ZeroKey: Point-Level Reasoning and Zero-Shot 3D Keypoint Detection from Large Language Models

Supplementary Material

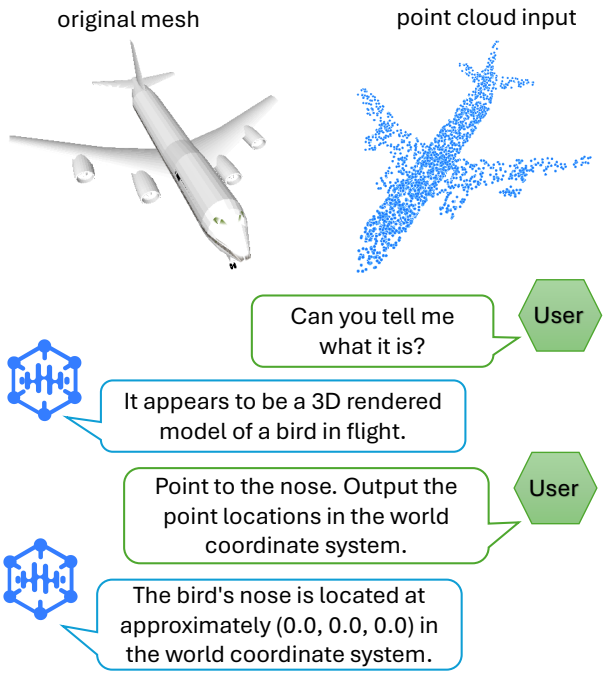

Figure 1. Qualitative comparison of keypoint localization using ShapeLLM on out-of-distribution 3D airplane shape. ShapeLLM struggles to accurately identify the shape and point to keypoints on this unseen pointcloud data.

1. 3D Multimodal Language Models

Recent general-purpose 3D language models (LLMs) are usually trained with specific objectives that align 3D data representations with vision-language models, allowing for global tasks like classification. However, these models are based on standard vision-language frameworks that have limited localization capabilities. Consequently, they are unable to accurately perform 3D keypoint localization based on language prompts. We evaluated the KeypointNet model against the state-of-the-art ShapeLLM [6]. The qualitative results are shown in Fig. 10 in the main paper. We observed that 3D LLMs are often sensitive to the alignment and rotation of the input point cloud and do not generalize well to new datasets. Even with perfectly aligned point clouds, we found that ShapeLLM still struggled to accurately understand out-of-distribution shapes and localize keypoints when the shapes were not seen during its training phase. These limitations highlight the challenges faced by current 3D language models in precise localization and generalization to unseen data in zero-shot settings.

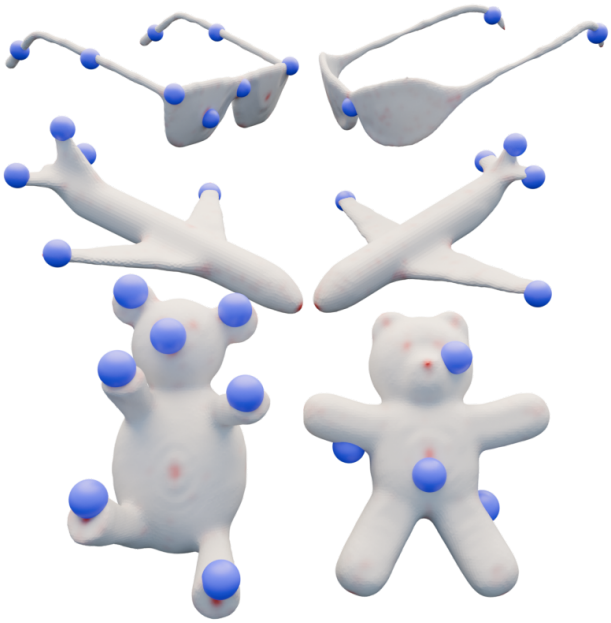

Figure 2. Superposition of the keypoints predicted by ZeroKey (blue) and the Schelling point distribution [3] (red).

2. Schelling Points

We make the hypothesis that the salient points selected by individuals in visual scenes often correspond to objects or features with clear, identifiable names in natural language. This leads us to believe that an effective Visual Language Model (VLM) should be capable of predicting such salient points by leveraging its combined understanding of language and vision.

To test this hypothesis, we conducted experiments using the dataset from Chen et al. [3], which captures human-selected Schelling points in various visual scenarios. In the context of coordination games, Schelling points refer to prominent locations or choices that people commonly select without prior communication, relying on mutual expectations and shared cultural references.

In Fig. 2, we showcase a visualization that superposes the distributions reported by the Schelling Points paper [3] to a set of salient points generated by ZeroKey, where the candidate keypoints were provided by ChatGPT using the following prompt:

"List possible salient keypoints (in text)."

