This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

PhysGS: Bayesian-Inferred Gaussian Splatting for Physical Property Estimation

Samarth Chopra[1, ✉], Jing Liang[2, *], Gershom Seneviratne[1], Dinesh Manocha[1]
[1]University of Maryland, College Park [2]Stanford University
sachopra@umd.edu, jinglms@stanford.edu, gershom@umd.edu, dmanocha@umd.edu

Abstract

Understanding physical properties such as friction, stiffness, hardness, and material composition is essential for enabling robots to interact safely and effectively with their surroundings. However, existing 3D reconstruction methods focus on geometry and appearance and cannot infer these underlying physical properties. We present PhysGS, a Bayesian-inferred extension of 3D Gaussian Splatting that estimates dense, per-point physical properties from visual cues and vision–language priors. We formulate property estimation as Bayesian inference over Gaussian splats, where material and property beliefs are iteratively refined as new observations arrive. PhysGS also models aleatoric and epistemic uncertainties, enabling uncertainty-aware object and scene interpretation. Across object-scale (ABO-500), indoor, and outdoor real-world datasets, PhysGS improves accuracy of the mass estimation by up to 22.8%, reduces Shore hardness error by up to 61.2%, and lowers kinetic friction error by up to 18.1% compared to deterministic baselines. Our results demonstrate that PhysGS unifies 3D reconstruction, uncertainty modeling, and physical reasoning in a single, spatially continuous framework for dense physical property estimation. Additional results are available at https://samchopra2003.github.io/physgs.

1. Introduction

Understanding the physical properties of real-world environments is critical for enabling robots to interact safely and effectively with their surroundings [5, 10, 38, 66]. This capability is essential across a wide range of domains such as navigation [36, 59], manipulation [1, 32], and surgical robotics [32, 41, 65], and is required in both complex indoor [42, 63, 66] and outdoor [10, 36, 50] environments. In particular, accurately estimating physical attributes such as friction [10, 36], elasticity [59], hardness [5], and density [63] is essential for safe and robust robot interaction with diverse and unstructured real-world scenarios.

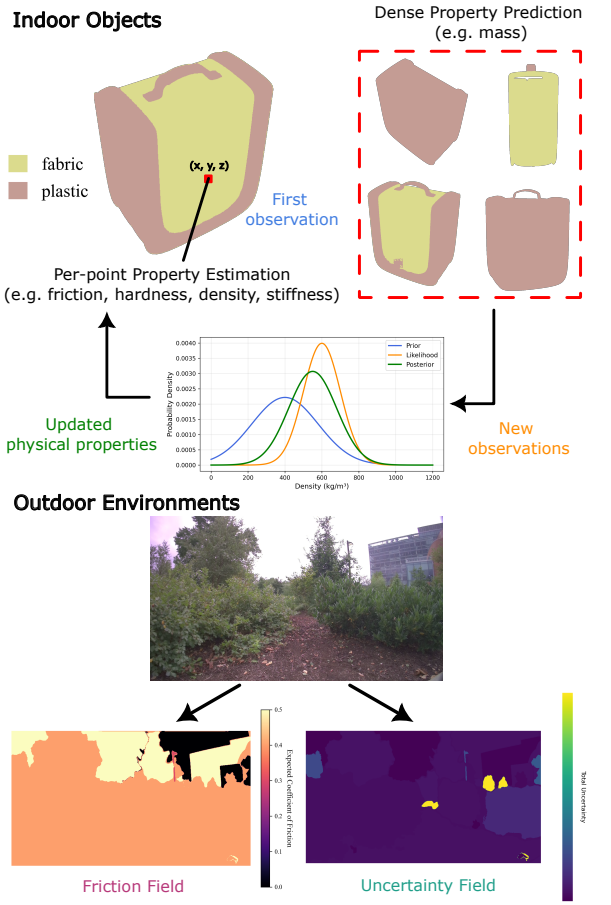


Figure 1. Overview of *PhysGS*. Top: Our method estimates per-point and dense physical properties (e.g., friction, hardness, density, stiffness, and mass) by combining vision–language material priors with Bayesian updates over 3D Gaussian splats. Bottom: *PhysGS* can also be deployed in outdoor environments to infer scene-level properties such as friction and predictive uncertainty; we visualize the total uncertainty (*aleatoric* + *epistemic*).

