

000
001 **GIQ: BENCHMARKING 3D GEOMETRIC REASONING**
002 **OF VISION FOUNDATION MODELS WITH SIMULATED**
003 **AND REAL POLYHEDRA**
004
005 **SUPPLEMENTARY MATERIAL**
006
007
008

009 **Anonymous authors**
010 Paper under double-blind review
011
012
013

014 This supplementary material provides additional details and extended experimental analyses to complement and further support the results presented in the main paper. Specifically, it includes:
015

- 016 • Extended summary of polyhedral groups: their key geometric features, and additional representative samples.
017
- 019 • Extended qualitative and quantitative evaluations of monocular 3D reconstruction methods across various polyhedral categories.
020
- 021 • Detailed statistics and descriptions of dataset splits for the 3D symmetry detection task, along with synthetic-to-wild generalization gaps across featurizers and symmetry types.
022
- 023 • Additional analyses and results from the Mental Rotation Test (MRT), covering both trivial and challenging setups, along with example test pairs from the hard split.
024
- 025 • Expanded zero-shot polyhedron classification results across synthetic and real-world images, with additional qualitative examples highlighting common reasoning failures.
026
- 027 • **Extended probing analysis comparing linear versus non-linear (MLP) heads for MRT and Symmetry Detection**, including results for three additional multi-view pretrained models (VGGT, DUS3R, MAS3R).
028
- 029 • **Ablation studies on zero-shot classification**, evaluating 3D-native VLMs (LLaVA-3D, ShapELLM, PointLLM) and the impact of Multi-View inputs and Chain-of-Thought prompting strategies.
030
- 031 • **Additional qualitative samples from the “Wild” dataset partition**, illustrating the diversity of indoor and outdoor environmental conditions.
032
- 033
- 034
- 035
- 036

037 **1 EXPANDED CHARACTERIZATION OF THE POLYHEDRAL DOMAIN**
038

039 To provide a more comprehensive understanding of the geometric domain covered by our GIQ
040 dataset, we present in Table 5 an extended summary of the polyhedral groups used in our study.
041 This table details families of polyhedra, from the well-known Platonic and Archimedean solids
042 to more complex groups like the Johnson solids and non-convex stellations. For each group, we
043 outline the defining geometric properties, specify the number of solids within that category, and
044 provide additional representative visual examples. Beyond this taxonomy, polyhedra exhibit many
045 further attributes—e.g., chirality, face-type distributions, stellation depth, or Rupert property—that
046 can be leveraged to design additional benchmarks, enabling targeted evaluation of geometric rea-
047 soning across tasks such as duality reasoning, convexity discrimination, component decomposition,
048 or face-type counting.
049

050 **2 EXTENDED MONOCULAR 3D RECONSTRUCTION ANALYSES**
051

052 This section provides additional quantitative and qualitative evaluations of state-of-the-art monocular
053 3D reconstruction methods, specifically Shap-E (Jun & Nichol, 2023), Stable Fast 3D (Boss et al.,
2024), and OpenLRM (He & Wang, 2023), across diverse polyhedral categories. We qualitatively

evaluate reconstructions using standard geometric similarity metrics: F-score (Tatarchenko et al., 2019), Chamfer Distance (Borgefors, 1986), and Hausdorff Distance (Aspert et al., 2002), which jointly capture point-level accuracy and surface coverage. Results showed in Table 9 consistently reveals low F-scores (below 0.6) across all tested methods and categories, underscoring the limitations of current approaches in capturing the complex geometric intricacies inherent in polyhedral structures. Additionally, qualitative assessments for selected shapes from Archimedean, compound, and stellation groups (Table 8) highlight clear deficiencies in preserving critical geometric details, symmetry properties, and overall structural coherence, indicating significant scope for methodological advancements.

3 DATASET SPLITS AND COMPOSITION FOR 3D SYMMETRY DETECTION

We provide a detailed description of the dataset structure and splits used for the 3D symmetry detection experiments. [To ensure robust evaluation, we employed a 5-fold cross-validation strategy.](#) [In each fold](#), the test split exclusively contains views from 26 unique polyhedral shapes not present in the training set, enabling rigorous evaluation of model generalization to unseen geometries. Table 6 summarizes the distribution of positive and negative examples and their ratios for central point reflection, 5-fold rotation, and 4-fold rotation symmetries [for a representative fold \(Fold 1\)](#).

Unlike the main paper, which reports only Wild performance, Table 7 reports balanced accuracies for linear probes trained on synthetic images and evaluated on both synthetic and Wild inputs for central point reflection, 5-fold rotation, and 4-fold rotation. As expected given the training domain, accuracies are higher on synthetic inputs, while the synthetic-to-Wild generalization gap depends on both featurizer and symmetry: for example, CLIP shows a minimal gap on 5-fold ($0.80 \rightarrow 0.78$), DINO suffers large drops—especially on 4-fold ($0.87 \rightarrow 0.61$) and also on 5-fold ($0.88 \rightarrow 0.71$) while DINOv2 generalizes strongly on 4-fold ($0.96 \rightarrow 0.93$) and ties for best Wild 5-fold (0.85).