Besides text prompt, we also input renderings of the 3D

scene to ChatGPT to obtain prompts simulates natural language descriptions that humans are likely to use. This process can simulate the way humans might describe these scenes in natural language, capturing the linguistic cues that people are likely to use when identifying salient features.

We see in this image that some of the salient points [3] do correspond to the predictions made by ZeroKey. Our findings reveal that several of the salient points identified by humans correspond to the predictions made by ZeroKey. These results support our hypothesis that salient points in visual scenes are closely tied to language and that VLMs that harness this connection can more effectively predict human-selected points of interest.

3. Point Describability and Consistency

In this section we present the algorithm used for the evaluation of the task of point describability and consistency, see Alg. 1.

Algorithm 1 Shape Selection and Saliency Prediction

Require: Random shape S, and a point $p \in S$. ▷ Choose a point $p \in S$ with 1) the highest reported saliency value; or 2) without semantic meaning (saliency value is 0).

Ensure: Absolute error shows that if p is the saliency point

1: Place a red marker at p on S, render several views.
2: Ask Molmo to describe the p with a short name l, retaining the most repeated prompt.
3: Render new views of S without any marker. ▷ Query these views with prompt l.
4: Infer a 3D point prediction $\hat{p} \leftarrow \text{ZeroKey}(S, l)$.
5: Store the absolute error $e \leftarrow |\hat{p} - p|$.

4. Setup and Dataset

We evaluate our method using the KeypointNet dataset. We adhered to the experiment setting of prior research, concentrating on the same three classes from the dataset during our evaluation: airplane, chair, and table. Our evaluation strategy follows the approach of You et al. [13], which computes the Intersection over Union (IoU) between predicted keypoints and ground-truth keypoints from the KeypointNet dataset, using varying distance thresholds. A match is counted if the geodesic distance between a ground-truth keypoint and a predicted keypoint is less than the specified threshold. The labels in KeypointNet only provide integer IDs for each keypoint. To detect keypoints from open vocabulary textual descriptions and evaluate them using KeypointNet, we manually annotated the labels of keypoint classes in the dataset. We then used this annotated text as a textual prompt for zero-shot 3D keypoint detection. Please refer to Sec. 8 for the exact prompt we used in each category.

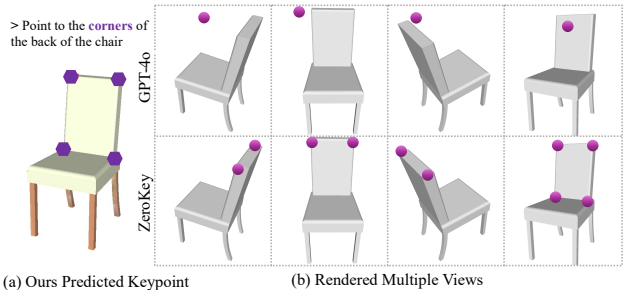

Figure 3. Renderings from the ablation study where the MLLM backbone is replaced with GPT-4o. As an MLLM trained solely on image-level tasks, it struggles with point-level reasoning, leading to less accurate and consistent keypoints compared to Molmo.

5. More Qualitative and Quantitative Analysis

In this supplementary document, we present additional quantitative evaluations. We compared our Zero-Shot method with both supervised and few-shot methods trained on KeypointNet. Our results demonstrate that the Zero-Shot method can even outperform some of the supervised methods. The IoU levels achieved by our method are comparable to those of the reference-based Few-Shot methods and the supervised methods specifically designed for this dataset. Qualitative results are provided in Table 3. Our method is evaluated using 26 rendered views, and we have also included the results from the same number of views of the few-shot method B2-3D as a reference.

Per-class breakdown As shown in the Table 1, ZeroKey performs well in different categories.

Thresh.	0.05	0.06	0.07	0.08	0.09	0.10
Airplane	42.45	49.37	58.75	66.64	74.59	81.13
Chair	52.17	58.44	63.21	65.94	68.78	71.85
Table	75.17	78.36	80.63	82.65	83.87	85.30

Table 1. Per-class threshold percentages breakdown

Moreover, we show case now several qualitative examples. Fig. 1 presents an example of keypoint localization using ShapeLLM. Then, on Fig. 3 we ablate our method by replacing Molmo with GPT-4o, we see that training with point-level reasoning enables accurate and consistent keypoints. Finally, Fig. 6 showcases a side-by-side comparison of ground truth annotations and our predictions.

6. More analysis on patchify and clustering

Patchify is based on our observation that the tokens generated by the VLMs follow a probability distribution close to the output coordinates. To aggregate this probability distribution across multiple views, we back-project patches and cluster them on the mesh based on their soft-voting weights.