In general, estimating physical properties from visual sensors remains a challenging task due to two primary reasons [10, 63]. First, visually similar but physically distinct regions (e.g., mud vs. asphalt, or grass vs. rock)

*Jing L. is affiliated with Stanford Robotic Center and Stanford Center of Longevity

are often difficult to distinguish, leading to brittle planning and control in complex scenes [10, 11]. While conventional 3D mapping approaches, such as occupancy grids [8], signed distance fields [40], or implicit neural representations [27, 39], can provide rich geometric detail, they primarily focus on recovering object shape, and typically fail to encode physical properties or material categories. Second, although recent work has made progress in estimating physical properties in indoor environments [42, 63, 66], outdoor scenes remain underexplored. Existing methods often focus on one or two specific physical properties of outdoor objects or terrains, such as friction [5, 10, 36], pliability [59], or stiffness [5], and are not easily generalizable to a broader set of physical attributes.

Recent advances in 3D vision and semantic perception have increasingly focused on linking visual appearance to underlying physical attributes, enabling models to distinguish between rigid, deformable, slippery, and compliant surfaces [5, 10, 11, 63, 66]. In parallel, vision-language models (VLMs) have demonstrated the ability to capture latent physical properties, such as friction and elasticity, with multimodal inputs [66]. These models can qualitatively reason about forces, materials, and object dynamics from language and imagery [42, 51, 58, 62, 67], highlighting their potential as semantic priors for physical inference.

However, a core challenge in estimating physical properties from visual sensors is managing the uncertainty inherent in both sensing and inference. Visual and depth observations are often degraded by sensor noise, lighting variation, occlusion, and calibration drift [6, 55–57]. This type of uncertainty, known as *aleatoric uncertainty*, captures measurement noise and perceptual ambiguity [3, 25]. Simultaneously, learning-based models trained on limited or domain-specific datasets frequently struggle to generalize to novel textures, materials, and environmental conditions [18, 25, 26]. This is referred to as *epistemic uncertainty*, which reflects the model's incomplete or imperfect knowledge of the world, often due to insufficient or biased training data [3, 25].

Main Contributions: We propose a 3D physical property estimation framework that integrates Bayesian inference into the Gaussian Splatting optimization process, treating each Gaussian primitive as a probabilistic entity whose properties are updated via posterior refinement. This enables PhysGS to estimate both point-level physical properties (e.g., friction, hardness, stiffness, density) and object-level quantities (e.g., total mass), while producing calibrated aleatoric and epistemic uncertainty. Our novel contributions include:

1. **Bayesian-Inferred Gaussian Splatting.** We embed Bayesian updates within the Gaussian Splatting pipeline, allowing each Gaussian's physical property values to be updated through confidence-weighted posterior refinement from observations.
2. **Unified multi-property estimation across scales.** A single Bayesian formulation supports diverse physical properties, including friction, hardness, stiffness, density, and mass, at both the point level and the object level, enabling fine-grained property mapping and global aggregation within the same framework.
3. **Generality across environments and object types.** PhysGS is broadly applicable to a wide range of indoor and outdoor scenes and operates on both rigid and deformable objects, including vegetation, soil, and everyday household materials, enabling consistent physical property estimation across heterogeneous real-world settings.

Across all datasets, including ABO-500, and a real-world friction–hardness dataset, *PhysGS* achieves strong gains over prior methods such as NeRF2Physics, CLIP-based recognition, and direct VLM regression. We observe improvements of up to 61.2% in Shore hardness error, 18.1% in kinetic friction error, and 22.8% in mass-density error.