4 EMBEDDING STRATEGIES AND ADDITIONAL MENTAL ROTATION RESULTS

We provide additional analyses for the Mental Rotation Test (MRT), reporting results for both simplified (trivial) and challenging (hard) experimental setups, with representative examples of the hard split shown in Table 12. Under the simplified scenario (Table 10), synthetic image pairs with an 80%-20% train-test split are used. Most models achieve high accuracy (93%-98%) when employing the absolute difference embedding method. In contrast, concatenation and raw subtraction embedding methods, which include randomized embedding ordering during training, yield near-random performance (50% accuracy). For the challenging scenario (Table 11), additional analyses using raw subtraction and concatenation embeddings further confirm their consistently inferior performance compared to the absolute difference embeddings.

5 ZERO-SHOT POLYHEDRON CLASSIFICATION: SYNTHETIC VS. REAL-WORLD

Finally, we provide expanded analyses for the zero-shot polyhedron classification experiments presented in the main paper. Table 13 compares classification performance of four leading vision-language models—ChatGPT o3, ChatGPT o4-mini-high, Gemini 2.5 Pro, and Claude 3.7 Sonnet—using both synthetic and real-world (“wild”) images. Results indicate only marginal performance differences between synthetic and real-world inputs, confirming consistent capabilities across these domains. However, polyhedral categories such as Catalan solids, Johnson solids, compound structures, stellations and uniform non-convex polyhedra remain particularly challenging, underscoring persistent limitations in geometric reasoning within current frontier vision-language models.

To further illustrate these challenges, Figure 1 presents additional qualitative examples of the models’ reasoning processes. These cases reveal a recurring failure mode: models correctly detect local cues (e.g., pentagonal faces and color/pattern) but miscompose them into the wrong global structure—hallucinating absent elements (hexagons), overlooking nonconvexity (e.g., labeling a noncon-

108 convex solid as Archimedean, though all Archimedean solids are convex), misreporting face types, and
 109 at times rationalizing the error by blaming viewpoint.
 110

111 6 EXTENDED PROBING ANALYSIS: LINEAR VS. NON-LINEAR

112
 113 In this section, we provide the complete results comparing linear versus non-linear (MLP) probing
 114 performance for both the Mental Rotation Test (MRT) and 3D Symmetry Detection tasks, as
 115 summarized in Table 1. Additionally, we include results for the newly added multi-view pretrained
 116 models: VGGT Wang et al. (2025), DUS3R Wang et al. (2024), and MAS3R Leroy et al. (2024).
 117 Regarding the probing architectures, the linear probe consists of a single affine layer mapping the input
 118 embedding dimension D to the target class dimension C . The non-linear probe is implemented as a
 119 Multi-Layer Perceptron (MLP) with one hidden layer. Specifically, it maps the input D to a
 120 hidden dimension $H = \min(1024, D)$, applies a ReLU activation function, and projects the result
 121 to the output dimension C . To ensure adequate convergence for the increased parameter count, non-
 122 linear probes were trained for twice the number of epochs compared to the linear baselines. Our
 123 results indicate that while non-linear probes yield significant gains for the Mental Rotation Test (en-
 124 abling SigLIP to reach 69% accuracy), they offer negligible improvements for Symmetry Detection,
 125 suggesting that symmetry cues are largely linearly separable. Furthermore, despite their native 3D
 126 training, multi-view models do not consistently outperform strong 2D baselines (e.g., DINOv2) on
 127 these discriminative geometric tasks.
 128

129 7 ABLATION STUDIES: MULTI-VIEW, CoT, AND 3D-NATIVE MODELS

130
 131 Here we provide detailed quantitative results for our additional ablation studies, including the eval-
 132 uation of 3D-native VLMs (LLaVA-3D Zhu et al. (2024), ShapeLLM Qi et al. (2024), PointLLM
 133 Guo et al. (2023)) and the comparison of Single-View (SV) vs. Multi-View (MV) inputs with Chain-
 134 of-Thought (CoT) prompting.

135 For the prompting experiments, we defined the specific queries as follows:

136 Baseline Prompt: “*What is the name of this polyhedron?*”
 137

138 Chain-of-Thought (CoT) Prompt: “*Let’s identify this polyhedron by thinking step-by-step: Convex-
 139 ity: First, analyze its overall shape. Is this a convex polyhedron or is it non-convex? Symmetry and
 140 Rotation: Second, describe its symmetries. What rotational symmetries does it have? Does it have
 141 planes of reflectional symmetry? Faces and Vertices: Third, describe its components. What is the
 142 shape of each face, and how many faces meet at each vertex? Conclusion: Based on this analysis of
 143 its convexity, symmetry, and components, what is the precise name of this polyhedron?*”

144 As detailed in Table 2, 3D-native models did not exhibit a significant advantage over 2D-pretrained
 145 VLMs. For instance, even with access to ground truth point clouds, PointLLM achieved only 40%
 146 accuracy on Platonic solids and failed to classify most complex categories (e.g., 0% on Stellations
 147 and Compounds). Furthermore, Table 3 indicates that Chain-of-Thought (CoT) prompting yielded
 148 negligible gains and often degraded performance due to hallucinations. Multi-view inputs provided
 149 a minor boost for low-symmetry shapes (e.g., Johnson solids), resolving projection ambiguities, but
 150 failed to improve recognition for high-symmetry classes where the failure stems from reasoning
 151 rather than viewpoint selection.