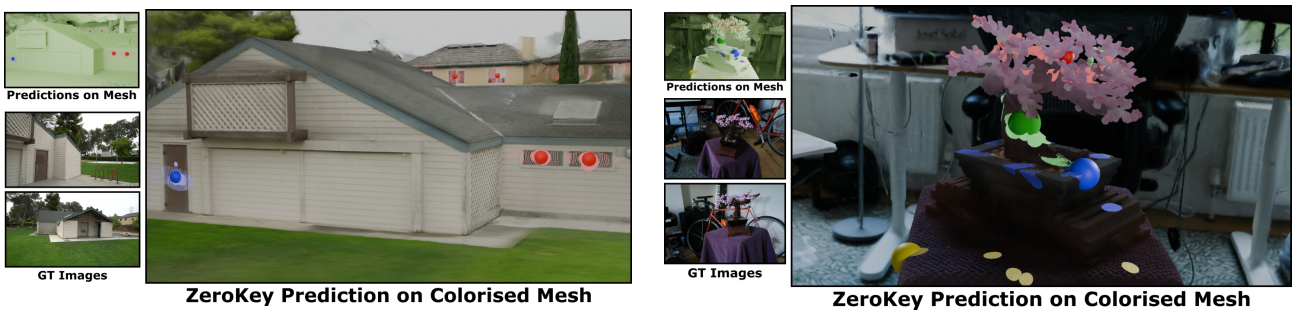

Figure 4. ZeroKey performs well on real-world data. We apply it to a mesh reconstructed from ground-truth images and query for **windows** and **doors** on the left; for **flowers**, **tree trunk**, **container** and table on the right. Small dots are corresponding patchify back-projected points. The method remains effective despite imperfections in the reconstruction. We will include more real-world scenes in the paper.

Token	4	4	.	6	”	4	0	.	0	”
Top 2	3	3	,	8	y	3	1	,	1	x
Top 3	5	5	=”	9	,”	6	2	=”	2	.”
Top 4	2	6	on	5	”	5	3	that	4	,”
Top 5	1	2	that	0	=”	2	4	in	3	)

Table 2. Top 5 Next-Token Predictions in Molmo Generation

In Sec. 6.4 and Fig. 7 of the paper, we illustrate the effects of removing the patchify process (see “Binary Back-projection”) and the clustering step (“w/o HDBSCAN Clustering”). Here, we have included a distribution of keypoint predictions following the patchify back-projection and prior to the clustering step in real-world cases (see Fig. 4 and 5).

7. Human3.6m dataset evaluation

In Fig. 5, we show results for actors 1, 6, 8, and 11 out of a total of 11 actors in the Human3.6m dataset. ZeroKey performs well on the scanned meshes.

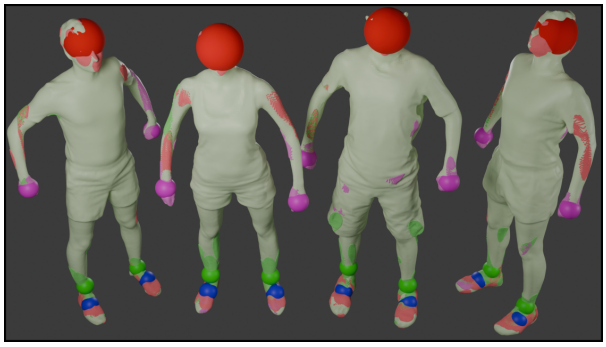

Figure 5. Human Dataset. Big dots: **Head**, **Hands**, **Ankles**, **Feet**. Small dots: Corresponding patchify back-projected points.

8. Keypoint Descriptions for Categories

In this supplementary material, we present a comprehensive text prompt outlining key points for 16 common object categories used in KeypointNet. Each object category is defined by a specific set of semantically meaningful manually annotated key points that capture its structural and functional characteristics. These key points are carefully selected to represent distinctive geometric features and important structural elements of each object. However, in KeypointNet, these key points are not accompanied by a text description; instead, they are assigned a unique semantic index. Some of these labels are derived from the source referenced in [11]. This information may also be useful for others aiming to detect key points based solely on textual descriptions.

