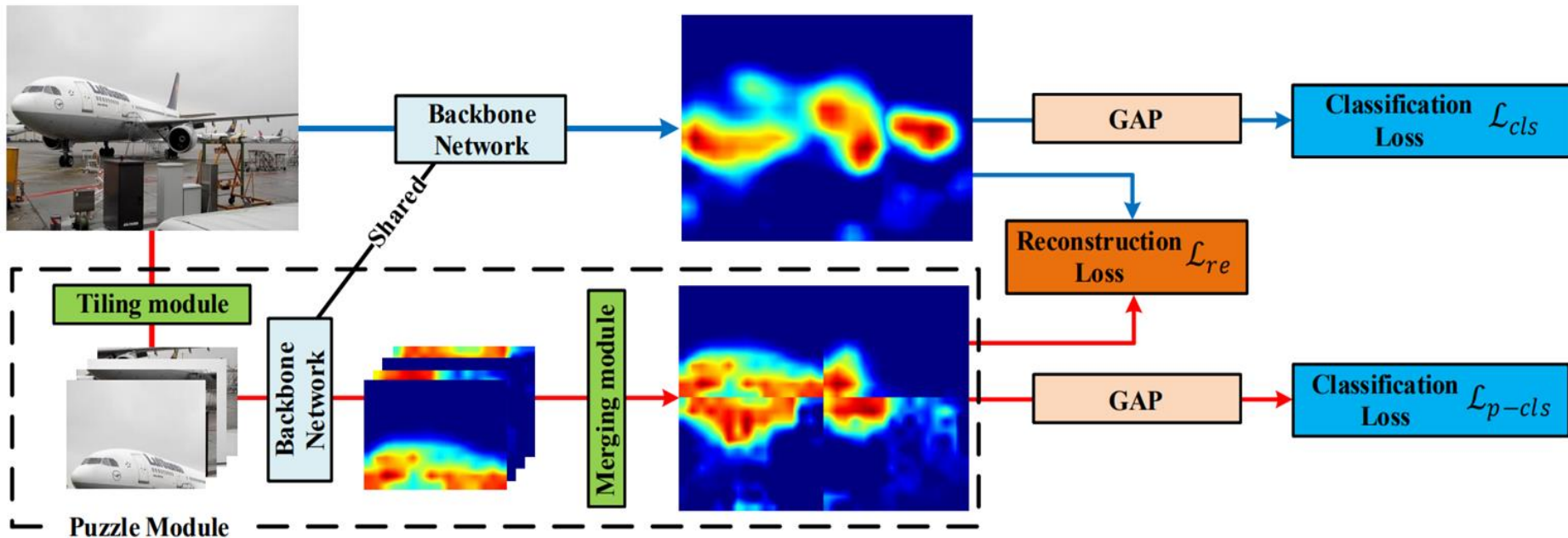
Innovations at OGQ: Vision AI

- ✓ Class Activation Map (CAM) highlights the most discriminative part of an object.
- ✓ Therefore, we aim to enlarge semantic pixels to replace ground truths with model outputs.



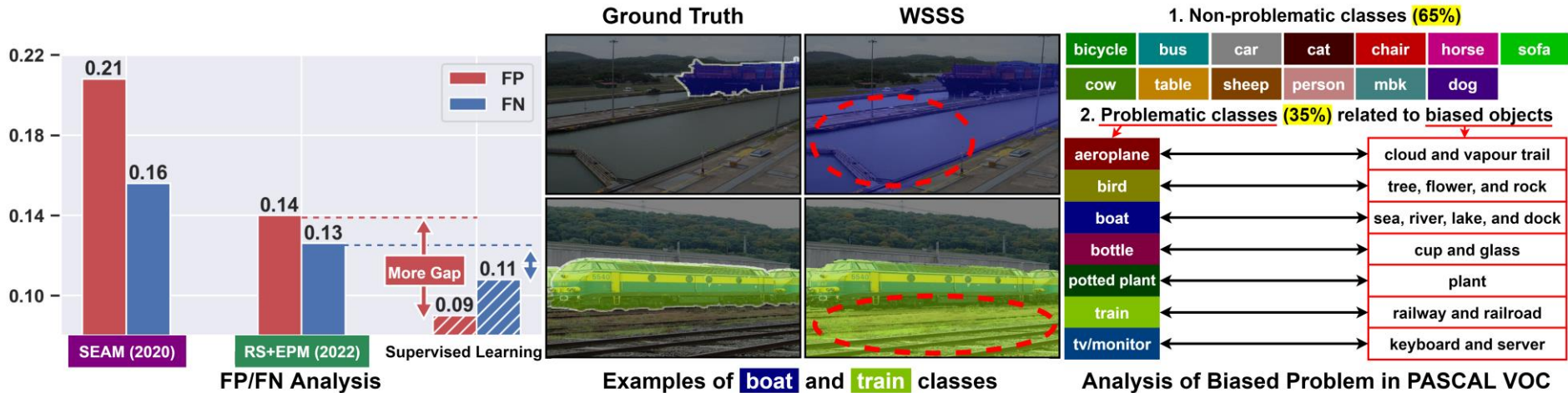
OGQ Vision AI / Puzzle-CAM (ICIP 2021, Citations: 74)

- ✓ Step 1: Tiling an image to image patches to divide into the attention.
- ✓ Step 2: Merging the feature maps from the network to produce the reconstructed features.
- ✓ Step 3: Matching partial and complete features with reconstructing regularization.



OGQ Vision AI / MARS (ICCV 2023, Citations: 3)

- ✓ Addressing the critical issue of False Positive (FP) in weakly supervised learning.
- ✓ 35% of classes in the VOC dataset are related to biased objects.



OGQ Vision AI / MARS (ICCV 2023, Citations: 3)

- ✓ Fully automatic elimination of biased objects without extra supervision or datasets.

