A Dataset Details

A.1 Part list

In Super-CLEVR-3D, the parts of each objects are listed in Tab. 2

A.2 Question templates

Part Questions we collect 9 part-based templates when generating the part based questions, as shown in Tab. 4. In the table, `<attribute>` means one attribute from shape, material, color or size to be queried, `<object>` (or `<object 1>`, `<object 2>`) means one object to be filtered with a combination of shape, material, color and size. Different from the pose and occlusion question, we don't query the size of object.

3D Pose questions We design 17 3D pose-based templates in question generation (as shown in table 5). The 17 templates are consisted of: 1 template of query of pose; 4 question of query of shape, material, color, size, where pose is in the filtering conditions; 12 templates about query of shape, material, color, size, where the relationship of pose is the filtering condition.

Occlusion Questions There are 35 templates in the occlusion question generation as shown in table 6, which is consisted of occlusion of objects and occlusion of parts.

The occlusion of objects is consisted of occlusion status and occlusion relationship. For the occlusion status of object, there are 4 templates to query the shape, color, material and size respectively. And there are 2 occlusion relationship of objects (occluded and occluding), each of them have 4 templates.

Similarly, we then create template about occlusion status and occlusion relationship for the parts. The only differences between object and part is that the parts only have 3 attribute to be queried: shape (name), material and color.

A.3 Statistics

As a result, a total of 314,988 part questions, 314,986 pose questions, and 228,397 occlusion questions and 314,988 occlusion questions with parts.

B Implementation details for the baselines

The FiLM and mDETR are trained with default setting as in the official implementation. FiLM is trained for 100k iterations with batch size 256. mDETR is trained for 30 epochs with batch size 64 using 2 GPUs for both the grounding stage and the answer classification stage.

For P-NSVQA, we first train a MaskRCNN for 30k iterations with batch size 16 to detect the objects and parts, then train the attribute extraction model (using Res50 backbone) for 100 epochs with batch size 64. Different fully connection (FC) layers are used for different type of question: the part question and occlusion question has 4 FC layers for the shape, material, color, size classification (as the parts also have size annotations in the dataset when generating scene files, but they are meaningless in the question answering). The pose question includes pose prediction of object, so we add a new FC layer with 1 output dimension to prediction the rotations, followed by a MSE loss during training. For different types of questions (part, pose and occlusion), the MaskRCNN and attribute extraction model are trained seperately.

In the PNSVQA+Projection baseline, we first train a MaskRCNN to detect all of the objects and predict their 3D pose (azimuth, elevation and theta) without category labels in the scene. This MaskRCNN is trained with batch size 8 and iteration 15000. We use an SGD optimizer with learning rate 0.02, momentum 0.9 and weigth decay 0.0001. Then, we use the same setting as our PO3D-VQA to train a CNN to classifiy the attributes of objects and parts.

C Detailed results of Analysis

As extension for section 5.4 in main paper, here we include the numerical value of accuracy and drop for the pose, part, occlusion + part question with reference to occlusion ratio or part size. The result is shown in Tab. 7, Tab. 9 and Tab. 8.

Table 2: List of objects and parts.

shape	part list
airliner	left door, front wheel, fin, right engine, propeller, back left wheel, left engine, back right wheel, left tailplane, right door, right tailplane, right wing, left wing
biplane	front wheel, fin, propeller, left tailplane, right tailplane, right wing, left wing
jet	left door, front wheel, fin, right engine, propeller, back left wheel, left engine, back right wheel, left tailplane, right tailplane, right wing, left wing
fighter	fin, right engine, left engine, left tailplane, right tailplane, right wing, left wing
utility bike	left handle, brake system, front wheel, left pedal, right handle, back wheel, saddle, carrier, fork, right crank arm, front fender, drive chain, back fender, left crank arm, side stand, right pedal
tandem bike	rearlight, front wheel, back wheel, fork, front fender, back fender
road bike	left handle, brake system, front wheel, left pedal, right handle, back wheel, saddle, fork, right crank arm, drive chain, left crank arm, right pedal
mountain bike	left handle, brake system, front wheel, left pedal, right handle, back wheel, saddle, fork, right crank arm, drive chain, left crank arm, right pedal
articulated bus	left tail light, front license plate, front right door, back bumper, right head light, front left wheel, left mirror, right tail light, back right door, back left wheel, back right wheel, back license plate, front right wheel, left head light, right mirror, trunk, mid right door, roof
double bus	left tail light, front license plate, front right door, front bumper, back bumper, right head light, front left wheel, left mirror, right tail light, back left wheel, back right wheel, back license plate, mid left door, front left door, front right wheel, left head light, right mirror, trunk, mid right door, roof
regular bus	left tail light, front license plate, front right door, front bumper, back bumper, right head light, front left wheel, left mirror, right tail light, back right door, back left wheel, back right wheel, back license plate, front right wheel, left head light, right mirror, trunk, mid right door, roof
school bus	left tail light, front license plate, front right door, front bumper, back bumper, right head light, front left wheel, left mirror, right tail light, back left wheel, back right wheel, back license plate, mid left door, front right wheel, left head light, right mirror, roof
truck	front left door, left tail light, left head light, back right wheel, right head light, front bumper, right mirror, front license plate, front right wheel, back bumper, left mirror, back left wheel, right tail light, hood, trunk, front left wheel, roof, front right door
suv	front left door, left tail light, left head light, back left door, back right wheel, right head light, front bumper, right mirror, front right wheel, back bumper, left mirror, back left wheel, right tail light, hood, trunk, front left wheel, back right door, roof, front right door
minivan	front left door, left tail light, left head light, back left door, back right wheel, right head light, front bumper, right mirror, front license plate, front right wheel, back bumper, left mirror, back left wheel, right tail light, hood, trunk, front left wheel, back right door, roof, front right door, back license plate
sedan	front left door, left tail light, left head light, back left door, back right wheel, right head light, front bumper, right mirror, front license plate, front right wheel, back bumper, left mirror, back left wheel, right tail light, hood, trunk, front left wheel, back right door, roof, front right door, back license plate
wagon	front left door, left tail light, left head light, back left door, back right wheel, right head light, front bumper, right mirror, front license plate, front right wheel, back bumper, left mirror, back left wheel, right tail light, hood, trunk, front left wheel, back right door, roof, front right door, back license plate
chopper	left handle, center headlight, front wheel, right handle, back wheel, center taillight, left mirror, gas tank, front fender, fork, drive chain, left footrest, right mirror, windscreen, engine, back fender, right exhaust, seat, panel, right footrest
scooter	left handle, center headlight, front wheel, right handle, back cover, back wheel, center taillight, left mirror, front cover, fork, drive chain, right mirror, engine, left exhaust, back fender, seat, panel
cruiser	left handle, center headlight, right headlight, right taillight, front wheel, right handle, back cover, back wheel, left taillight, left mirror, left headlight, gas tank, front cover, front fender, fork, drive chain, left footrest, license plate, right mirror, windscreen, left exhaust, back fender, right exhaust, seat, panel, right footrest
dirtbike	left handle, front wheel, right handle, back cover, back wheel, gas tank, front cover, front fender, fork, drive chain, left footrest, engine, right exhaust, seat, panel, right footrest