2. Related Work

2.1. Visual Property Fields

A major line of work estimates physical properties by associating scene objects with open-vocabulary physical semantics, querying where specific physical property appear in observed spaces. LERF grounds CLIP embeddings [47] within NeRF, distilling multi-scale language features into a dense, queryable 3D field that produces 3D relevancy maps for text prompts [28]. Closely related open-vocabulary 3D mapping approaches propagate physical–semantic features into 3D reconstructions for zero-shot recognition and retrieval [22, 45].

With the rise of 3D Gaussian Splatting (3DGS)[27], several methods directly inject language or semantic features into Gaussian primitives, yielding fast, explicit, and queryable 3D fields. LangSplat [46] distills 2D CLIP features into a 3D language field over Gaussians for open-vocabulary search. Related efforts [16, 21] assign semantic Gaussians for open-vocabulary 3D understanding, showing that explicit Gaussian fields are well-suited for encoding and rendering high-dimensional properties beyond color.

Based on visual–semantic cues, recent work has further incorporated language-context features to enhance physical property estimation [63, 66]. NeRF2Physics[66] constructs a language-embedded 3D feature space and performs zero-shot kernel regression to estimate per-point physical properties, while GaussianProperty[63] extends this idea to 3D Gaussians.

2.2. Uncertainty-Aware and Probabilistic Scene Understanding

In vision and robotics, uncertainty is typically decomposed into aleatoric (data or sensor noise) and epistemic (model or knowledge) components. This formulation has become standard for loss design, model calibration, and risk-sensitive decision making in perception systems [3, 4, 25]. Prior work formalizes these notions for vision-based tasks and demonstrates how to jointly learn uncertainty with outputs such as depth or segmentation, or to approximate Bayesian inference via dropout or ensembles, thereby improving robustness and out-of-distribution behavior [14, 26, 33].

Estimating the physical properties of real-world environments from sensors inherently involves uncertainty. Several approaches introduce probabilistic maps, tail-risk measures, and confidence-aware policies to quantify and mitigate this uncertainty [5, 9, 12, 13, 43]. STEP [12] models traversability as a stochastic variable and plans using a CVaR-based risk formulation, validated across diverse field environments and the DARPA SubT Challenge. Evidential and Bayesian formulations extend this concept by outputting full distributions rather than point estimates, enabling online belief updates from new observations [3, 4, 10, 11]. EVORA [3] learns evidential traction distributions, explicitly separating aleatoric and epistemic components to assess motion risk, while Ewen et al. [10, 11] maintain joint beliefs over semantics and continuous properties to predict physical property maps (e.g. friction).

Beyond 2D perception, uncertainty has been integrated into NeRF and 3D Gaussian Splatting (3DGS) frameworks to quantify ambiguity arising from sparse observations, occlusions, and under-constrained geometry, enabling tasks such as active perception and view planning through information-theoretic measures and Bayesian modeling of scene primitives [24, 34, 35, 49]. Recent works estimate spatial uncertainty post hoc for pre-trained NeRFs [15, 34], propose probabilistic NeRFs [19], or directly model uncertainty in 3DGS [29] via variational or evidential objectives, including dynamic and 4D settings [30]. Complementary efforts have explored uncertainty-aware on variational Gaussian splatting and SLAM pipelines that propagate uncertainty in pose and structure, demonstrating that per-Gaussian uncertainty can substantially enhance mapping robustness and downstream reasoning [20, 48, 49].

3. Proposed Approach

In this section, we outline our Bayesian framework for dense physical property estimation. We begin with a Dirichlet–Categorical model for fusing confidence-weighted material labels across views (Sec. 3.1, 3.2), then extend it to continuous properties using a Normal–Inverse–Gamma prior to obtain calibrated aleatoric and epistemic uncertainty (Sec. 3.3). We then describe how these Bayesian updates integrate with 3D Gaussian Splatting, segmentation, and VLM prompting to produce per-point property fields and object-level estimates (Sec. 3.4).