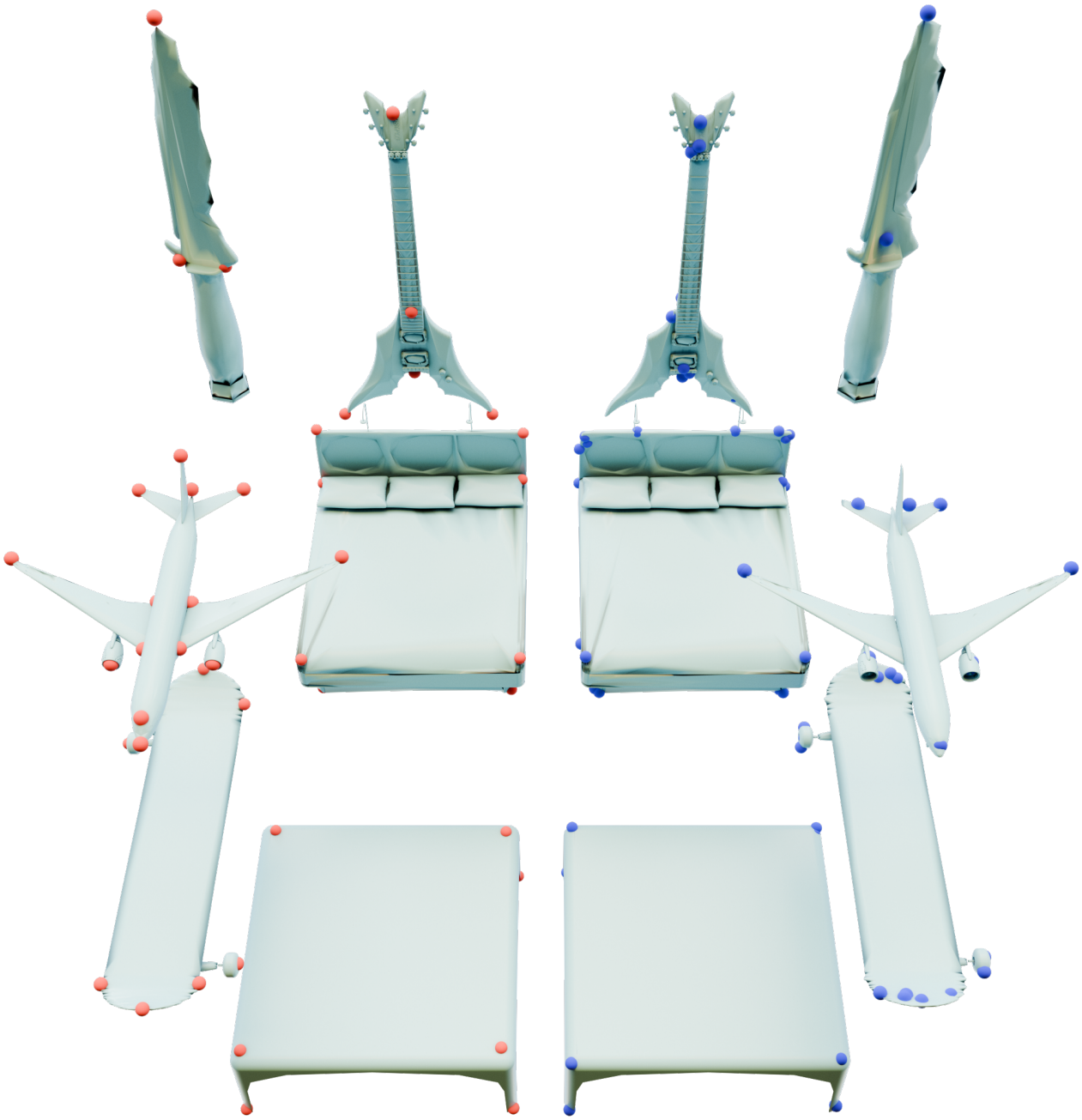

Figure 6. Comparing the ground truth KeypointNet dataset annotations (in red) to our method's predictions (in blue). This figure showcases our results on the KeypointNet dataset, illustrating the effectiveness of our approach in keypoint detection. The close alignment of the red and blue dots demonstrates the effectiveness of our approach in accurately detecting keypoints and highlights its precision in point-level reasoning.

Method	IoU (in %)										
Distance Thresholds	0.001	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
HARRIS-3D [9]	0.15	0.76	2.19	3.96	6.16	8.88	11.91	15.13	18.63	22.10	25.69
SIFT-3D [7]	0.29	1.05	2.62	4.38	6.65	9.42	12.38	15.65	19.39	22.71	26.15
ISS [14]	0.49	1.10	2.69	4.67	7.27	10.14	13.29	16.56	20.02	23.58	27.20
USIP [5]	0.83	1.70	3.25	5.24	8.00	11.76	15.98	20.63	25.56	30.20	33.67
D3FEAT [2]	2.36	3.86	7.27	10.37	13.37	17.12	21.41	26.12	30.39	34.82	38.41
UKPGAN [13]	3.95	6.54	12.27	17.86	22.42	26.55	30.28	34.32	38.46	42.65	46.49
FPS	6.34	10.55	17.68	23.18	26.91	30.27	33.57	36.85	40.47	44.42	48.38
FSKD [1]	7.00	7.94	11.17	16.74	23.99	31.14	38.14	43.97	49.30	53.62	57.03
B2-3D (62 views)	12.41	20.29	35.55	46.25	52.74	57.72	61.61	64.19	66.56	68.55	70.57
B2-3D (26 views)	8.60	14.39	25.10	33.31	39.29	44.11	47.85	50.83	53.17	55.11	56.90
PaliGemma 2 [10]	0.00	0.34	1.03	2.98	4.25	5.70	6.62	8.17	8.92	11.49	11.65
RedCircle [8]	0.21	0.34	0.64	1.16	1.90	3.05	4.81	7.55	11.06	14.92	18.50
GPT-4o	0.18	0.48	1.20	2.61	4.19	6.04	8.48	10.89	14.03	17.03	20.73
CLIP-DINOiser [12]	0.73	1.41	3.00	4.94	7.31	9.80	12.66	15.52	18.52	21.76	25.56
StablePoints [4]	3.66	5.80	9.94	13.32	16.58	19.91	23.48	27.26	31.28	34.89	38.22
Binary Back-projection	3.69	5.84	10.91	17.58	23.72	29.81	35.52	40.82	45.72	50.24	54.64
ZeroKey (Ours)	**9.12**	**13.16**	**24.01**	**37.43**	**48.43**	**56.60**	**62.06**	**67.53**	**71.74**	**75.75**	**79.43**