152 8 ADDITIONAL REAL-WORLD QUALITATIVE SAMPLES

153
 154 To further illustrate the diversity and complexity of the real-world data distribution in GIQ, we pro-
 155 vide expanded qualitative samples in Table 4. These images highlight the significant domain shift
 156 introduced by the “Wild” split, encompassing diverse indoor environments with artificial lighting
 157 (Rows 1–4) as well as outdoor settings featuring natural illumination, shadows, and varied back-
 158 grounds (Rows 5–9).
 159

Featurizer	Syn-Wild MRT	4-fold rot.	5-fold rot.	CPF
Linear probe				
DINOv2	0.43	0.93	0.84	0.80
SigLIP	<u>0.54</u>	<u>0.81</u>	0.87	0.64
ConvNeXt	0.60	0.80	<u>0.86</u>	0.64
CLIP	0.39	0.74	0.78	0.70
DreamSim	0.49	0.74	0.78	0.65
DeiT III	0.41	0.69	0.74	<u>0.71</u>
DINO	<u>0.54</u>	0.71	0.81	0.56
SAM	0.42	0.66	0.75	0.58
MASt3R	0.48	0.64	0.70	0.58
DUSt3R	0.53	0.63	0.67	0.49
MAE	0.42	0.62	0.63	0.53
VG GT	0.47	0.51	0.56	0.49
Mean (linear)	0.478	0.69	0.75	0.62
Nonlinear probe				
DINOv2	<u>0.67</u>	0.91	0.81	0.75
SigLIP	0.69	<u>0.78</u>	0.84	<u>0.73</u>
ConvNeXt	0.59	0.69	0.80	0.71
CLIP	0.53	0.69	0.77	0.68
DreamSim	0.60	0.62	0.60	0.68
DeiT III	0.50	0.71	0.72	0.68
DINO	0.63	0.67	0.70	0.66
SAM	0.52	0.62	0.71	0.59
MASt3R	0.57	0.56	0.56	0.52
DUSt3R	0.58	0.55	0.55	0.54
MAE	0.61	0.66	0.61	0.60
VG GT	0.52	0.54	0.53	0.50
Mean (nonlinear)	0.56	0.67	0.70	0.64

Table 1: Extended Probing Analysis (Linear vs. Nonlinear) with multi-view pretrained models. We compare performance across standard 2D foundation models and newly added multi-view pretrained models (VG GT, DUSt3R, MASt3R). Featurizers are sorted by the combined rank of accuracies. **Bold** indicates the best result per column/section; underline indicates the second best.

216

217

218

Method	Plat.	Arch.	Cat.	John.	KP	Stel.	Comp.	NonConv.
Simple prompt								
LLaVA-3D (Syn)	20%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LLaVA-3D (Wild)	20%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
ShapeLLM	40%	7.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
PointLLM	40%	7.6%	7.6%	1.0%	25%	0.0%	0.0%	3.8%
CoT prompt								
LLaVA-3D (Syn)	0%	0.0%	0.0%	1.0%	0.0%	0.0%	0.0%	0.0%
LLaVA-3D (Wild)	20%	7.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
ShapeLLM	20%	7.6%	0.0%	1.8%	0.0%	0.0%	0.0%	0.0%
PointBind & PointLLM	0.0%	15.3%	0.0%	0.0%	0.0%	0.0%	0.0%	1.9%

Table 2: Zero-Shot Classification with 3D-aware VLMs. We evaluate LLaVA-3D (using synthetic/wild images) and point-cloud-based models (ShapeLLM, PointLLM, utilizing ground truth point clouds). Performance is reported for both Simple and Chain-of-Thought (CoT) prompting strategies.