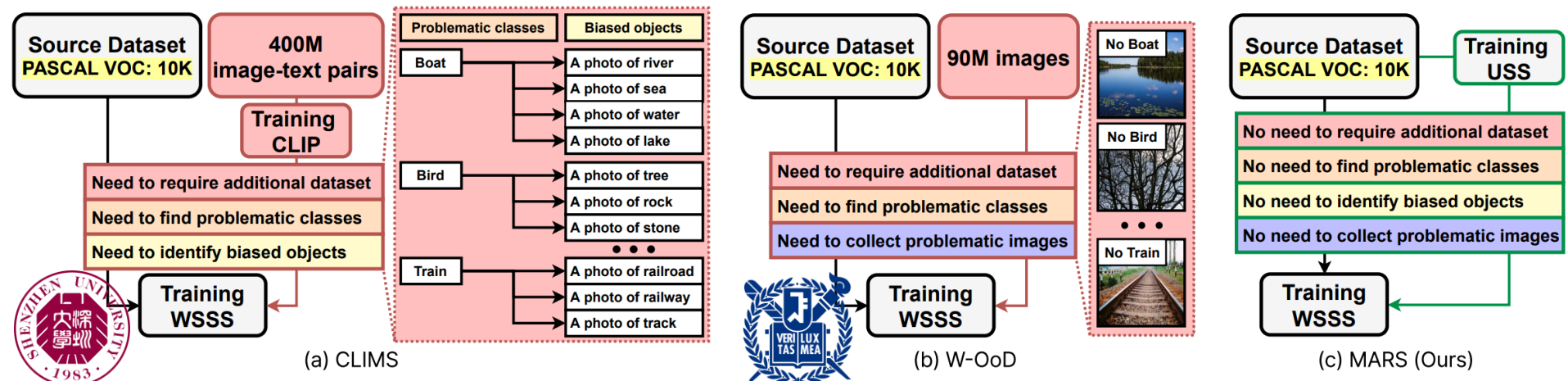


Figure 4. Conceptual comparison of three WSSS requirements. **(a)**: Using the CLIP’s knowledge trained on image-text pairs dataset alleviates the biased problem by finding problematic classes and identifying biased objects. **(b)**: Human annotators manually collect problematic images from the Open Images dataset [24] to train biased objects directly. **(c)**: The proposed MARS first applies an existing USS approach to remove biased objects without additional supervision, achieving the fully-automatic biased removal.

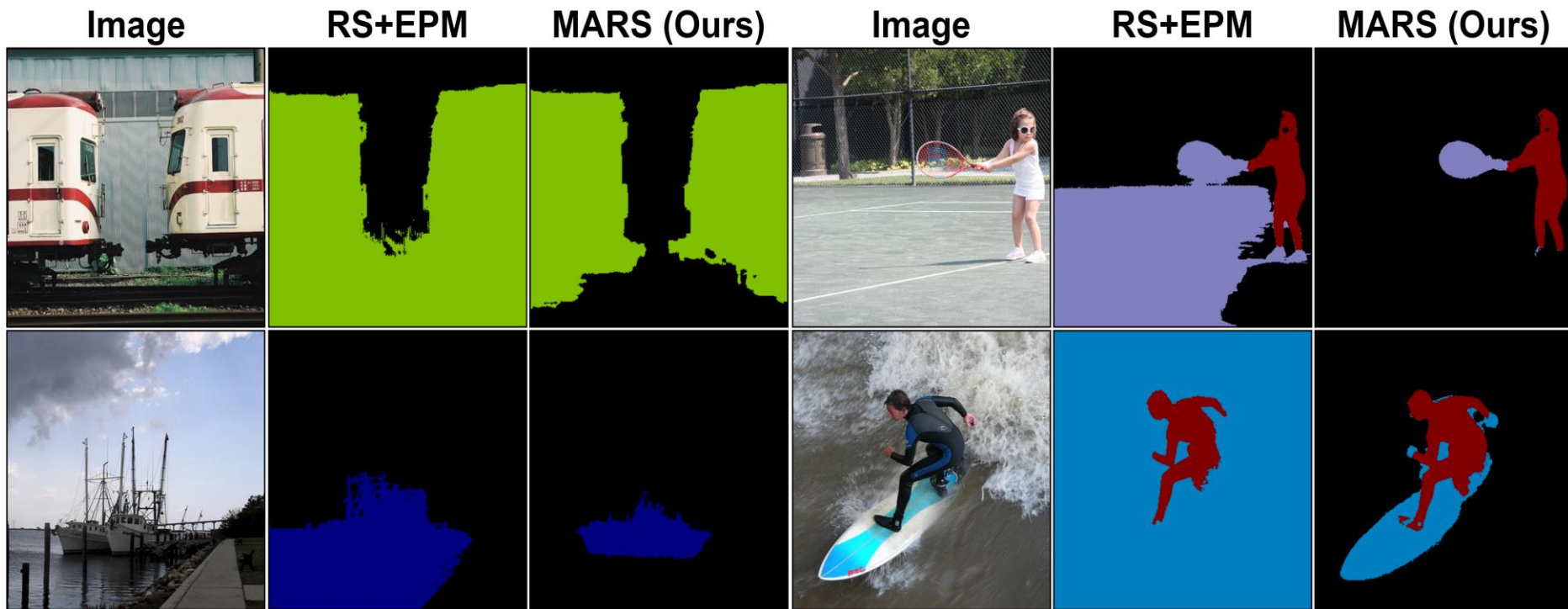
OGQ Vision AI / MARS (ICCV 2023, Citations: 3)

