GaussianProperty: Integrating Physical Properties to 3D Gaussians with LMMs

Supplementary Material

A. Derivation of Grasping Force

In the derivation below, we assume physical properties are uniform distributed over the entire object to grasp, which can be easily extended to more generic situations.

The lower bound of the grasping force $F_{\min}$ is the minimal sufficient force applied on the gripper to lift the object without slipping.

$$mg\cos\theta = \mu(2F_{\min} + mg\sin\theta)$$

$$\begin{aligned} F_{\min} &= \frac{1}{2}mg\left(\frac{\cos\theta}{\mu} - \sin\theta\right) \\ &= \frac{1}{2}\rho V g\left(\frac{\cos\theta}{\mu} - \sin\theta\right) \end{aligned}$$

where m, ρ and V are the mass, the density and the volume of the object respectively, θ is the lifting angle of the gripper with upward at 0 degree, μ is the friction coefficient between the gripper finger tips and the object surface, and $g \approx 9.8m/s^2$ is the gravity constant.

The upper bound of the grasping force $F_{\max}$ is the maximal force that does not cause any damage resulted by exceeding the yield stress σ_y or any undesirable deformation over some maximum allowable bending curvature $\kappa_{\max}$ of the object. Following the formula of bending stress

$$\frac{\sigma}{y} = \frac{E}{R}$$

the corresponding maximum stress applied on the force bearing surface at curvature $\kappa_{\max}$ is

$$\sigma_c = \frac{Ey(s)}{R_{\min}} = \frac{1}{2}Ed\kappa_{\max}$$

Therefore, the maximal grasping force

$$\begin{aligned} F_{\max} &= A\sigma_{\max} \\ &= \min\left[A\sigma_y, A\sigma_c\right] \\ &= \min\left[A\sigma_y, \frac{1}{2}AEd\kappa_{\max}\right] \end{aligned}$$

where A is the area of the force bearing surface of the object (or equivalently the area of one side of the robot gripper finger tips), σ is the bending stress at a point of the object at perpendicular distance y from the neutral axis, s is the outmost point of the force bearing surface, d is the thickness of the force bearing surface of the object, $R = 1/\kappa$ is the radius of curvature of the neutral axis, and E is Young's modulus of electricity of the object material.

Figure 1. The robot platform (left) and the robotic gripper (right) utilized in robot grasping experiments.

To maximize the grasping reliability, a reasonable choice of grasping force would be $\bar{F} = (F_{\min} + F_{\max})/2$. Additionally, the grasping force must be confined within the input bounds of the robotic gripper, and we also attempt to avoid the gripper executing commands close to its input bounds, with preferably $0 \leq \eta \leq 1$ margin. These three principals yield an optimal choice of grasping force

$$F^* = \begin{cases} \left[\bar{F}\right]_{[F_{\min}]_G+\eta\Delta F}^{[F_{\max}]_G-\eta\Delta F} & F_{\min} < F_{\max} \\ \left[\bar{F}\right]_G & F_{\min} \geq F_{\max} \end{cases}$$

with $[f]_G$ clipping a force f between the minimum and the maximum grasping forces of robotic gripper G, $[f]_{f_{\min}}^{f_{\max}}$ clipping f between $f_{\min}$ and $f_{\max}$, and $\Delta F = \max\left[0, [F_{\max}]_G - [F_{\min}]_G\right]$. In reality, it is possible to observe $F_{\min} > F_{\max}$, rendering infeasibility to picked up an object without damaging it. And F^* remains optimality in such situations.

B. Robot Grasping Experiment Details

In robot grasping experiments, we utilized a Jacobi.ai JSR-1 robot platform equipped with a TEK CTAG2F90-C robotic gripper (see Figure 1). The force-bearing surface at the tip of the gripper is measured to encompass an area of $A = 110\text{mm}^2 = 0.00011\text{m}^2$. And a maximum allowable bending curvature $\kappa_{\max} = 0.5$ is used.

