GOO: A Dataset for Gaze Object Prediction in Retail Environments

Henri Tomas[1]*, Marcus Reyes[1]*, Raimarc Dionido[1]*, Mark Ty[1], Jonric Mirando[1], Joel Casimiro[1], Rowel Atienza[1], and Richard Guinto[2]

[1]University of the Philippines, [2]Samsung R&D Institute Philippines

University of the Philippines

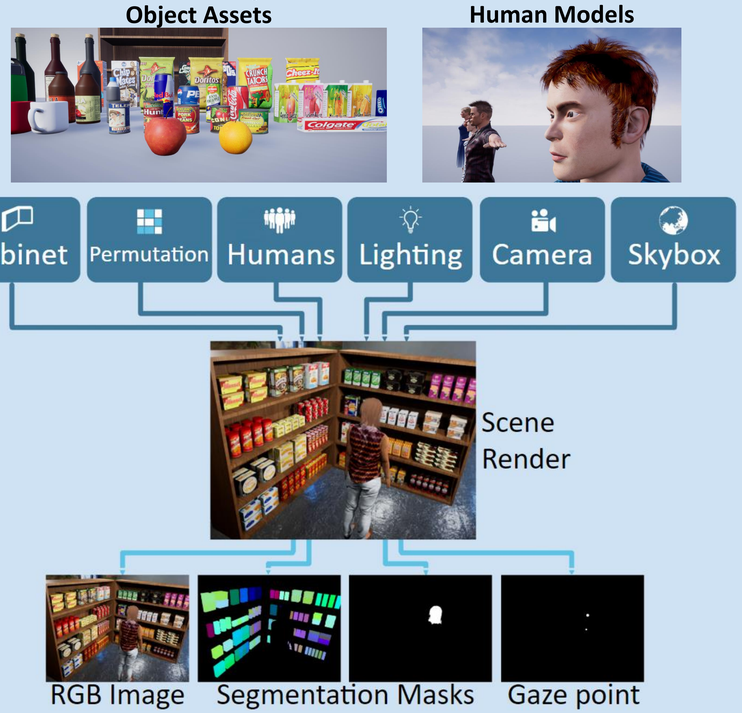
Samsung Research

CVPR VIRTUAL JUNE 19-25

Why it matters?

- **Gaze object prediction** = gaze following with object detection
- **GOO** is:
  - a **new dataset** for gaze following and gaze object prediction.
  - a combination of **synthetic + real images**
  - **considerably larger** than existing datasets.
- Applications: **gaze following**, gaze object prediction, object detection, segmentation, and **domain adaptation**

GOO-Synth Data Generation



GOO-Real Data Collection and Annotation

- Volunteers were instructed to look at specific items in the scene
- Ground truth was manually annotated based on the instructions.

Dataset Comparison

	Ground Truth	Perspective	Size
iSUN	Point	1st Person	20,000
SALICON	Point	1st Person	10,000
CAT2000	Point	1st Person	4,000
EYEDIAP	Point	2nd Person	N/A
GazeFollow	Point	3rd Person	122,143
GOO (Ours)	**Object**	**3rd Person**	**201,552**

	GazeFollow	GOO
Size	122,143	201,552 (192,000 synth + 9,552 real)
Type	Real	Synthetic & Real
Annotations	Head Bbox, Gaze point	Head Bbox, Gaze object Object Segmentation
Context	Varied	Retail
People per Image	Varied	1
Objects per Image	Few	Many
Domain Adaptation?		✓

Annotations

Gaze Following and Domain Adaptation

Key Takeaways

- GOO hopefully inspires novel architectures and training methods for gaze systems to infer the specific object being looked at.
- Gaze object prediction still lacks metrics for measuring performance including correctness of the bounding box and the class of the object

GOO Github