- ✓ Superior performance of our MARS (OGQ) in conventional benchmarks

Method	Backbone	Sup.	VOC		COCO
			<i>val</i>	<i>test</i>	<i>val</i>
DSRG CVPR'18 [16]	R101	$\mathcal{I}+\mathcal{S}$	61.4	63.2	26.0*
W-OoD CVPR'22 [29]	R101	$\mathcal{I}+\mathcal{D}$	69.8	69.9	-
L2G CVPR'22 [19]	R101	$\mathcal{I}+\mathcal{S}$	72.1	71.7	44.2
RCA CVPR'22 [58]	R101	$\mathcal{I}+\mathcal{S}$	72.2	72.8	36.8*
PPC CVPR'22 [12]	R101	$\mathcal{I}+\mathcal{S}$	72.6	73.6	-
SSDD ICCV'19 [43]	R101	\mathcal{I}	64.9	65.5	-
OAA ICCV'19 [18]	R101	\mathcal{I}	63.9	65.6	-
CONTA [56]	R101	\mathcal{I}	66.1	66.7	32.8
AdvCAM CVPR'21 [28]	R101	\mathcal{I}	68.1	68.0	-
RIB NeurIPS'21 [26]	R101	\mathcal{I}	68.3	68.6	43.8
AMN CVPR'22 [30]	R101	\mathcal{I}	69.5	69.6	44.7
SANCE CVPR'22 [32]	R101	\mathcal{I}	70.9	72.2	44.7†
RS+EPM Arxiv'22 [21]	R101	\mathcal{I}	74.4	73.6	46.4
MARS (Ours)	R101	\mathcal{I}	77.7	77.2	49.4
FSSS	R101	\mathcal{F}	80.6	81.0	61.8

OGQ Vision AI / MARS (ICCV 2023, Citations: 3)

- ✓ Our MARS automatically removes all biased objects in model predictions.



OGQ Vision AI / MARS (ICCV 2023, Citations: 3)

✓ Our paper is accepted to ICCV 2023 (2,162 accepted papers / 8,068 submitted papers)

Google Scholar

Top publications

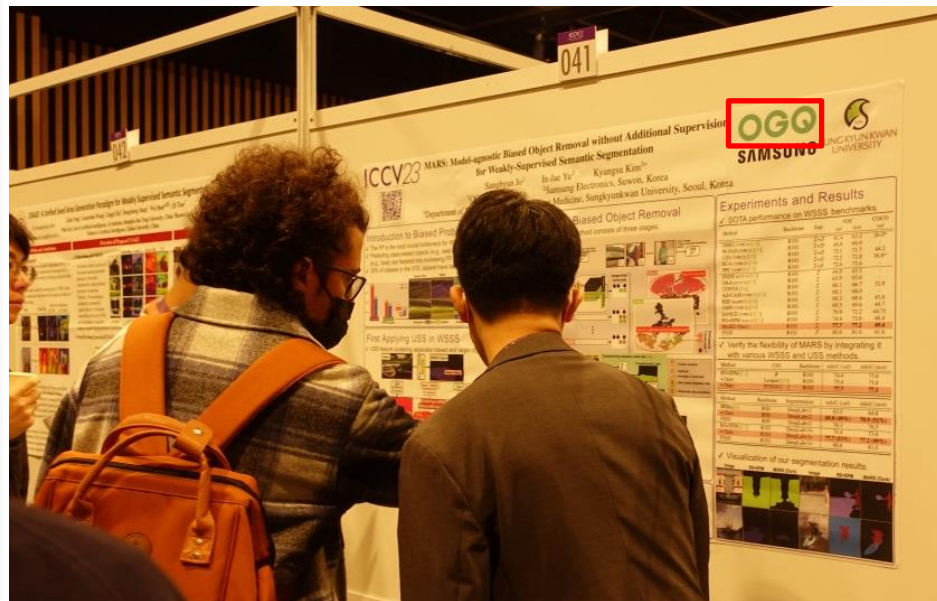
Categories > Engineering & Computer Science > Computer Vision & Pattern Recognition >

Publication	h5-index	h5-median
1. IEEE/CVF Conference on Computer Vision and Pattern Recognition	422	681
2. European Conference on Computer Vision	238	390
3. IEEE/CVF International Conference on Computer Vision	228	366
4. IEEE Transactions on Pattern Analysis and Machine Intelligence	179	318
5. IEEE Transactions on Image Processing	138	199
6. Pattern Recognition	111	160
7. IEEE/CVF Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)	108	176
8. Medical Image Analysis	103	171
9. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)	95	150
10. International Journal of Computer Vision	88	165
11. Pattern Recognition Letters	80	138
12. British Machine Vision Conference (BMVC)	77	128
13. IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)	66	102
14. IEEE International Conference on Image Processing (ICIP)	61	90
15. Asian Conference on Computer Vision (ACCV)	60	92
16. International Conference on Pattern Recognition	58	84
17. Journal of Visual Communication and Image Representation	51	73
18. Computer Vision and Image Understanding	49	94
19. International Conference on 3D Vision (3DV)	47	83
20. IEEE International Conference on Automatic Face & Gesture Recognition	47	79

Wednesday
4th 10:30
AM-12:30 PM

Segmentation,
grouping and
shape analysis

MARS: Model-agnostic Biased Object Removal without Additional Supervision for Weakly-Supervised Semantic Segmentation
Sanghyun Jo, In-Jae Yu, Kyungsu Kim
Paper ID:10811 - Location: Room "Nord" - 041



Generative AI at OGQ (Image)