B.1. Grasping Force Calibration

The robotic gripper employed in this study offers the capability to specify the grasping force on a normalized scale $0 \leq N_{\text{GF}} \leq 100$. Prior to conducting the grasping experiments, we performed a calibration on its grasping force, where 5 measurements are taken for each normalized input data point. The calibration curve is shown in Figure 2. We

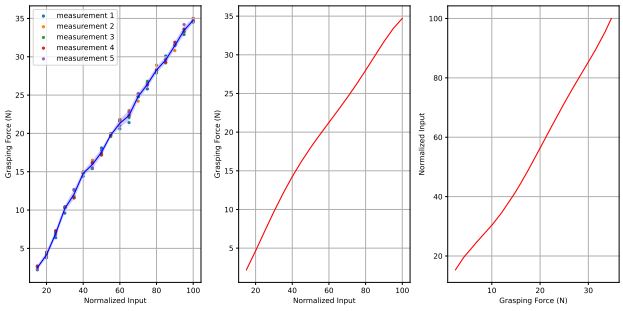

Figure 2. Calibration curve of robotic gripper grasping force (left) and its 5th-order polynomial smoothings (middle and right).

also note that there is a minimum enabling normalized input, and the robotic gripper is only enabled with normalized input $N_{\text{GF}} \geq 15$.

B.2. Full Object List and Experiment Results

We collected real-world 16 objects for the robot grasping experiments, as illustrated in Figure 3. This collection represents a diverse range of weights and materials, including plastic, ceramic, paper, steel, wood, and glass, etc. These objects are commonly encountered in everyday life, and the material properties of their different parts exhibit significant variability. Consequently, naive grasping strategies that do not account for material adaptability may find struggling to grasp all of these items in an effective and safe manner.

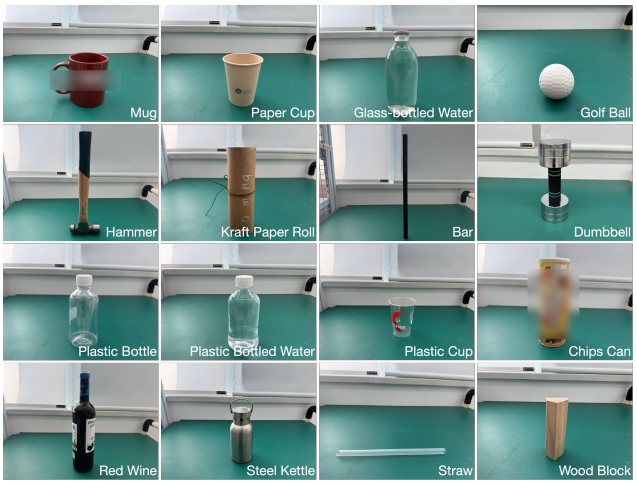

Figure 3. List of selected objects for robot grasping experiments.

We compare our proposed method on integrating *GaussianProperty* to material-sensitive robot grasping with three baselines, namely MinGF (with the minimum grasping force, $N_{GF} = 15$), MidGF (with medium grasping force, $N_{GF} = 60$) and MaxGF (with maximum grasping force, $N_{GF} = 100$). Table 2 in the main PDF listed the detailed experiment results. As summarized in Table 2, our method outperforms all the baselines and achieves a success rate of 100% on all the test objects, by successfully picking them up without slippery or causing any damage or undesirable deformation to them. Figure 4 shows the results of the complete robot grasping experiment.

B.3. Visualization Of Full Pipeline

Fig. 5 presents two examples showing the full pipeline including 3D reconstruction, material prediction, and grasping force estimation. The second example involves a transparent plastic cup and illustrates the challenge of geometry recovery. As expected, Gaussian Splatting struggles with such objects. We include this case to highlight current limitations and motivate future research.