3.1. Preliminaries

Dirichlet–Categorical formulation. We model discrete material labels produced by the VLM using a Categorical distribution and place a Dirichlet prior over its parameters. The Dirichlet distribution is the conjugate prior to the Categorical likelihood, enabling closed-form Bayesian updates as new observations are incorporated across views.

The Categorical distribution parameterized by $\boldsymbol{\theta} \in [0,1]^K$ represents the probability that an observation belongs to class i:

$$f(z = i \mid \boldsymbol{\theta}) = \theta_i. \tag{1}$$

The Dirichlet distribution defines a continuous K-variate prior over $\boldsymbol{\theta}$, parameterized by $\boldsymbol{\alpha} \in \mathbb{R}^K_{>0}$, as

$$f(\boldsymbol{\theta} \mid \boldsymbol{\alpha}) = \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K \theta_k^{\alpha_k - 1}, \tag{2}$$

where $\Gamma(\cdot)$ is the Gamma function.

Given a set of n observed material labels $\mathcal{Z} = \{z_1, \ldots, z_n\}$ drawn from a Categorical distribution, the posterior predictive probability that a new observation belongs to material class i is

$$f(z = i \mid \mathcal{Z}, \boldsymbol{\alpha}) = \int_{\boldsymbol{\theta}} f(z = i \mid \boldsymbol{\theta})\, f(\boldsymbol{\theta} \mid \mathcal{Z}, \boldsymbol{\alpha})\, d\boldsymbol{\theta}. \tag{3}$$

Using conjugacy of the Dirichlet prior and Categorical likelihood, the integral simplifies to the closed-form expression

$$f(z = i \mid Z, \boldsymbol{\alpha}) = \frac{\tilde{\alpha}_i}{\sum_{j=1}^K \tilde{\alpha}_j}, \tag{4}$$

where the posterior parameters are recursively updated as

$$\tilde{\alpha}_i \leftarrow \alpha_i(0) + \sum_{m:\, c_m = i} \lambda\, p_m, \tag{5}$$

with λ controlling the evidence strength contributed by each observation and p_m denoting the confidence provided by the VLM for the m-th prediction.

3.2. Bayesian Inference for Material Property Estimation

We introduce our hierarchical Bayesian framework for estimating material-specific physical properties from confidence-weighted observations. Building on the Dirichlet–Categorical model of [11], we extend it with a continuous posterior to jointly infer material class and properties such as friction, density, and hardness.