Generative Models



NeRF



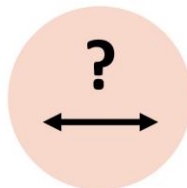
2D GAN



3D GAN



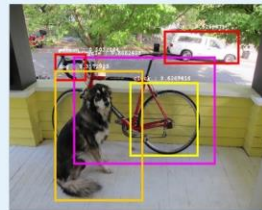
Text-to-Image



Computer Vision



Image Classification



Object Detection



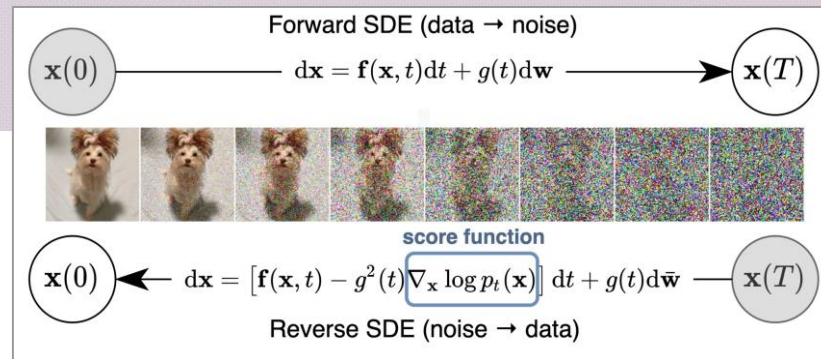
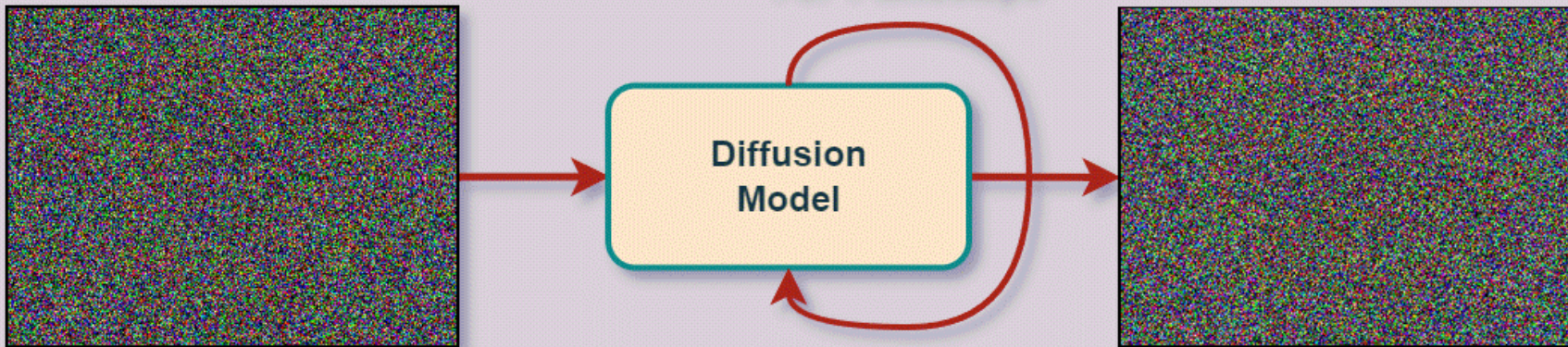
Semantic Segmentation



Pose Estimation

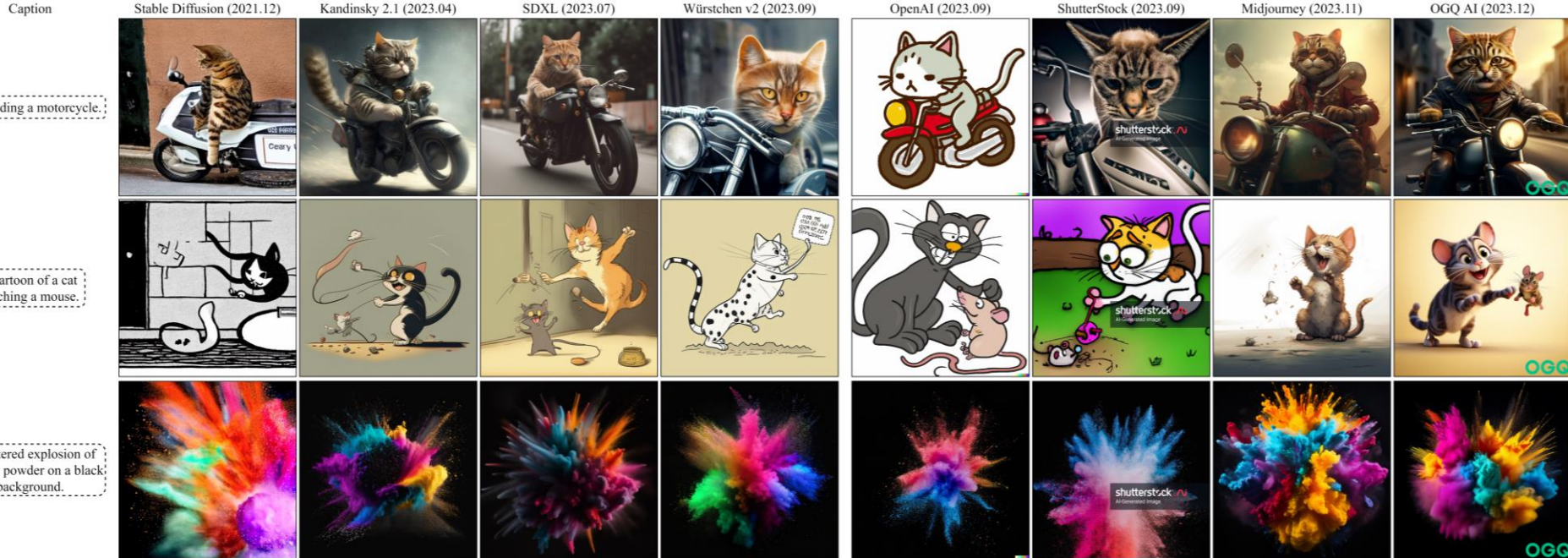
Generative AI at OGQ (Image)

For T timesteps



Generative AI at OGQ (Image)

✓ Qualitative analysis showcasing OGQ's edge over competitors and public models.