C. More Results of Experiments

C.1. Datasets

For mass estimation, we use the ABO dataset, which provides mass data for each object. Since the NeRF2Physics method does not include a corresponding hardness dataset, we constructed our own dataset for hardness estimation using a similar methodology. Our dataset includes 10 household items, each captured in a realistic home setting. It features multi-view images paired with Shore hardness measurements. We captured the images and their corresponding poses with an iPhone 13 camera. For each object, hardness was measured at 10 specific points using a hardness tester, with each measurement averaged over three trials. Each measurement point is annotated with pixel coordinates in the images. Notably, Shore A and Shore D hardness testers use different indenters: Shore A measures within a range of 0-100, while Shore D spans a range of 100-200.

C.2. Evaluation Metrics

We report the following metrics, where p is the ground-truth mass/hardness and $\hat{p}$ is the estimated mass/hardness:

- Absolute difference error (ADE): $|p - \hat{p}|$,
- Absolute log difference error (ALDE): $|\ln p - \ln \hat{p}|$,
- Absolute percentage error (APE): $\left|\frac{p-\hat{p}}{p}\right|$,
- Min ratio error (MnRE): $\min\left(\frac{p}{\hat{p}}, \frac{\hat{p}}{p}\right)$, and
- Pairwise Relationship Accuracy (PRA):

$$\text{PRA} = \frac{1}{N_{\text{pairs}}} \sum_{i \neq j} \mathbb{1}\big((p_i > p_j) \iff (\hat{p}_i > \hat{p}_j)\big),$$

where N_{pairs} is the total number of object pairs, and $\mathbb{1}(\cdot)$ is the indicator function, which equals 1 if the condition inside is true, and 0 otherwise.