Table 3. Comparison of IoU between the predicted and ground-truth keypoints from KeypointNet using different methods across various geodesic distance thresholds. The IoU of our method is highlighted in bold.

Object	Index	Description	Short Description
Table	0	corner of the table top	different corners
	1	corner of the table top	different corners
	2	corner of the table top	different corners
	3	corner of the table top	different corners
	4	end of the table leg	end of the legs
	5	end of the table leg	end of the legs
	6	end of the table leg	end of the legs
	7	end of the table leg	end of the legs
	8	end of the stretcher between the table legs	-
	9	end of the stretcher between the table legs	-
	10	end of the stretcher between the table legs	-
	11	end of the stretcher between the table legs	-
Mug	0	mug rim	-
	1	mug rim	-
	2	mug rim	-
	3	mug rim	-
	4	mug base rim	-
	5	mug base rim	-
	6	mug base rim	-
	7	mug base rim	-
	8	upper attachment point for the mug handle	top of the handle
	9	center of the mug handle	handle
	10	lower attachment point for the mug handle	bottom of the handle
Bottle	0	lower rim of the bottle	-
	1	lower rim of the bottle	-
	2	lower rim of the bottle	-
	3	lower rim of the bottle	-
	4	edge between the bottle body and the bottle neck	-
	5	edge between the bottle body and the bottle neck	-
	6	edge between the bottle body and the bottle neck	-
	7	edge between the bottle body and the bottle neck	-
	8	bottle cap center	-
	9	edge of the bottle cap	-
	10	edge of the bottle cap	-
	11	edge of the bottle cap	-
	12	edge of the bottle cap	-
	13	lower edge of the bottle neck	-
	14	lower edge of the bottle neck	-
	15	lower edge of the bottle neck	-
	16	lower edge of the bottle neck	-
	17	lower edge of the bottle neck	-

Object	Index	Description	Short Description
Vessel	0	bow of a boat	front
	1	lower part of the end of the bow of a boat (starboard)	bottom front
	2	upper part of the end of the bow of a boat (starboard)	-
	3	upper part of the end of the bow of a boat (port)	-
	4	lower part of the end of the bow of a boat (port)	bottom front
	5	lower corner of the boat stern (starboard	back
	6	upper corner of the boat stern (starboard)	back left corner
	7	upper corner of the boat stern (port)	back left corner
	8	lower corner of the boat stern (port)	back
	9	mast step of a sailboat	-
	10	mast top of a sailboat	-
	11	lower front corner of the deckhouse (starboard)	right front of the cabin
	12	lower front corner of deckhouse (port)	left front of the cabin
	13	lower aft corner of the deckhouse (starboard)	-
	14	lower aft corner of the deckhouse (port)	-
	15	upper front corner of the deckhouse (starboard)	-
	16	upper front corner of the deckhouse (port)	-
	17	upper aft corner of the deckhouse (starboard)	-
	18	upper aft corner of the deckhouse (port)	-
	19	foremost elevation point on the deck of the ship	-
	20	lower corner of the windshield of the ship (starboard)	-
	21	lower corner of the windshield of the ship (port)	-
	22	upper corner of the windshield of the ship (starboard)	-
	23	upper corner of the windshield of the ship (port)	-
Chair	0	upper corner of the back of the chair	top corners of the back rest
	1	upper corner of the back of the chair	top corners of the back rest
	2	lower corner of the back of the chair	different corners of the seat
	3	lower corner of the back of the chair	different corners of the seat
	4	front corner of the seat of the chair	different corners of the seat
	5	front corner of the seat of the chair	different corners of the seat
	6	front corner of the armrest of the chair	end of both armrests
	7	back corner of the armrest of the chair	end of both armrests
	8	front corner of the armrest of the chair	end of both armrests
	9	back corner of the armrest of the chair	end of both armrests
	10	lower end of the front leg of the chair	tip of all legs
	11	lower end of the back leg of the chair	tip of all legs
	12	lower end of the front leg of the chair	tip of all legs
	13	lower end of the back leg of the chair	tip of all legs
	14	lower end of the leg of the chair	tip of all legs
	15	top of the chair cylinder	cylinder
	16	bottom of the chair cylinder	cylinder
	17	lower end of the front leg of the chair	-
	18	lower end of the back leg of the chair	-
	19	lower end of the back leg of the chair	-
	20	lower end of the front leg of the chair	-