Method	Plat.	Arch.	Cat.	John.	KP	Stel.	Comp.	NonConv.
Synthetic Images								
Baseline Prompt								
GPT-4o (SV)	80.0%	7.6%	0.0%	0.0%	50.0%	0.0%	10.0%	0.0%
GPT-4o (MV)	100.0%	7.6%	0.0%	4.0%	25.0%	0.0%	0.0%	1.9%
GPT-5-mini (SV)	60.0%	7.6%	0.0%	1.4%	0.0%	0.0%	0.0%	0.0%
GPT-5-mini (MV)	80.0%	15.3%	7.6%	2.1%	0.0%	2.9%	10.0%	0.0%
Chain-of-Thought (CoT) Prompt								
GPT-4o (SV)	100.0%	0.0%	0.0%	0.0%	50.0%	2.9%	0.0%	0.0%
GPT-4o (MV)	100.0%	0.0%	0.0%	2.9%	25.0%	2.9%	0.0%	1.9%
GPT-5-mini (SV)	60.0%	7.6%	0.0%	1.4%	25.0%	0.0%	10.0%	0.0%
GPT-5-mini (MV)	100.0%	23.0%	15.3%	5.0%	25.0%	2.9%	10.0%	0.0%
Wild Images								
Baseline Prompt								
GPT-4o (SV)	20.0%	7.6%	0.0%	2.7%	50.0%	0.0%	16.7%	0.0%
GPT-4o (MV)	40.0%	0.0%	0.0%	3.4%	25.0%	0.0%	16.7%	1.9%
GPT-5-mini (SV)	20.0%	23.0%	7.6%	0.0%	0.0%	0.0%	0.0%	0.0%
GPT-5-mini (MV)	20.0%	23.0%	7.6%	3.4%	25.0%	0.0%	16.7%	0.0%
Chain-of-Thought (CoT) Prompt								
GPT-4o (SV)	20.0%	7.6%	0.0%	0.0%	50.0%	0.0%	0.0%	0.0%
GPT-4o (MV)	40.0%	15.3%	0.0%	2.7%	25.0%	0.0%	16.7%	1.9%
GPT-5-mini (SV)	0.0%	0.0%	7.6%	2.0%	25.0%	0.0%	0.0%	0.0%
GPT-5-mini (MV)	0.0%	30.7%	15.3%	6.9%	50.0%	0.0%	16.7%	0.0%

Table 3: Ablation Study on Input Modality and Prompting. We compare zero-shot classification accuracy for ChatGPT-4o and ChatGPT-5-mini using Single-View (SV) vs. Multi-View (MV) inputs, and Baseline vs. Chain-of-Thought (CoT) prompts across synthetic and wild domains.