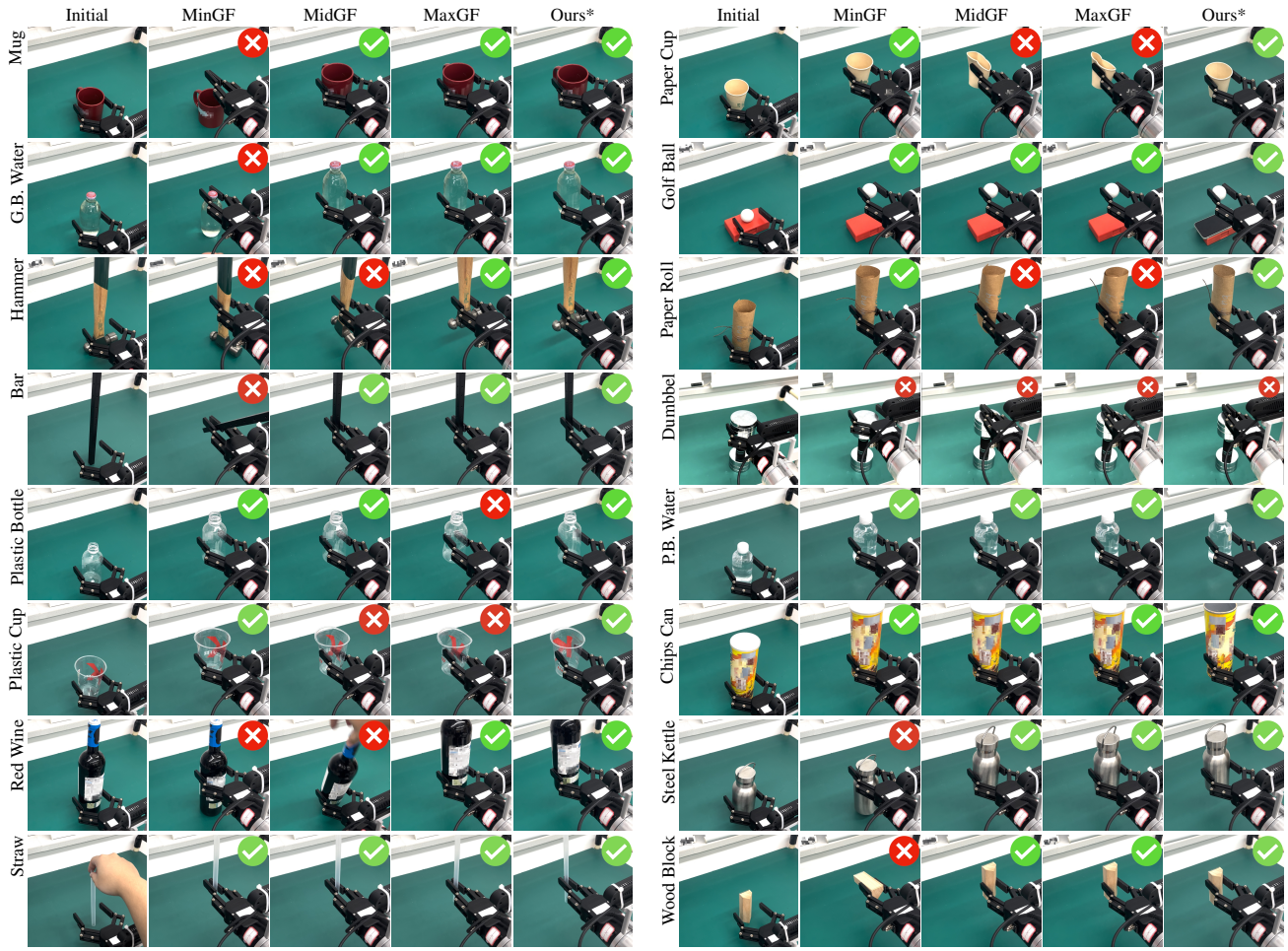

Figure 4. **Complete robot grasping experiment results.** The 16 test cases along with results in robot grasping experiments are listed. We compare our proposed method (right) against three baselines (middle columns), starting from initial configurations (left). **You can view the MP4 videos of the experiments in our supplementary material.**

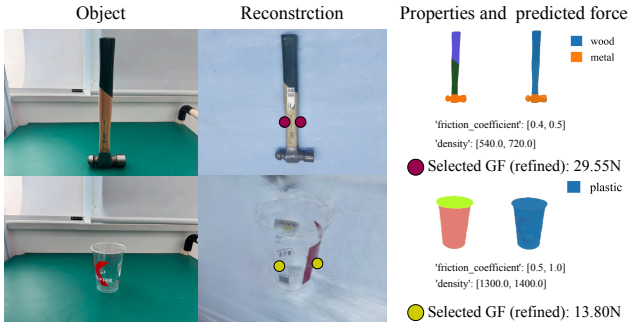

Figure 5. **Reconstruction, properties, predicted forces.**

C.3. Hardness Estimation

Table 1 presents the quantitative results of our method and NeRF2Physics on the hardness estimation task. Our approach outperforms NeRF2Physics across all metrics, demonstrating a significantly improved capability to accurately assess object attributes. This improvement can be attributed to the integration of LMMs, our method can have a more accurate understanding of each part of the object and form an accurate and clear-cut hardness estimation. Figure 6 illustrates the hardness estimation results produced by our method on the same case without the application of voting .

Table 1. Estimation of per-point Shore hardness on the real-captured in-house collected dataset (10 objects, 100 points). **Bold**: best model.

Method	ADE (↓)	ALDE (↓)	APE (↓)	MnRE (↑)	PRA (↑)
NeRF2Physics	35.917	0.328	0.294	0.748	0.575
Ours*	**28.583**	**0.220**	**0.198**	**0.820**	**0.686**

Figure 6. **Qualitative comparison of hardness prediction**. Compared to NeRF2Physics, our method provides more accurate hardness prediction with clear boundaries.

C.4. Mass Estimation

3D Gaussian object reconstruction allows for the estimation of the volumes of various parts composing an object. By integrating this with material property prediction where densities of different object parts are inferred, we can derive an overall estimation of object mass. We subsequently compare our mass estimation with the baseline NeRF2Physics, Table 2 demonstrating that our method provides more accurate quality assessments and significantly outperforms the baseline across most indicators.

Table 2. Mass estimation on ABO dataset. **Bold**: best results.