Table 4: Templates of parts questions

Templates	Count
What is the <attribute> of the <part> of the <object>?	3
What is the <attribute> of the <object> that has a <part>?	3
What is the <attribute> of the <part 1> that belongs to the same object as the <part 2>?	3

Table 5: Templates of pose questions

Templates	Count
Which direction the <object> is facing?	1
What is the <attribute> of the <object> which face to the <O>?	4
What is the <attribute> of the <object 1> that faces the **same direction** as a <object 2>	4
What is the <attribute> of the <object 1> that faces the **opposite direction** as a <object 2>	4
What is the <attribute> of the <object 1> that faces the **vertical direction** as a <object 2>	4

Table 6: Templates of occlusion questions

Templates	Count
What is the <attribute> of the <object> that is occluded?	4
What is the <attribute> of the <object 1> that is occluded by the <object 2> ?	4
What is the <attribute> of the <object 1> that occludes the <object 2>?	4
Is the <part> of the <object> occluded?	1
Which part of the <object> is occluded?	1
What is the <attribute> of the <object> whose <part> is occluded?	4
What is the <attribute> of the <part> which belongs to an occluded <object>?	3
What is the <attribute> of the <part 1> which belongs to the <object> whose <part 2> is occluded?	3
Is the <part> of the <object 1> occluded by the <object 2>	1
What is the <attribute> of the <object 1> whose <part> is occluded by the <object 2> ?	4
What is the <attribute> of the <part> which belongs to <object 1> which is occluded by the <object 2>	3
What is the <attribute> of the <part 1> which belongs to the same object whose <part 2> is occluded by the <object 2>?	3

Table 7: Accuracy value and relative drop for pose questions wrt. occlusion ratio

	Occlusion Ratio	**0**	**5**	**10**	**15**	**20**	**25**	**30**
PNSVQA	Accuracy	87.43	74.09	74.09	63.16	62.01	60.33	58.52
	Drop	0.00%	15.26%	15.26%	27.76%	29.08%	31.00%	33.07%
PNSVQA + Projection	Accuracy	86.30	74.61	67.20	66.78	60.26	56.52	55.56
	Drop	0.00%	13.54%	22.13%	22.62%	30.17%	34.51%	35.63%
Ours	Accuracy	**86.43**	**86.05**	**84.32**	**75.00**	**79.44**	**73.22**	**67.98**
	Drop	0.00%	**0.44%**	**2.44%**	**13.22%**	**8.09%**	**15.28%**	**21.35%**

Table 8: Accuracy value and relative drop for occlusion + part wrt. part size

	Part Size	**max**	**300**	**150**	**100**	**50**	**20**
PNSVQA	Accuracy	58.18	54.98	54.05	52.09	45.20	21.28
	Drop	0.00%	5.49%	7.10%	10.47%	22.31%	63.43%
PNSVQA + Projection	Accuracy	61.85	50.64	56.77	53.97	55.29	45.83
	Drop	0.00%	18.11%	8.20%	12.74%	**10.60%**	**25.89%**
Ours	Accuracy	**81.68**	**75.32**	**77.20**	**71.54**	**67.00**	**53.19**
	Drop	0.00%	**7.78%**	**5.49%**	**12.41%**	17.97%	34.88%

Table 9: Accuracy value and relative drop for part wrt. part size

	Part Size	**max**	**300**	**150**	**100**	**50**	**20**
PNSVQA	Accuracy	57.31	51.00	37.50	44.18	40.85	29.73
	Drop	0.00%	11.02%	34.57%	22.92%	28.73%	48.12%
PNSVQA + Projection	Accuracy	58.89	57.54	42.64	43.20	46.73	38.67
	Drop	0.00%	2.30%	27.60%	26.65%	**20.65%**	34.34%
Ours	Accuracy	**64.04**	**64.80**	**60.16**	**57.03**	**49.05**	**55.41**
	Drop	0.00%	**-1.19%**	**6.06%**	**10.94%**	23.41%	**13.48%**