Object	Index	Description	Short Description
Bed	0	lower end of the front leg of the bed	-
	1	lower end of the headboard leg of the bed	-
	2	lower end of the headboard leg of the bed	-
	3	lower end of the front leg of the bed	-
	4	foot corner of the mattress	-
	5	head corner of the mattress	-
	6	head corner of the mattress	-
	7	foot corner of the mattress	-
	8	foot corner of the upper mattress of the bunk bed	-
	9	head corner of the upper mattress of the bunk bed	-
	10	head corner of the upper mattress of the bunk bed	-
	11	foot corner of the upper mattress of the bunk bed	-
	12	upper corner of the bed frame of the bunk bed	-
	13	upper corner of the bed frame of the bunk bed	-
	14	upper corner of the bed frame of the bunk bed	-
	15	upper corner of the bed frame of the bunk bed	-
	16	upper corner of the footboard of the bed	-
	17	upper corner of the headboard of the bed	-
	18	upper corner of the headboard of the bed	-
	19	upper corner of the footboard of the bed	-
Skateboard	0	corner of the tail of the skateboard	-
	1	corner of the tail of the skateboard	-
	2	corner of the nose of the skateboard	-
	3	corner of the nose of the skateboard	-
	4	wheel of the skateboard	wheel
	5	wheel of the skateboard	wheel
	6	wheel of the skateboard	wheel
	7	wheel of the skateboard	wheel
	8	nose of the skateboard	end
	9	tail of the skateboard	end
Knife	0	tip of the knife	tip of the blade
	1	butt of the knife	bottom
	2	shoulder of the knife	-
	3	heel of the knife	-
	4	end of the crossguard of the knife	-
	5	end of the crossguard of the knife	-
Cap	0	top button of the cap	top
	1	front of the visor of the cap	front of the visor
	2	corner of the visor of the cap	-
	3	corner of the visor of the cap	-
	4	center of the back of the cap	middle of the strap
	5	center of the visor attachment of the cap	-
Laptop	0	corner of the laptop body	bottom left corner
	1	hinge of the laptop	bottom left corner of the screen
	2	hint of the laptop	bottom right corner of the screen
	3	corner of the laptop body	bottom right corner
	4	corner of the laptop screen	top left corner of the screen
	5	corner of the laptop screen	top right corner of the screen

Object	Index	Description	Short Description
Motorcycle	0	headlight of the motorcycle	headlight
	1	end of the handlebar of the motorcycle	right part of the steering wheel
	2	end of the handlebar of the motorcycle	left part of the steering wheel
	3	seat of the motorcycle	seat
	4	axle of the front wheel of the motorcycle	front wheel
	5	axle of the back wheel of the motorcycle	back wheel
	6	bottom part of the frame of the motorcycle	bottom of the engine
	7	axle of the front wheel of the motorcycle	front wheel
	8	axle of the back wheel of the motorcycle	back wheel
	9	front part of the front wheel of the motorcycle	front of the front wheel
	10	back part of the back wheel of the motorcycle	back of the back wheel
	11	top of the fuel tank of the motorcycle	top of the tank
	12	front end of the front fender of the motorcycle	top of the front wheel
	13	back end of the rear fender of the motorcycle	top of the back wheel
Airplane	0	nose of the airplane	nose of the airplane
	1	cockpit of the airplane	top of the cockpit
	2	tail of the airplane	back end
	3	the end of the left wing of the airplane	tip of all the wings
	4	wing root of the left wing of the airplane	different corners of the intersection between the body and the wings
	5	the end of the right wing of the airplane	tip of all the wings
	6	wing root of the right wing of the airplane	different corners of the intersection between the body and the wings
	7	wingtip of the left horizontal stabilizer of the airplane	tip of all the wings
	8	wingtip of the right horizontal stabilizer of the airplane	tip of all the wings
	9	top of vertical stabilizer of the airplane	-
	10	wingtip of the left diagonal stabilizer of the airplane	tip of all the wings
	11	wingtip of the right diagonal stabilizer of the airplane	tip of all the wings
	12	wing root of the left wing of the airplane	different corners of the intersection between the body and the wings
	13	wing root of the right wing of the airplane	different corners of the intersection between the body and the wings
	14	inner turbine of the left wing of the airplane	center of all turbines
	15	inner turbine of the right wing of the airplane	center of all turbines
	16	outer turbine of the left wing of the airplane	center of all turbines
	17	outer turbine of the right wing of the airplane	center of all turbines
	18	outer turbine of the right wing of the airplane	-