Method	ADE (↓)	ALDE (↓)	APE (↓)	MnRE (↑)
NeRF2Physics	12.761	0.803	**0.589**	0.498
Ours*	**5.960**	**0.744**	1.609	**0.559**

C.5. Evaluation with Different LMMs

As shown in Table 3, our method yields consistent results across different LMMs. GPT-4o achieves the highest mIoU, while the open-source Qwen-VL-72B performs competitively, confirming the method's robustness and generality.

Table 3. Comparison of different LMMs on per-material mIoU.

LMM Backbone	Wood	Metal	Plastic	Fabric	Ceramic	Avg. mIoU↑
Qwen-VL-72B	57.22	33.47	38.46	56.90	88.40	54.89
GPT-4V	61.53	33.41	38.26	67.57	78.40	55.83
GPT-4o	**70.43**	**37.32**	**38.51**	**67.10**	**96.19**	**61.91**

C.6. Prompting Format and Ablation

We conducted an ablation study comparing three prompting strategies: (1) querying GPT-4V with only part-level crops; (2) first showing the full image before asking about the part (global-then-local); and (3) our structured global-to-local approach. As shown in Table 4, our method significantly outperforms simpler alternatives, confirming the effectiveness of our design.

Table 4. Ablation of prompting strategies for material reasoning.

Prompt Strategy	Wood	Metal	Plastic	Fabric	Ceramic	Avg. mIoU↑
Part-only	37.90	13.82	17.03	27.40	14.69	22.17
Global-then-local (2-step)	52.03	26.17	31.28	44.95	63.71	43.63
Ours (global-to-local)	**61.53**	**33.41**	**38.26**	**67.57**	**78.40**	**55.83**

C.7. Effect of SAM in NeRF2Physics

We implemented a SAM-enhanced variant of NeRF2Physics. As shown in Table 5, it improves performance (25.59 → 30.31 mIoU) but remains below ours (55.83). This shows that our gains come not from segmentation alone, but from the full pipeline, including multimodal reasoning and 3D aggregation. Fig. 7 further shows that while SAM improves part boundaries, material prediction remains unreliable without multimodal inference.

Table 5. NeRF2Physics vs. its SAM-enhanced variant.

Method	Wood	Metal	Plastic	Fabric	Ceramic	Avg. mIoU↑
NeRF2Physics (original)	27.87	13.01	8.38	40.26	38.44	25.59
NeRF2Physics + SAM	**31.41**	**15.11**	**18.69**	**45.54**	**40.78**	**30.31**

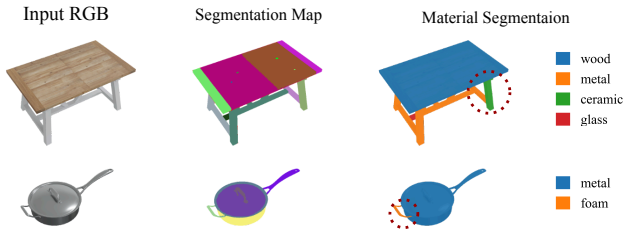

Figure 7. **SAM-enhanced variant performance.**

C.8. Effect of Voting Strategy.

Running SAM independently on each view may introduce occasional segmentation errors. Our voting-based multi-view aggregation mitigates this by projecting masks into 3D and suppressing inconsistent predictions via majority consensus, improving overall robustness. A representative example is shown in Fig. 8. Since Gaussian Splatting reconstructs only visible surfaces, unobserved regions (e.g., object interiors) are not represented and thus not assigned physical properties. This is consistent with surface-based reconstruction pipelines.

Figure 8. *Case of voting for material consistency.*

D. Additional details of Our Method

D.1. Segmentation Process Using SAM at Different Levels

We employ the Segment Anything Model (SAM) to generate segmentations at three levels of granularity: large-level,

middle-level, and small-level (Figure 9). Large-level segmentation simplifies object grouping but lacks detail, while small-level segmentation captures fine details at the cost of increased computational complexity. To balance object understanding and efficiency, we select the middle-level segmentation, which preserves meaningful part-level details without excessive fragmentation, making it ideal for our tasks.