268

269

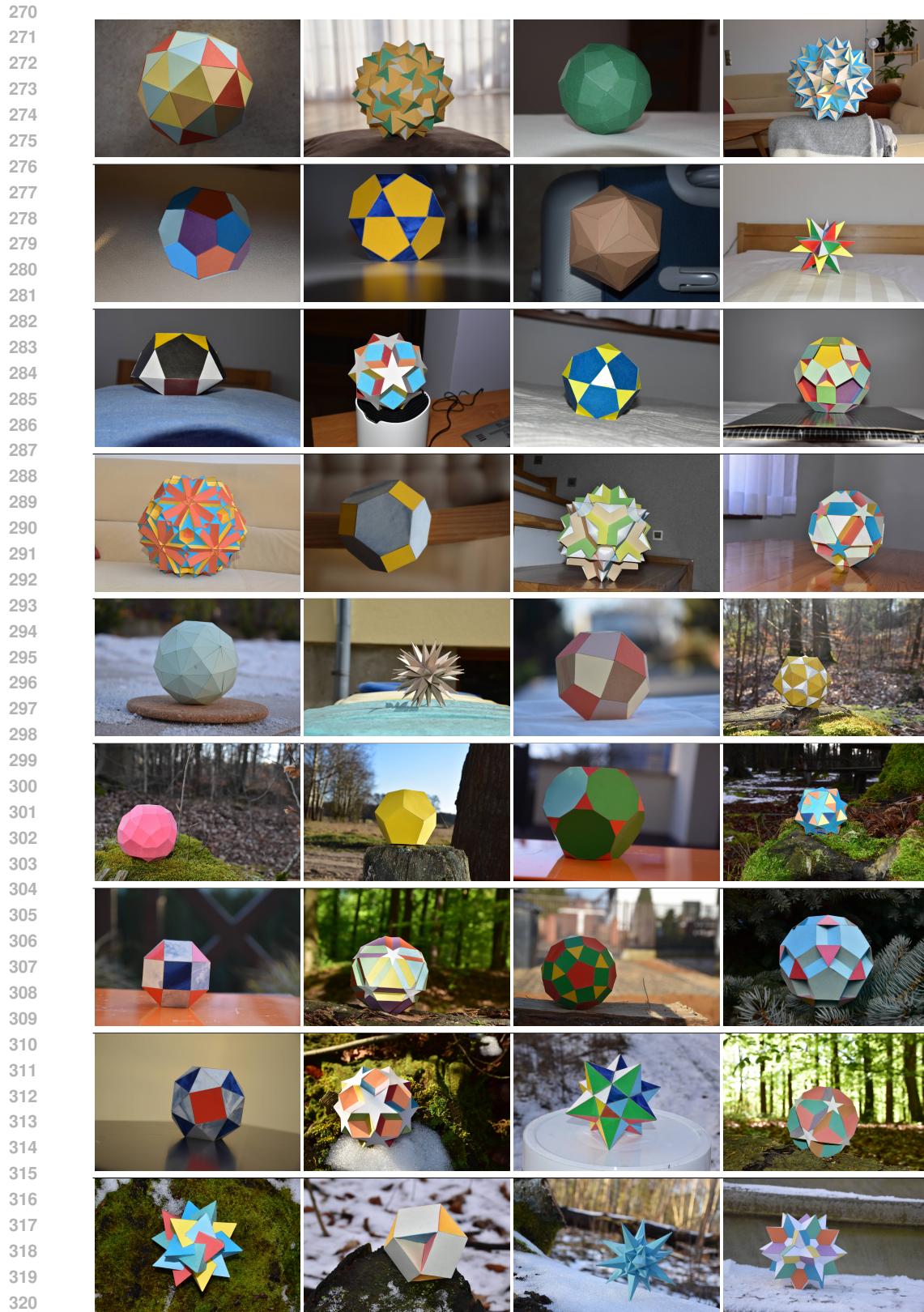


Table 4: **Additional Real-World Examples.** Selected samples from the Wild dataset demonstrating diverse environmental conditions, including indoor, outdoor scenes, and varying lighting.

Group	#	Key Features	Examples from GIQ
Platonic	5	Congruent regular polygonal faces; vertex-transitive, face-transitive	
Archimedean	13	Vertex-transitive; faces are regular polygons of different types	
Catalan	13	Duals of Archimedean solids; face-transitive, but not vertex-transitive	
Johnson	92	Convex polyhedron; regular polygonal faces	
Stellations	48	Formed by extending faces or edges; Nonconvex	
Kepler-Poinsot	4	Regular star polyhedra; specific stellations of Platonic solids	
Compounds	10	Symmetric combination of multiple polyhedra	
Uniform non-convex	53	Nonconvex; regular polygonal faces, vertex-transitive	

Table 5: Extended summary of polyhedral groups in the GIQ dataset, with counts of 3D shapes, key geometric features, and additional representative samples.

Symmetry Element	Split	Positives	Negatives	Pos./Neg. Ratio
Central point reflection	Train	1168	480	2.43
5-fold rotation	Train	432	1216	0.36
4-fold rotation	Train	1424	224	6.36
Central point reflection	Test	752	180	4.18
5-fold rotation	Test	148	784	0.19
4-fold rotation	Test	716	216	3.31

Table 6: Detailed composition of training and test dataset splits used for 3D symmetry detection experiments (representative statistics from Fold 1 of the 5-fold cross-validation). For each considered symmetry element (central point reflection, 5-fold rotation, and 4-fold rotation), the number of positive and negative samples, as well as the corresponding positive-to-negative ratio, is provided.

378
379
380
381
382
383
384
385
386
387
388
389
390
391
392

Featurizer	Central point reflection		5-fold rotation		4-fold rotation	
	Syn	Wild	Syn	Wild	Syn	Wild
DINOv2	0.85	0.73	0.97	0.85	0.96	0.93
CLIP	0.82	0.74	0.80	0.78	0.74	0.69
ConvNext	0.76	0.62	0.93	0.85	0.90	<u>0.76</u>
SigLip	0.69	0.59	0.86	<u>0.79</u>	0.78	0.74
MAE	0.70	0.66	0.82	0.71	0.74	0.71
DeiT III	0.75	0.65	0.82	0.72	0.70	0.65
DreamSim	0.77	0.71	0.78	0.71	0.68	0.63
SAM	0.74	0.73	0.79	0.68	0.68	0.59
DINO	0.66	0.62	0.88	0.71	0.87	0.61

393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431

Table 7: Extension of main-paper figure: balanced accuracies ($0.5 \cdot \frac{TP}{P} + 0.5 \cdot \frac{TN}{N}$) for linear probes trained on synthetic images and evaluated on both synthetic (Syn) and real-world (Wild) images for central point reflection, 5-fold rotation, and 4-fold rotation. Including Syn results quantifies the Syn→Wild generalization gap, which varies by featurizer and symmetry. Featurizers are sorted by the combined rank of Wild-set accuracies across the three symmetry types. **Bold** indicates best accuracy; underline indicates second best.

Input	Frontier Model Reasoning
	ChatGPT o4-mini-high: It has many faces that seem like pentagons , but I also see some hexagons . The polyhedron in the image seems like it could be a truncated icosahedron , due to its faces resembling hexagons and pentagons, but something about it doesn't match exactly . It's hard to tell without a clearer view of its exact pattern .
	Claude 3.7 Sonnet: I can see a colorful polyhedron in the image. This appears to be a cuboctahedron , which is an Archimedean solid . It has 8 triangular faces and 6 square faces , with 12 identical vertices where 2 triangles and 2 squares meet. The model shown has a checkered pattern with light blue/teal .

Figure 1: Examples of failure cases illustrating errors in reasoning by frontier models that led to misclassification of polyhedra. These cases highlight systematic mistakes in geometric recognition, such as misidentification of face geometry, convexity, and compound structures. Text highlighted in green indicates correct statements, while text in red indicates incorrect reasoning.

