Table 1: Quantitative comparison results for mesh visual and geometry quality over **FULL** GSO dataset. We report the metrics of PSNR, SSIM, LPIPS and Clip-Similarity, ChamferDistance (CD), Volume IoU and F-score. The original paper use a Blender CYCLES renderer + 2048 resolution, but rendering 24 views of a sample takes over 10 minutes. This experiment is Blender EEVEE + 1024 resolution to ensure that it can finish before the rebuttal ended. So there is a numerical difference in the visual metrics (the metrics scores are higher for all methods).

Method	**PSNR↑**	**SSIM↑**	**LPIPS↓**	**Clip-Sim↑**	**CD↓**	**Vol. IoU↑**	**F-Score↑**
One-2-3-45	16.1058	0.8874	0.1812	0.7782	0.0313	0.4142	0.5518
OpenLRM	18.0433	0.8957	0.1560	0.8416	0.0336	0.3947	0.5354
Wonder3D	18.0932	0.8995	0.1536	0.8535	0.0261	0.4663	0.6016
InstantMesh	<u>18.8262</u>	<u>0.9111</u>	<u>0.1283</u>	**0.8795**	0.0161	0.5083	0.6491
CRM	18.4407	0.9088	0.1366	0.8639	**0.0141**	<u>0.5218</u>	<u>0.6574</u>
Unique3D	**20.0611**	**0.9222**	**0.1070**	<u>0.8787</u>	<u>0.0143</u>	**0.5416**	**0.6696**
Unique3D w/o ET	20.0383	0.9199	0.1129	0.8675	0.0158	0.5320	0.6594
Wonder3D+ISOMER	18.6131	0.9026	0.1470	0.8621	0.0244	0.4743	0.6088

Table 2: Quantitative comparison results for ablation on 100 random samples with random rotation on GSO dataset.

Method	**PSNR↑**	**SSIM↑**	**LPIPS↓**	**Clip-Sim↑**	**CD↓**	**Vol. IoU↑**	**F-Score↑**
Unique3D	19.6744	0.9217	0.1101	0.8864	0.0118	0.5463	0.6833

Figure 1: **Ablation on Colorize**. We show a comparison of whether or not to apply ExplicitTarget in coloring, and we can see that the group that does not use ExplicitTarget has significant artifacts, as there is no precise consistency across multiple views.

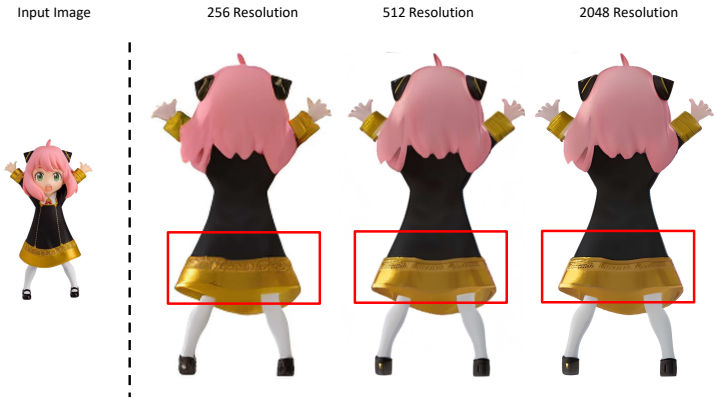

Figure 2: **Ablation on Resolution**. The visualization of the generated multi-views images at different stages is shown. Multi-level super-resolution does not change the general structure, but only improves the detail resolution, allowing the model to remain well-detailed.