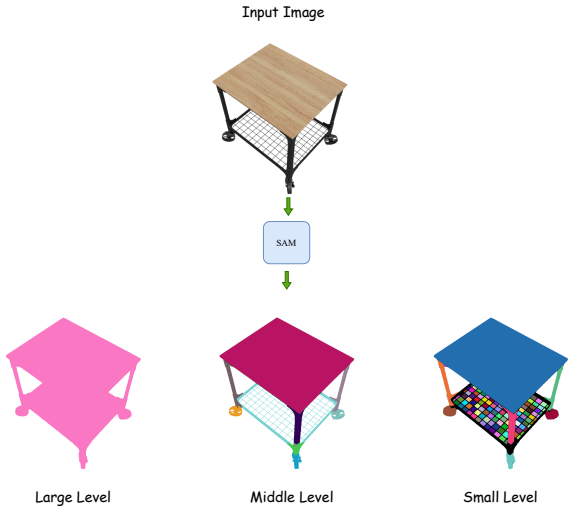

Figure 9. **Segmentation process using SAM at different levels** of granularity. From left to right: the input image, large-level segmentation, middle-level segmentation, and small-level segmentation. For our model, we selected the middle-level of SAM prediction to balance part-level object understanding and computational efficiency.

E. Detail of Data Labeling

We utilize the open-source interactive segmentation tool EISeg to annotate certain views of each object from ABO and MVImgNet, as shown in Figure 10. Since some materials are difficult to distinguish by the naked eye, such as aluminum and iron within the metal category. We established ten precise and unambiguous labels for a fair comparison. The labels are: wood, metal, plastic, glass, fabric, foam, marble, ceramic, concrete, and leather.

F. Prompting Details

We provide the prompts used for material proposal with other physical propertie such as hardness, density,Young's modulus and Poisson's Ratio in Figure 11.

G. Effects of Frequency-based Voting Strategy

Figure 12 showcases that implementing a frequency-based voting strategy can enhance the accuracy of property estimation. By projecting to multi-view images, we can determine the most frequently occurring material for each part. This frequency-based approach ensures consistency and reliability in the predicted properties, effectively aggregating information from different viewpoints, minimizing errors and improving overall prediction accuracy.

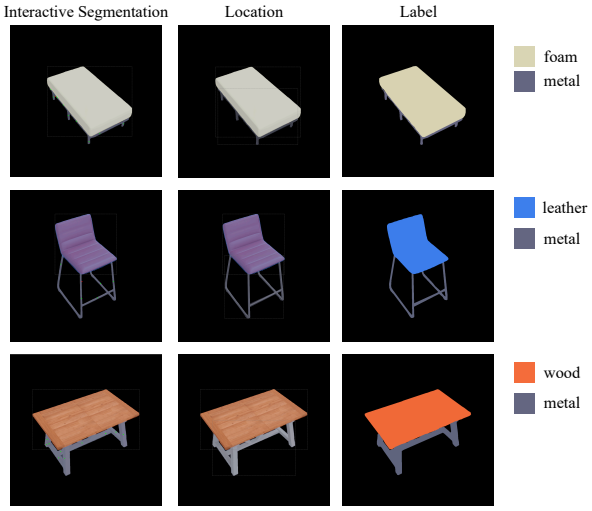

Figure 10. **Examples of data labeling**. These objects are sourced from the ABO-500 dataset.

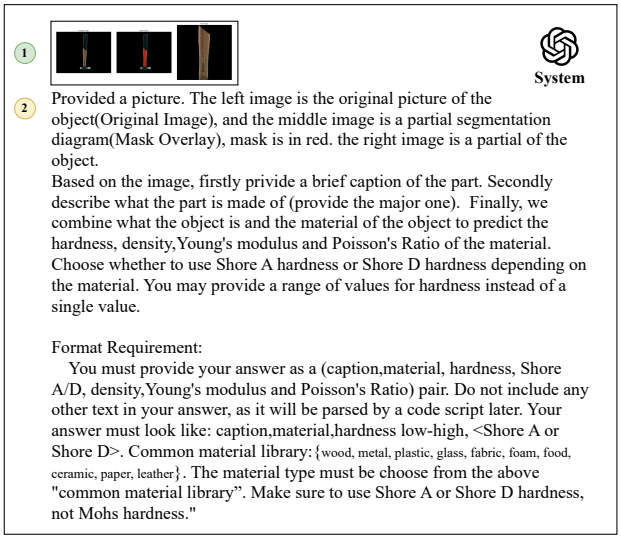

Figure 11. Prompt used for proposing materials and other physical properties.