Object	Index	Description	Short Description
Car	0	front right wheel of the car	front right wheel
	1	rear right wheel of the car	back right wheel
	2	rear left wheel of the car	back left wheel
	3	front left wheel of the car	front left wheel
	4	corner of the front bumper of the car	right part of the front bumper
	5	corner of the rear bumper of the car	left part of the back bumper
	6	corner of the rear bumper of the car	left part of the back bumper
	7	corner of the front bumper of the car	right part of the front bumper
	8	top corner of the windshield of the car	top of right window
	9	top corner of the windshield of the car	top right corner of back windshield
	10	top corner of the rear window of the car	left corner of back windshield
	11	top corner of the rear window of the car	top left corner of front windshield
	12	bottom corner of the windshield of the car	bottom left corner of front windshield
	13	bottom corner of the windshield of the car	bottom right corner of front windshield
	14	bottom corner of the rear window of the car	bottom right corner of back windshield
	15	bottom corner of the rear window of the car	bottom left corner of back windshield
	16	right headlight of the car	front right headlight
	17	right taillight of the car	right back headlight
	18	left taillight of the car	left back headlight
	19	left headlight of the car	front left headlight
	20	right side mirror of the car	right rear view
	21	left side mirror of the car	left rear view
Guitar	0	head of the guitar	top of the headstock
	1	connection of the body and neck of the guitar	middle of the fretboard
	2	corner of the body of the guitar	top left corner of the body
	3	corner of the body of the guitar	top right corner of the body
	4	corner of the body of the guitar	bottom left corner of the body
	5	corner of the body of the guitar	bottom right corner of the body
	6	strap pin on the bottom of the body of the guitar	bottom of the body
	7	waist of the body of the guitar	middle left of the body
	8	waist of the body of the guitar	middle right of the body

Object	Index	Description	Short Description
Bathtub	0	top outer corner of the bathtub	-
	1	top outer corner of the bathtub	-
	2	top outer corner of the bathtub	-
	3	top outer corner of the bathtub	-
	4	bottom outer corner of the bathtub	-
	5	bottom outer corner of the bathtub	-
	6	bottom outer corner of the bathtub	-
	7	bottom outer corner of the bathtub	-
	8	top inner corner of the bathtub	-
	9	top inner corner of the bathtub	-
	10	top inner corner of the bathtub	-
	11	top inner corner of the bathtub	-
	12	bottom inner corner of the bathtub	-
	13	bottom inner corner of the bathtub	-
	14	bottom inner corner of the bathtub	-
	15	bottom inner corner of the bathtub	-
	16	center of the short side upper edge of the bathtub	-
	17	center of the short side upper edge of the bathtub	-
	18	center of the short side lower edge of the bathtub	-
	19	center of the short side lower edge of the bathtub	-
	20	center of the short side upper edge of the bathtub	-
	21	center of the short side upper edge of the bathtub	-
	22	center of the short side lower edge of the bathtub	-
	23	center of the short side lower edge of the bathtub	-
Helmet	0	top of the helmet	top
	1	front of the helmet peak	front
	2	back of the helmet peak	back
	3	left side of the helmet peak	left
	4	right side of the helmet peak	right
	5	right bottom of the helmet	bottom right
	6	left bottom of the helmet	bottom left
	7	bottom of the visor of the helmet	-
	8	bottom of the back of the helmet	bottom back

References

[1] Souhaib Attaiki and Maks Ovsjanikov. Ncp: Neural correspondence prior for effective unsupervised shape matching. *Advances in Neural Information Processing Systems*, 35:28842–28857, 2022. 5

[2] Xuyang Bai, Zixin Luo, Lei Zhou, Hongbo Fu, Long Quan, and Chiew-Lan Tai. D3feat: Joint learning of dense detection and description of 3d local features. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6359–6367, 2020. 5

[3] Xiaobai Chen, Abulhair Saparov, Bill Pang, and Thomas Funkhouser. Schelling points on 3d surface meshes. *ACM Transactions on Graphics (TOG)*, 31(4):1–12, 2012. 1, 2

[4] Eric Hedlin, Gopal Sharma, Shweta Mahajan, Xingzhe He, Hossam Isack, Abhishek Kar, Helge Rhodin, Andrea Tagliasacchi, and Kwang Moo Yi. Unsupervised keypoints from pretrained diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22820–22830, 2024. 5