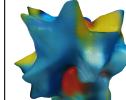
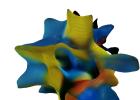
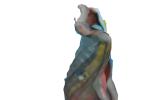
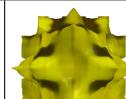
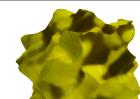
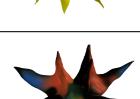
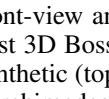
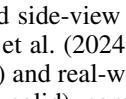
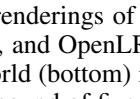
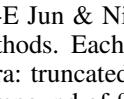
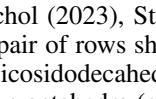
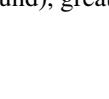
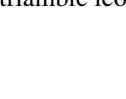
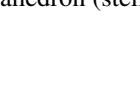
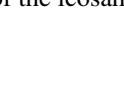
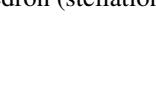
Input	Shap-E		Stable Fast 3D		OpenLRM	
	Front	Side	Front	Side	Front	Side
						
						
						
						
						
						
						
						
						
						
						
						
						

Table 8: Additional qualitative results for monocular 3D reconstruction, supplementing evaluations presented in the main paper. Columns depict, from left to right: the input 2D image, followed by front-view and side-view renderings of reconstructions from Shap-E Jun & Nichol (2023), Stable Fast 3D Boss et al. (2024), and OpenLRM He & Wang (2023) methods. Each pair of rows shows synthetic (top) and real-world (bottom) images of selected polyhedra: truncated icosidodecahedron (Archimedean solid), compound of five tetrahedra (compound), compound of five octahedra (compound), great triambic icosahedron (stellation), and final stellation of the icosahedron (stellation).

486
487
488
489
490
491
492
493
494

Category	Metric	Synthetic			Wild		
		Shap-E	SF3D	OpenLRM	Shap-E	SF3D	OpenLRM
Platonic	F-score \uparrow	0.367	0.521	0.626	0.380	0.378	0.224
	Hausdorff Distance \downarrow	0.083	0.047	0.043	0.158	0.123	0.173
	Chamfer Distance \downarrow	0.001	0.001	0.001	0.006	0.004	0.008
Archimedean	F-score \uparrow	0.355	0.477	0.586	0.309	0.348	0.175
	Hausdorff Distance \downarrow	0.083	0.052	0.062	0.134	0.137	0.215
	Chamfer Distance \downarrow	0.001	0.002	0.001	0.008	0.007	0.009
Catalan	F-score \uparrow	0.361	0.478	0.597	0.257	0.365	0.172
	Hausdorff Distance \downarrow	0.086	0.059	0.050	0.156	0.120	0.184
	Chamfer Distance \downarrow	0.001	0.002	0.002	0.009	0.005	0.007
Stellations	F-score \uparrow	0.259	0.231	0.239	0.162	0.096	0.191
	Hausdorff Distance \downarrow	0.119	0.179	0.158	0.313	0.292	0.297
	Chamfer Distance \downarrow	0.002	0.011	0.007	0.051	0.020	0.009
Kepler-Poinsot	F-score \uparrow	0.255	0.292	0.258	0.257	0.124	0.245
	Hausdorff Distance \downarrow	0.115	0.116	0.120	0.147	0.254	0.218
	Chamfer Distance \downarrow	0.002	0.001	0.002	0.002	0.017	0.006
Compounds	F-score \uparrow	0.272	0.253	0.252	0.220	0.113	0.184
	Hausdorff Distance \downarrow	0.110	0.136	0.135	0.172	0.294	0.271
	Chamfer Distance \downarrow	0.001	0.002	0.002	0.003	0.023	0.013
Uniform Nonconvex	F-score \uparrow	0.263	0.194	0.250	0.232	0.120	0.182
	Hausdorff Distance \downarrow	0.122	0.144	0.119	0.145	0.257	0.186
	Chamfer Distance \downarrow	0.002	0.005	0.002	0.004	0.021	0.008

527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
759
760
761
762
763
764
765
766
767
768
769
769
770
771
772
773
774
775
776
777
778
779
779
780
781
782
783
784
785
786
787
788
789
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
809
810
811
812
813
814
815
816
817
818
819
819
820
821
822
823
824
825
826
827
828
829
829
830
831
832
833
834
835
836
837
838
839
839
840
841
842
843
844
845
846
847
848
849
849
850
851
852
853
854
855
856
857
858
859
859
860
861
862
863
864
865
866
867
868
869
869
870
871
872
873
874
875
876
877
878
879
879
880
881
882
883
884
885
886
887
888
889
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
909
910
911
912
913
914
915
916
917
918
919
919
920
921
922
923
924
925
926
927
928
929
929
930
931
932
933
934
935
936
937
938
939
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
959
960
961
962
963
964
965
966
967
968
969
969
970
971
972
973
974
975
976
977
978
979
979
980
981
982
983
984
985
986
987
988
989
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1969<br