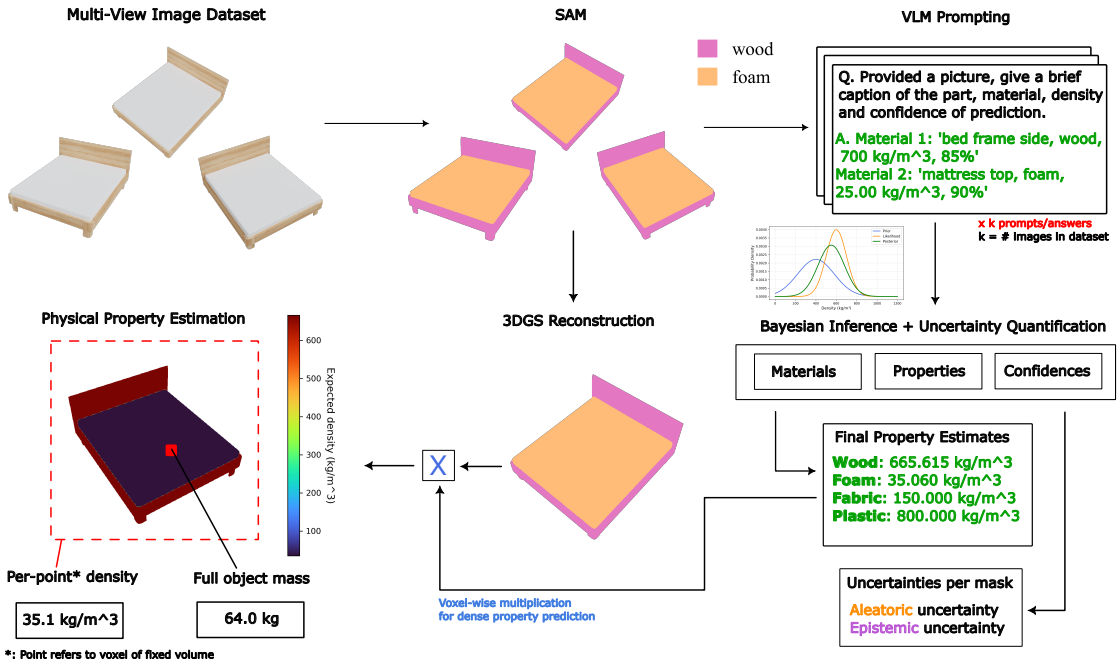


Figure 2. *PhysGS* architecture. Given multi-view images, SAM provides part-level segmentations that are used for 3D Gaussian Splatting (3DGS) reconstruction. For each segmented part, a VLM produces material labels, density estimates, and confidence scores across multiple views. These observations are fused using Bayesian inference with uncertainty quantification to obtain final per-material property distributions. By propagating the estimated densities over the reconstructed 3D Gaussian field, *PhysGS* predicts per-point density and full-object mass.

Continuous property estimation. While the Dirichlet–Categorical formulation governs the discrete class probabilities, we also require an estimate of the continuous physical property ψ associated with each material. For each material class i, we maintain confidence-weighted accumulators that enable incremental computation of the first and second moments using a running mean and variance formulation proposed by [60] and generalized by [44]:

$$W_i = \sum_m p_m, \qquad S_i = \sum_m p_m \psi_m, \qquad Q_i = \sum_m p_m \psi_m^2, \tag{6}$$

representing the total weight, first moment, and second moment, respectively. These accumulators allow efficient online updates without requiring access to past observations, which is particularly beneficial in streaming or on-the-fly reconstruction settings.

The posterior mean and variance for material i are then estimated as

$$\mu_i = \frac{S_i}{W_i}, \qquad \sigma_i^2 = \max\left(\frac{Q_i}{W_i} - \mu_i^2, \epsilon\right), \tag{7}$$

yielding a Gaussian posterior

$$p(\psi_i \mid Z) = \mathcal{N}(\mu_i, \sigma_i^2), \tag{8}$$

which represents the system's belief over the continuous physical property for material i given all confidence-weighted evidence Z. This formulation integrates naturally with the Dirichlet update by providing confidence-weighted, incremental, and uncertainty-aware refinement as new observations become available.

Hierarchical posterior. The resulting model is hierarchical in nature, jointly estimating the discrete material identity z and continuous physical property ψ:

$$p(z, \psi \mid Z, \boldsymbol{\alpha}) = p(\psi \mid z, Z)\, p(z \mid Z, \boldsymbol{\alpha}), \tag{9}$$

where $p(z \mid Z, \boldsymbol{\alpha})$ is the Dirichlet–Categorical posterior and $p(\psi \mid z, Z)$ is the Gaussian posterior. Applying the Law of Total Probability as in [11], the overall predictive distribution over physical properties is

$$f(\psi \mid Z, \boldsymbol{\alpha}) = \sum_{i=1}^{K} f(\psi \mid z = i)\, f(z = i \mid Z, \boldsymbol{\alpha}). \tag{10}$$

Substituting Eq. (8) into Eq. (10) gives a closed-form multimodal Gaussian mixture for the predicted material properties:

$$f(\psi \mid Z, \boldsymbol{\alpha}) = \sum_{i=1}^{K} \frac{\tilde{\alpha}_i}{\sum_{j=1}^{K} \tilde{\alpha}_j} \mathcal{N}(\mu_i, \sigma_i^2). \tag{11}$$

This mixture formulation expresses the full posterior as a weighted sum of unimodal Gaussian components, where each mode corresponds to a material class and is weighted by its recursively updated class likelihood from the Dirichlet posterior.