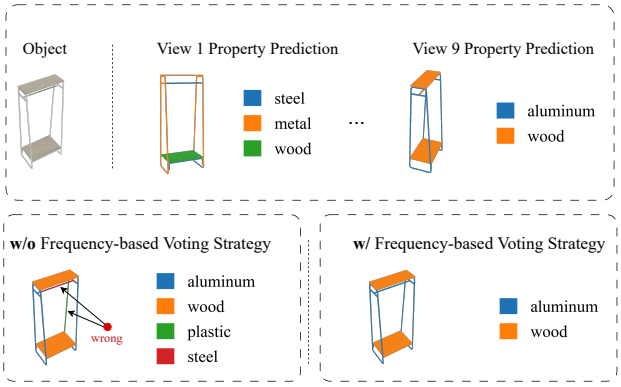

Figure 12. **Effects of Frequency-based Voting Strategy**. We provide an example to demonstrate the effectiveness of the frequency-based voting strategy. The result misclassified the "aluminum" and "wood" into "plastic" and "'steel' without voting strategy.

H. More qualitative results of Material Segmentation

In the supplementary material, we provide additional performance comparisons with the baseline model Nerf2Physics. As shown in Figure 13, our method predicts the physical properties of objects more accurately. We also show some cases on MVImgNet dataset in Figure 14.

I. Failure cases

However, our method still has limitations. For instance, when the surface texture of an object is ambiguous, it can lead to incorrect classification of material categories, as illustrated in Figure 15.