540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556

557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572

Featurizer	Concat ($e_1 \parallel e_2$)	Subtraction ($e_1 - e_2$)	Absolute ($ e_1 - e_2 $)
CLIP	0.5065	0.5786	0.9400
ConvNext	0.5188	0.5468	0.9306
DeiT III	0.4804	0.5537	0.9473
DINO	0.4616	0.5380	0.9674
DINOv2	0.4991	0.5693	0.9706
DreamSim	0.5221	0.5983	0.9843
MAE	0.5240	0.5247	0.9768
SAM	0.5035	0.5516	0.9519
SigLip	0.5399	0.5615	0.9594

573 Table 10: Balanced accuracy of linear probing approaches on pairwise embeddings for the Mental
 574 Rotation Test (MRT). Given two embeddings e_1, e_2 from each featurizer, we form input features by
 575 concatenation ($e_1 \parallel e_2$), subtraction ($e_1 - e_2$), or absolute difference ($|e_1 - e_2|$), followed by a linear
 576 classifier. Results are reported on a simplified “trivial” setting, with synthetic-only image pairs and
 577 shapes randomly split into 80% train and 20% test.

578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

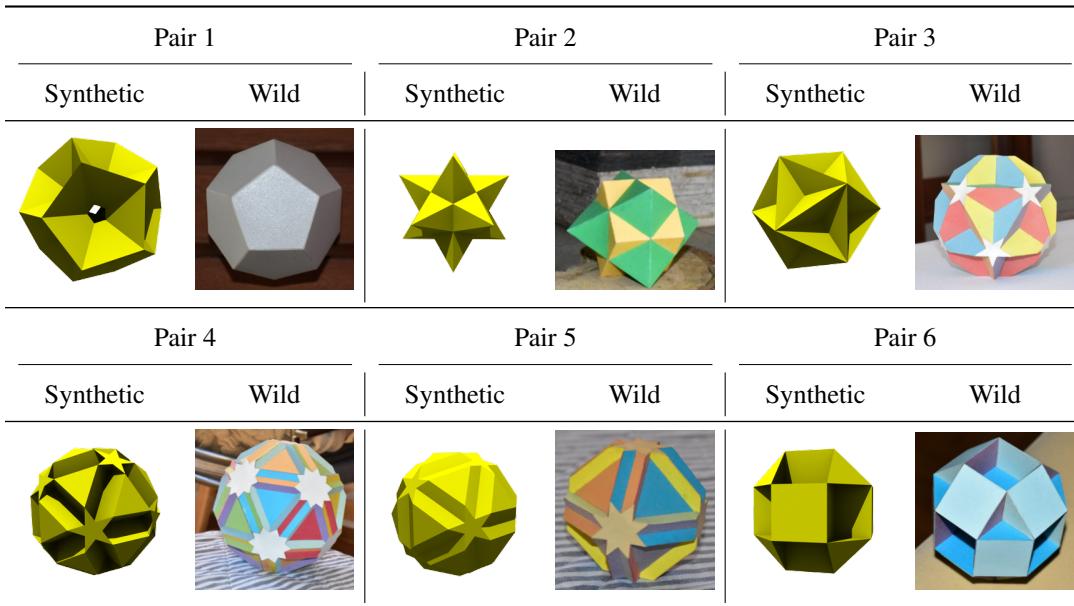
646

647

Featurizer	Train (syn,syn)→Test (syn,syn)	Train (syn,syn)→Test (syn,wild)	Train (syn,wild)→Test (syn,wild)
Absolute Difference $e_1 - e_2$			
CLIP	0.64	0.44	0.55
ConvNext	0.70	0.47	0.65
DeiT III	0.70	0.48	0.56
DINO	0.74	0.44	0.58
DINOv2	0.79	0.44	0.60
DreamSim	0.77	0.44	0.62
MAE	0.76	0.44	0.58
SAM	0.77	0.44	0.57
SigLip	0.76	0.44	0.64
Subtraction $(e_1 - e_2)$			
CLIP	0.36	0.39	0.56
ConvNext	0.51	0.49	0.54
DeiT III	0.45	0.46	0.53
DINO	0.48	0.49	0.49
DINOv2	0.53	0.51	0.45
DreamSim	0.38	0.34	0.55
MAE	0.37	0.34	0.53
SAM	0.47	0.48	0.50
SigLip	0.50	0.44	0.54
Concatenation $(e_1 \ e_2)$			
CLIP	0.35	0.56	0.58
ConvNext	0.46	0.50	0.52
DeiT III	0.41	0.54	0.54
DINO	0.52	0.52	0.54
DINOv2	0.57	0.48	0.60
DreamSim	0.38	0.55	0.55
MAE	0.41	0.52	0.53
SAM	0.37	0.56	0.59
SigLip	0.36	0.44	0.54

Table 11: Accuracy on the Mental Rotation Test (MRT) evaluated on the *hard* test set, where only pairs of visually similar shapes are considered. “Train (X)→Test (Y)” denotes the training and testing domains. Results are presented for absolute difference ($|e_1 - e_2|$), raw subtraction ($e_1 - e_2$), and feature concatenation ($e_1 \| e_2$). Absolute difference performed best overall.