Generative AI at OGQ (Image)

- ✓ Breakthroughs in copyright object removal using text-driven inpainting models.



remove the gray kite at the left



remove the street light at the left



remove the man at the right of the man



remove the red car at the left of the tall ladder



remove the colorful train at the right



remove the boat at the right of the small boat

Generative AI at OGQ (Image)

- ✓ Integration of OGQ's labeling technology with leading conditional diffusion models.



A painting of a couple holding a yellow umbrella in a street on a rainy night. The woman is wearing a white dress and the man is wearing a blue suit.

DenseDiffusion **NAVER CLOVA**

StableDiffusion



An oil painting of a cute hamster holding a sunflower and a red rose.

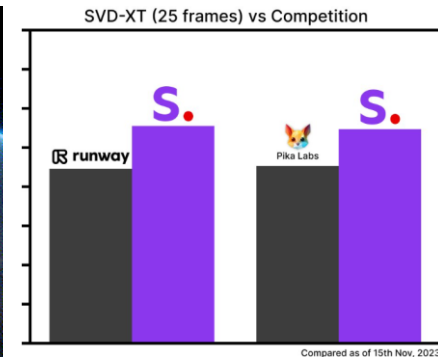
Generative AI at OGQ (Image)

- ✓ Support for OGQ creators in generating diverse content variants with OGQ AI models.



Generative AI at OGQ (Image)

- ✓ Stability AI releases Stable Video Diffusion (SVD), achieving state-of-the-art performance.

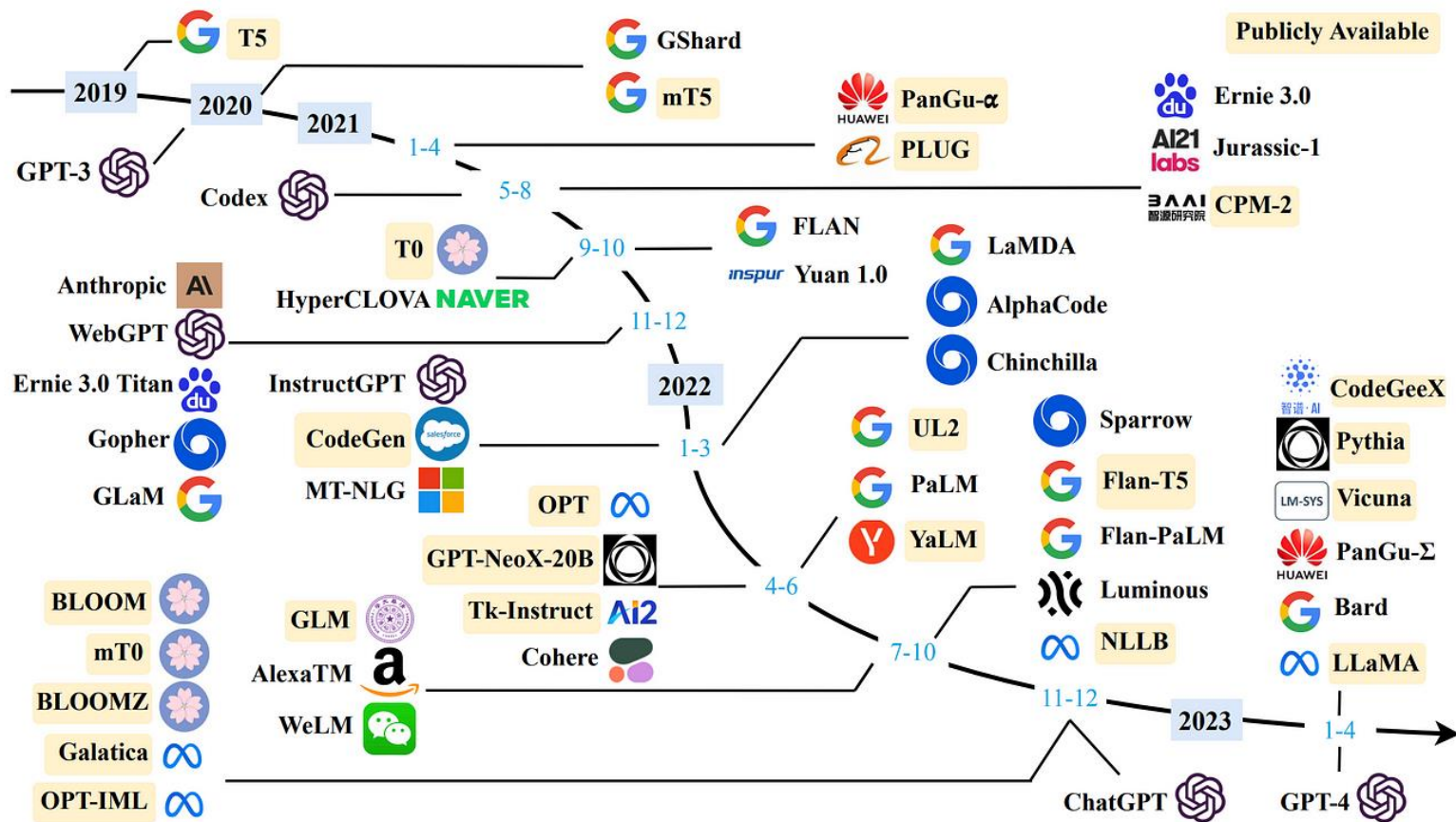


Method	LPIPS↓	PSNR↑	CLIP-S
SyncDreameer [55]	0.18	15.29	0.88
Zero123 [54]	0.18	14.87	0.87
Zero123XL [13]	0.20	14.51	0.87
Scratch-MV	0.22	14.20	0.76
SD2.1-MV [68]	0.18	15.06	0.83
SVD-MV (ours)	0.14	16.83	0.89

Table 1. Comparison of our dataset before and after filtering with publicly available research datasets.