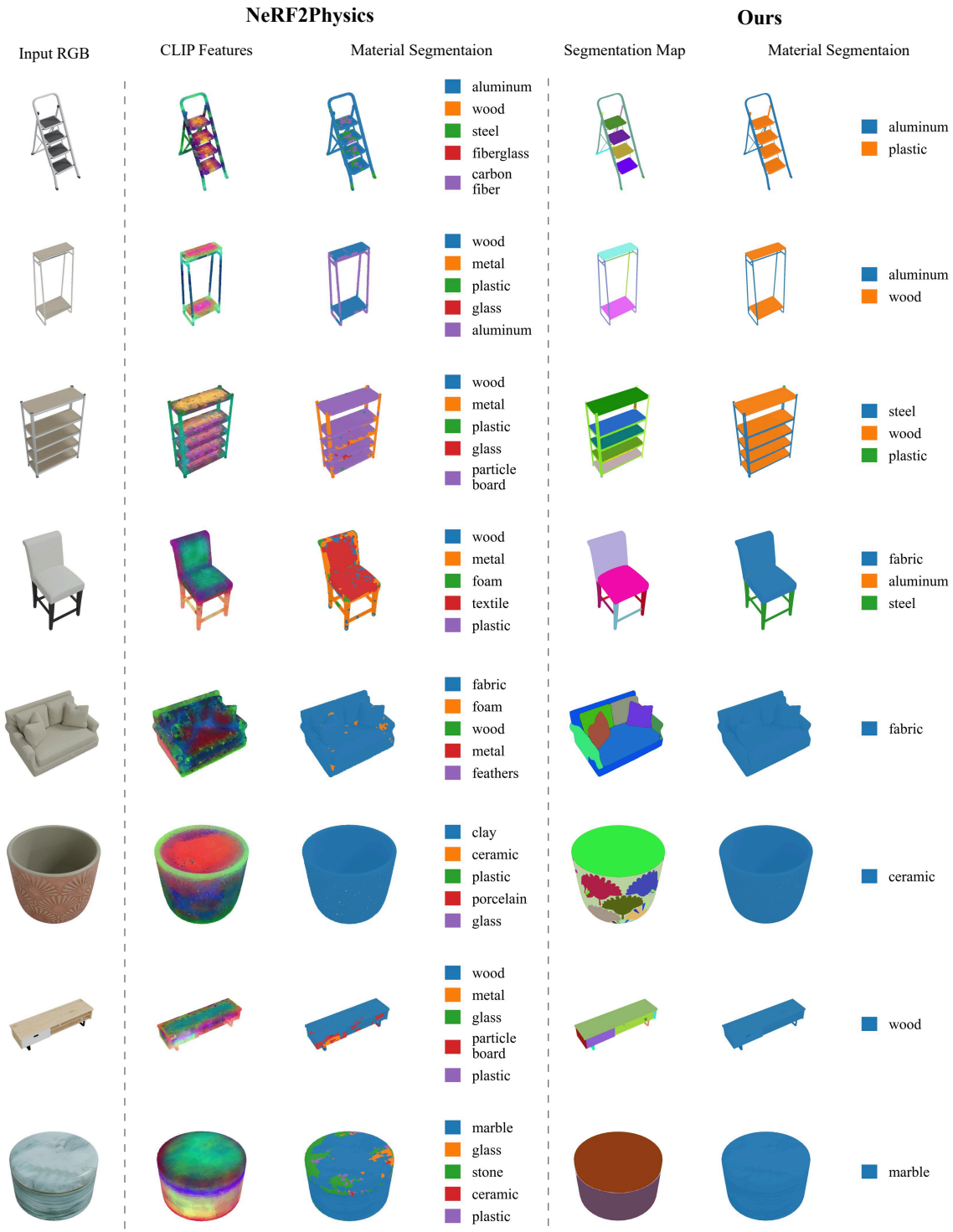

Figure 13. **Qualitative comparison of Material Segmentation**. These objects are sourced from the ABO-500 dataset.

Figure 14. **Qualitative results of object material segmentation** on MVImgNet. Our model makes reasonable and boundary-accurate material predictions for objects with multiple or single materials.

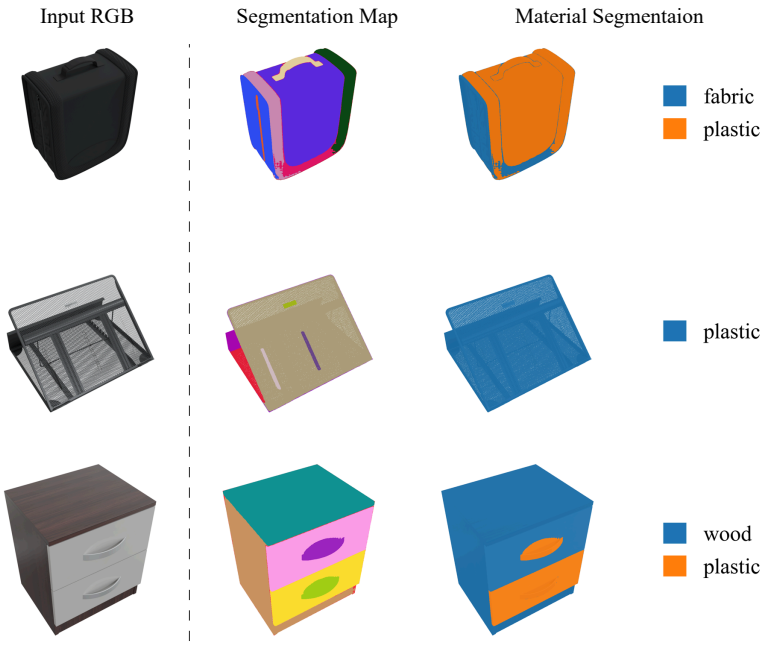

Figure 15. **Examples of Challenging Material Segmentation Cases**. These objects are sourced from the ABO-500 dataset.