648
649
650
651
652
653
654
655
656
657
658
659
660



661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701

Table 12: Samples of visually and geometrically similar synthetic-wild shape pairs used in the *hard* test set. Pairs were manually selected based on structural and visual similarities, such as shared symmetries, vertex configurations, and derivation from common polyhedra (e.g., pair 6: Small cubicuboctahedron and Small rhombihexahedron, both derived from the rhombicuboctahedron).

Category	Gemini 2.5 Pro		ChatGPT o4-mini-high		ChatGPT o3		Claude 3.7 Sonnet	
	Syn	Wild	Syn	Wild	Syn	Wild	Syn	Wild
Platonic	0.60	0.80	0.60	0.60	1.00	1.00	0.60	0.60
Archimedean	0.53	0.61	0.23	0.31	0.61	0.54	0.21	0.23
Catalan	0.15	0.15	0.15	0.15	0.08	0.08	0.09	0.08
Johnson	0.20	0.18	0.11	0.12	0.21	0.18	0.11	0.11
Kepler-Poinsot	1.00	1.00	0.25	0.25	0.25	0.25	0.25	0.25
Stellations	0.41	0.43	0.05	0.07	0.23	0.21	0.27	0.26
Compounds	0.16	0.17	0.00	0.00	0.16	0.33	0.33	0.33
Uniform non-convex	0.08	0.09	0.00	0.00	0.05	0.06	0.00	0.00

Table 13: Accuracy (%) of various frontier models on 0-shot classification across polyhedron categories, reported on synthetic (Syn) and in-the-wild (Wild) images. The Syn→Wild gap is generally small across categories and models, indicating comparable performance across domains.

702 REFERENCES
703704 Nicolas Aspert, Diego Santa-Cruz, and Touradj Ebrahimi. Mesh: Measuring errors between surfaces
705 using the hausdorff distance. In *Proceedings. IEEE international conference on multimedia and*
706 *expo*, volume 1, pp. 705–708. IEEE, 2002.707 Gunilla Borgefors. Distance transformations in digital images. *Computer vision, graphics, and*
708 *image processing*, 34(3):344–371, 1986.709 Mark Boss, Zixuan Huang, Aaryaman Vasishta, and Varun Jampani. Sf3d: Stable fast 3d
710 mesh reconstruction with uv-unwrapping and illumination disentanglement. *arXiv preprint*
711 *arXiv:2408.00653*, 2024.712 Ziyu Guo, Renrui Zhang, Xiangyang Zhu, Yiwen Tang, Xianzheng Ma, Jiaming Han, Kexin Chen,
713 Peng Gao, Xianzhi Li, Hongsheng Li, et al. Point-bind & point-llm: Aligning point cloud
714 with multi-modality for 3d understanding, generation, and instruction following. *arXiv preprint*
715 *arXiv:2309.00615*, 2023.716 Zexin He and Tengfei Wang. Openlrm: Open-source large reconstruction models. <https://github.com/3DTopia/OpenLRM>, 2023.717 Heewoo Jun and Alex Nichol. Shap-e: Generating conditional 3d implicit functions. *arXiv preprint*
718 *arXiv:2305.02463*, 2023.719 Vincent Leroy, Yohann Cabon, and Jérôme Revaud. Grounding image matching in 3d with mast3r.
720 In *European Conference on Computer Vision*, pp. 71–91. Springer, 2024.721 Zekun Qi, Runpei Dong, Shaochen Zhang, Haoran Geng, Chunrui Han, Zheng Ge, Li Yi, and
722 Kaisheng Ma. Shapellm: Universal 3d object understanding for embodied interaction. In *Euro-
723 pean Conference on Computer Vision*, pp. 214–238. Springer, 2024.724 Maxim Tatarchenko, Stephan R Richter, René Ranftl, Zhuwen Li, Vladlen Koltun, and Thomas
725 Brox. What do single-view 3d reconstruction networks learn? In *Proceedings of the IEEE/CVF
726 conference on computer vision and pattern recognition*, pp. 3405–3414, 2019.727 Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David
728 Novotny. Vggt: Visual geometry grounded transformer. In *Proceedings of the Computer Vision
729 and Pattern Recognition Conference*, pp. 5294–5306, 2025.730 Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jerome Revaud. Dust3r: Ge-
731 ometric 3d vision made easy. In *Proceedings of the IEEE/CVF Conference on Computer Vision
732 and Pattern Recognition*, pp. 20697–20709, 2024.733 Chenming Zhu, Tai Wang, Wenwei Zhang, Jiangmiao Pang, and Xihui Liu. Llava-3d: A simple
734 yet effective pathway to empowering lmms with 3d-awareness. *arXiv preprint arXiv:2409.18125*,
735 2024.

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755