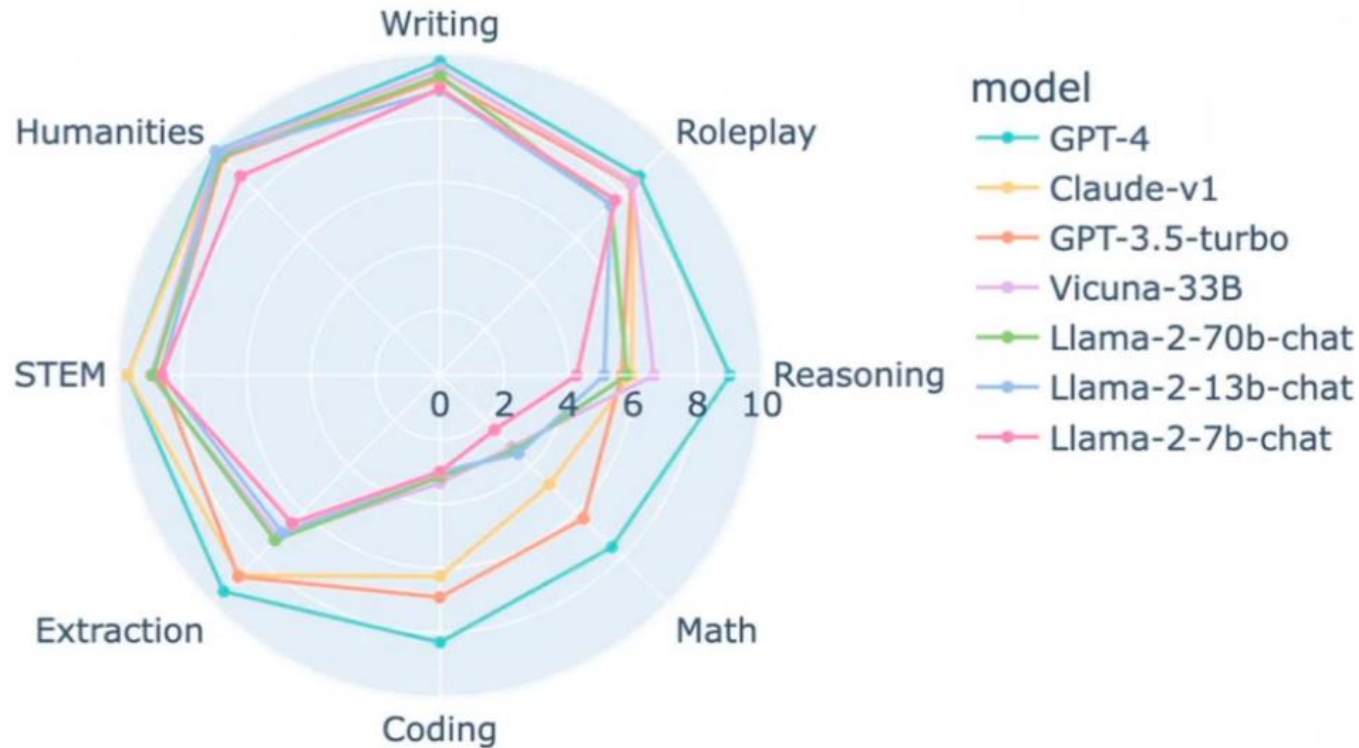
	LVD	LVD-F	LVD-10M	LVD-10M-F	WebVid	InternVid
#Clips	577M	152M	9.8M	2.3M	10.7M	234M
Clip Duration (s)	11.58	10.53	12.11	10.99	18.0	11.7
Total Duration (y)	212.09	50.64	3.76	0.78	5.94	86.80
Mean #Frames	325	301	335	320	-	-
Mean Clips/Video	11.09	4.76	1.2	1.1	1.0	32.96
Motion Annotations?	✓	✓	✓	✓	✗	✗

Generative AI at OGQ (Text)



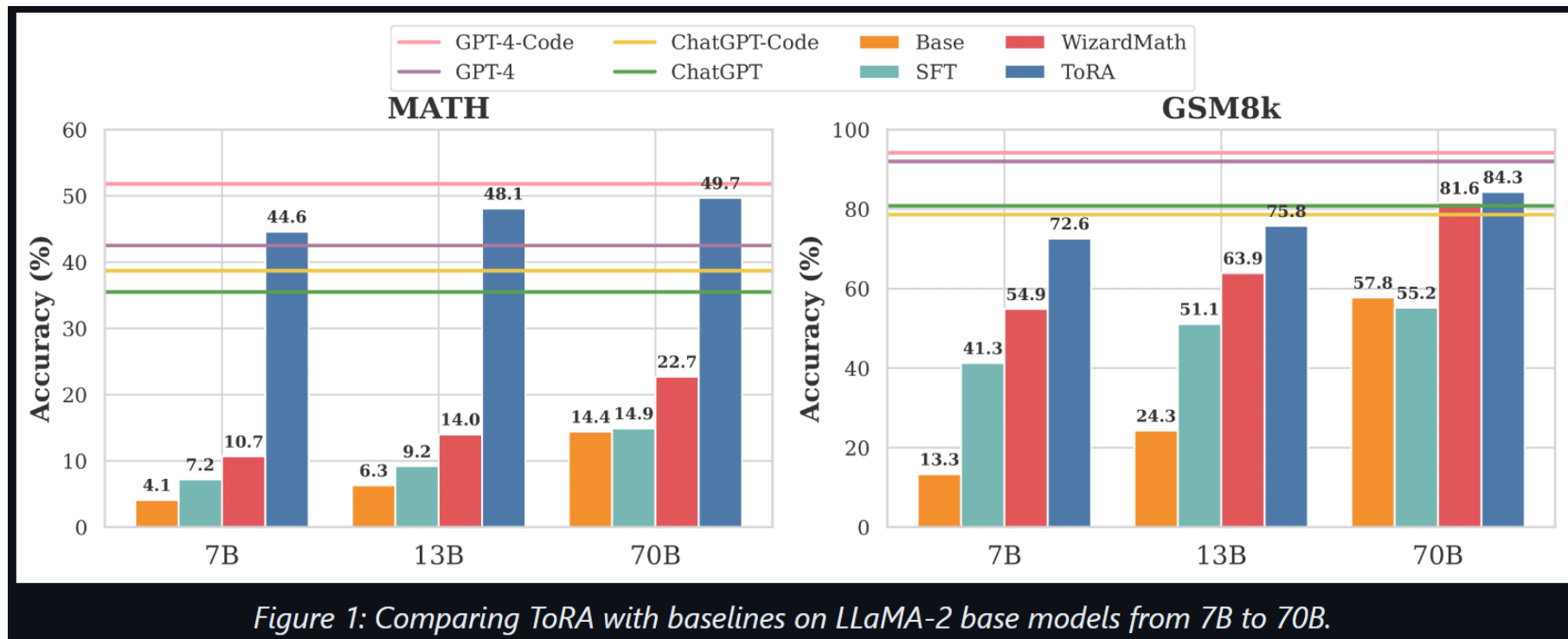
Generative AI at OGQ (Text)