[5] Jiaxin Li and Gim Hee Lee. Usip: Unsupervised stable interest point detection from 3d point clouds. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 361–370, 2019. 5

[6] Zekun Qi, Runpei Dong, Shaochen Zhang, Haoran Geng, Chunrui Han, Zheng Ge, Li Yi, and Kaisheng Ma. Shapellm: Universal 3d object understanding for embodied interaction. In *Computer Vision – ECCV 2024: 18th European Conference, Milan, Italy, September 29–October 4, 2024, Proceedings, Part XLIII*, page 214–238, Berlin, Heidelberg, 2024. Springer-Verlag. 1

[7] Blaine Rister, Mark A Horowitz, and Daniel L Rubin. Volumetric image registration from invariant keypoints. *IEEE Transactions on Image Processing*, 26(10):4900–4910, 2017. 5

[8] Aleksandar Shtedritski, Christian Rupprecht, and Andrea Vedaldi. What does clip know about a red circle? visual prompt engineering for vlms. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11987–11997, 2023. 5

[9] Ivan Sipiran and Benjamin Bustos. Harris 3d: a robust extension of the harris operator for interest point detection on 3d meshes. *The Visual Computer*, 27:963–976, 2011. 5

[10] Gemma Team, Thomas Mesnard, Cassidy Hardin, Robert Dadashi, Surya Bhupatiraju, Shreya Pathak, Laurent Sifre, Morgane Rivière, Mihir Sanjay Kale, Juliette Love, Pouya Tafti, Léonard Hussenot, Pier Giuseppe Sessa, Aakanksha Chowdhery, Adam Roberts, Aditya Barua, Alex Botev, Alex Castro-Ros, Ambrose Slone, Amélie Héliou, Andrea Tacchetti, Anna Bulanova, Antonia Paterson, Beth Tsai, Bobak Shahriari, Charline Le Lan, Christopher A. Choquette-Choo, Clément Crepy, Daniel Cer, Daphne Ippolito, David Reid, Elena Buchatskaya, Eric Ni, Eric Noland, Geng Yan, George Tucker, George-Christian Muraru, Grigory Rozhdestvenskiy, Henryk Michalewski, Ian Tenney, Ivan Grishchenko, Jacob Austin, James Keeling, Jane Labanowski, Jean-Baptiste Lespiau, Jeff Stanway, Jenny Brennan, Jeremy Chen, Johan Ferret, Justin Chiu, Justin Mao-Jones, Katherine Lee, Kathy Yu, Katie Millican, Lars Lowe Sjoesund, Lisa Lee, Lucas Dixon, Machel Reid, Maciej Mikuła, Mateo Wirth, Michael Sharman, Nikolai Chinaev, Nithum Thain, Olivier Bachem, Oscar Chang, Oscar Wahltinez, Paige Bailey, Paul Michel, Petko Yotov, Rahma Chaabouni, Ramona Comanescu, Reena Jana, Rohan Anil, Ross McIlroy, Ruibo Liu, Ryan Mullins, Samuel L Smith, Sebastian Borgeaud, Sertan Girgin, Sholto Douglas, Shree Pandya, Siamak Shakeri, Soham De, Ted Klimenko, Tom Hennigan, Vlad Feinberg, Wojciech Stokowiec, Yu hui Chen, Zafarali Ahmed, Zhitao Gong, Tris Warkentin, Ludovic Peran, Minh Giang, Clément Farabet, Oriol Vinyals, Jeff Dean, Koray Kavukcuoglu, Demis Hassabis, Zoubin Ghahramani, Douglas Eck, Joelle Barral, Fernando Pereira, Eli Collins, Armand Joulin, Noah Fiedel, Evan Senter, Alek Andreev, and Kathleen Kenealy. Gemma: Open models based on gemini research and technology, 2024. 5

[11] Thomas Wimmer, Peter Wonka, and Maks Ovsjanikov. Back to 3d: Few-shot 3d keypoint detection with back-projected 2d features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4154–4164, 2024. 3

[12] Monika Wysoczanska, Oriane Siméoni, Michaël Ramamonjisoa, Andrei Bursuc, Tomasz Trzcinski, and Patrick Pérez. Clip-dinoiser: Teaching clip a few dino tricks for open-vocabulary semantic segmentation. In *Proc. ECCV*, 2024. 5

[13] Yang You, Wenhai Liu, Yanjie Ze, Yong-Lu Li, Weiming Wang, and Cewu Lu. Ukpgan: A general self-supervised keypoint detector. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 2, 5

[14] Yu Zhong. Intrinsic shape signatures: A shape descriptor for 3d object recognition. In *2009 IEEE 12th international conference on computer vision workshops, ICCV Workshops*, pages 689–696. IEEE, 2009. 5