- ✓ ChatGPT (i.e., GPT-4) significantly outperforms Reasoning, Math, and Coding performance versus open-source models.

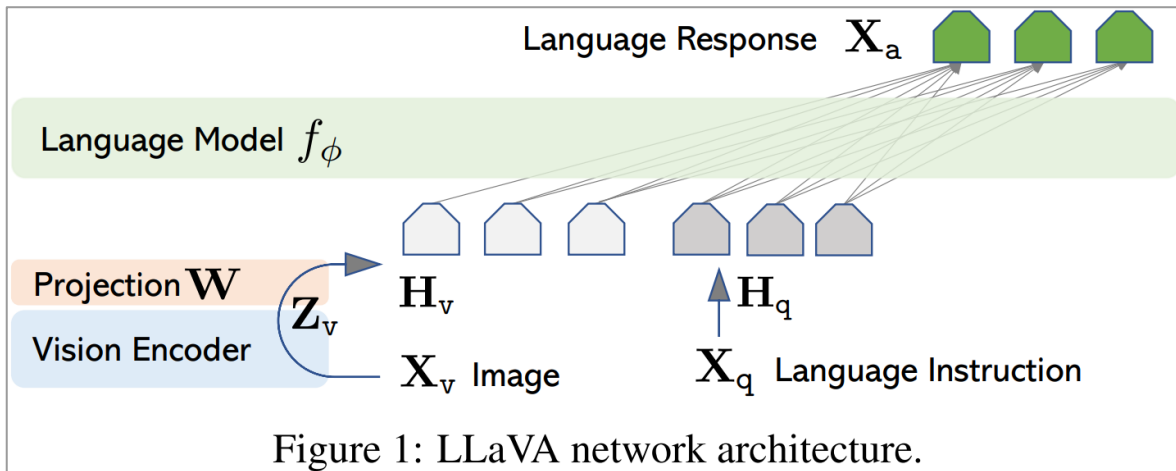
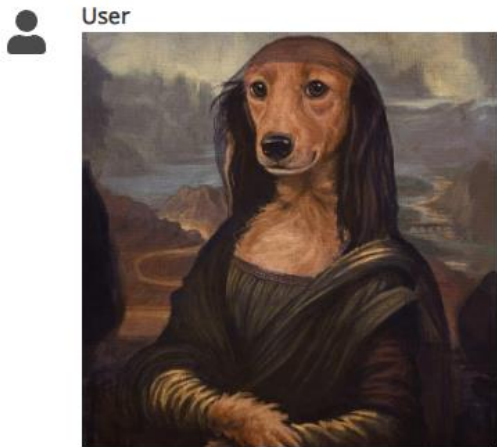


Generative AI at OGQ (Text)

- ✓ Closing the performance gap between GPT-4 and open-source models (e.g., ToRA).



Generative AI at OGQ (Text)



User

Do you know who drew this painting?

LLaVA

The painting depicts a dog in a humorous situation, where it appears to be wearing a woman's dress and standing in front of a landscape. Although the artist of the painting is not explicitly identified, it could be a reference to the famous artist Da Vinci, known for his artistic works, including the Mona Lisa. The painting is likely a creative and comical take on the traditional portrait style, replacing a human figure with a dog dressed in a woman's clothing, making it an interesting and amusing visual piece.

Generative AI at OGQ (Text)



User



User

What are the meals that I can cook with these?



LLaVA

With the variety of food items stored in the refrigerator, you can prepare several meals. For example, you can create a fruit salad using the strawberries, blueberries, and carrots. Additionally, you can make a delicious smoothie using the fruits and vegetables, which would be a healthy and nutritious snack or breakfast option. The milk and eggs can be used to create various dishes, such as a milk-based smoothie, a baked casserole, or scrambled eggs. The juice can also be used to make a fresh fruit and juice mix or as a refreshing homemade popsicle. The possibilities are vast, and the meals you can create will depend on your preferences and dietary needs.

Generative AI at OGQ (Text)



User

I would like to do the fruit salad. Show me the recipe.



LLaVA

Certainly! Here's a simple fruit salad recipe using the items in your refrigerator:

Ingredients:

- 2 cups fresh strawberries, sliced
- 1 cup fresh blueberries
- 1 large carrot, grated
- Juice of half a lemon
- A handful of chopped fresh parsley or mint
- A few ice cubes
- Optional: honey or maple syrup (adjust to taste)

Instructions:

1. In a large bowl, combine the sliced strawberries, blueberries, and carrot. Gently toss them together to mix.
2. Drizzle the juice of half a lemon over the fruit and vegetables to prevent them from oxidizing and add a refreshing citrus flavor.
3. Sprinkle the chopped fresh parsley or mint over the fruit and gently toss once more.
4. If desired, add a drizzle of honey or maple syrup for a touch of sweetness. Adjust the amount according to your taste preferences.
5. Cover the bowl and refrigerate the fruit salad for at least 30 minutes or up to a few hours. This allows the flavors to meld and creates a more refreshing salad.
6. Serve the fruit salad in cups or bowls, garnished with a sprinkle of fresh parsley or mint, if desired.

This fruit salad recipe is perfect for a healthy snack or side dish, and with the variety of ingredients available in your refrigerator, it should be easy to put together quickly.

Future Trajectory at OGQ: Unimodal AI

1) Cross-Modal Retrieval

Audio



Crackle of a Fire

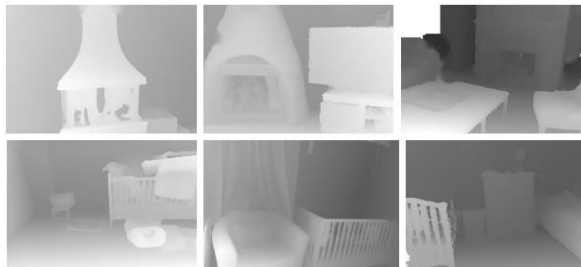


Images & Videos



Baby Cooing

Depth



Text

“A fire crackles while a pan of food is frying on the fire.”

“Fire is crackling then wind starts blowing.”

“Firewood crackles then music...”

“A baby is crying while a toddler is laughing.”

“A baby is laughing while an adult is laughing.”

“A baby laughs and something...”

2) Embedding-Space Arithmetic



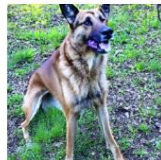
Waves



3) Audio to Image Generation



Dog



Engine



Fire

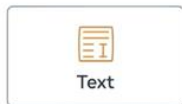


Rain

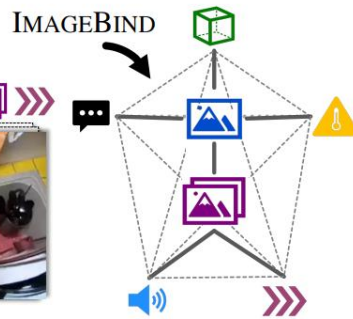
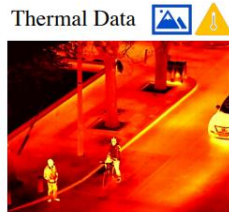


Figure 1. IMAGEBIND’s joint embedding space enables novel multimodal capabilities. By aligning six modalities’ embedding into a common space, IMAGEBIND enables: **1)** Cross-Modal Retrieval, which shows *emergent* alignment of modalities such as audio, depth or text, that aren’t observed together. **2)** Adding embeddings from different modalities naturally composes their semantics. And **3)** Audio-to-Image generation, by using our audio embeddings with a pre-trained DALLE-2 [61] decoder designed to work with CLIP text embeddings.

Future Trajectory at OGQ: Unimodal AI



Meta AI



Future Trajectory at OGQ: Unimodal AI

Objects and Attributes

- 1 dog, pub dog, a brown and white dog
- 2 dog collar, black color, chain collar
- 3 bell, cowbell
- 4 steps, stairs, the steps of a building
- 5 sack, a large white bag with black writing



Relationships and Landmarks

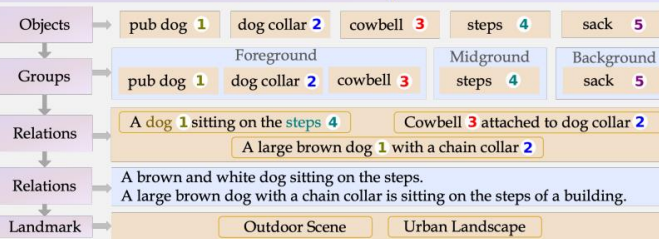
A dog sitting on the steps

A large brown dog wearing a chain collar

Cowbell attached to dog collar

Landmarks: Outdoor - Urban Landscape

Scene Graph



Dense Grounded Caption

A large brown dog is sitting on the steps of a building. It is wearing a black chain dog collar. The collar has a cowbell attached to it. There is a bag in the background with black writings on it.

Extra Context

Dogs, especially pugs and bulldogs, have been a part of human families for thousands of years, serving as loyal companions. They have been bred for specific traits, making them popular pets. Dogs have been trained for various tasks, including assisting people with disabilities and serving as search and rescue animals. Dog collars, often bearing identification tags, are essential for keeping pets safe and ensuring they can be returned home if lost. Cowbells, once used to signal the arrival of a cow, have been repurposed as dog collars, providing a distinct sound to help locate a dog if it wanders off. In outdoor urban landscape, dogs are often found sitting on steps, as they may choose to rest in spots that offer a good view of their surroundings.

Level-1

Object localization and attributes

- Image Tagging and Object Detection
- Open Vocabulary Detection
- Region Attribute Detection

Level-2

Relationships

- Short Captions and Phrase extraction
- Grounding expression
- Landmarks

Level-3

Scene Graph & Dense Captioning

- Hierarchical Scene Graph
- In-contex Learning with LLM
- Verification Pipeline

Level-4

Extra Contextual Insights

- Landmark Details
- History and Background
- Precautionary Measures

Q & A

shjo.april@ogqcorp.com

ICCV23 MARS
PARIS



Puzzle-CAM



RSEPM