3.3. Uncertainty-Aware Property Fields

Uncertainty modeling via the Normal–Inverse–Gamma prior. For each material i, the joint prior over the mean μ_i and variance σ_i^2 of the property ψ is given by

$$p(\mu_i, \sigma_i^2 \mid \tau_i, \kappa_i, \alpha_i, \beta_i) = \mathcal{N}\Big(\mu_i \,\Big|\, \tau_i, \tfrac{\sigma_i^2}{\kappa_i}\Big) \text{Inv-Gamma}\big(\sigma_i^2 \,\big|\, \alpha_i, \beta_i\big), \tag{12}$$

where τ_i denotes the prior mean, κ_i controls the precision on μ_i (i.e., the strength of accumulated evidence), and (α_i, β_i) are the shape and scale parameters governing the uncertainty in σ_i^2.

Predictive Uncertainty update. Given a new observation ψ_m associated with material class i and its confidence p_m, the posterior parameters $(\tilde{\tau}_i, \tilde{\kappa}_i, \tilde{\alpha}_i, \tilde{\beta}_i)$ can be updated in closed-form, allowing sequential fusion of confidence-weighted evidence without storing past data. This conjugate formulation provides closed-form expressions for the predictive mean and variance of the property ψ_i.

The total predictive uncertainty decomposes into two components:

$$\text{Var}[\psi_i] = \underbrace{\mathbb{E}[\sigma_i^2]}_{\text{aleatoric}} + \underbrace{\text{Var}[\mu_i]}_{\text{epistemic}}. \tag{13}$$

The first term represents *aleatoric uncertainty*, which arises from inherent noise in the observations and variability within each material class. The second term represents *epistemic uncertainty*, corresponding to uncertainty in the estimated mean that decreases as more high-confidence evidence is incorporated. For implementation, we compute these moments directly from the NIG parameters:

$$\mathbb{E}[\sigma_i^2] = \frac{\tilde{\beta}_i}{\tilde{\alpha}_i - 1}, \qquad \text{Var}[\mu_i] = \frac{\mathbb{E}[\sigma_i^2]}{\tilde{\kappa}_i}. \tag{14}$$

The resulting aleatoric, epistemic, and total predictive uncertainties provide interpretable measures of confidence in both the per-class property estimates and the overall reconstruction. Regions with high aleatoric uncertainty reflect sensor or perceptual noise, while high epistemic uncertainty indicates insufficient or conflicting evidence about the underlying material properties.

3.4. Learning Semantics and Physical Properties

Semantic Segmentation. Given a multi-view image dataset, we employ SAM [31] to produce pixel-accurate masks that decompose each object into hierarchical levels (whole, part, and sub-part), facilitating fine-grained semantic understanding. The model outputs multiple candidate masks at different granularities, which we refine by discarding redundant or low-confidence predictions using SAM's built-in IoU and stability measures. The resulting segmentation maps capture precise object boundaries and semantically coherent regions, forming the basis for our downstream physical property estimation.

VLM Prompting. For each segmented image, we construct a vision–language prompt comprising a triplet of images arranged side by side, following the design of [63]. The left image presents the complete object, the middle image overlays the segmentation mask, and the right image isolates the masked region of interest. Given an input image I, this process yields k visual prompts corresponding to the k masks predicted by SAM. We additionally condition the VLM with a structured textual query that instructs it to (i) provide a concise caption of the segmented part, (ii) identify its predominant material, and (iii) infer relevant physical properties such as friction, density, etc. The model is further asked to report a normalized confidence score within $[0, 1]$, representing its belief in the prediction.

3D Gaussian Splatting. Given the VLM responses and the refined physical property estimates from our Bayesian inference scheme, we construct a material legend assigning each material a unique color. The corresponding scene images are recolored accordingly and used as semantic inputs for 3DGS reconstruction. This yields a semantic splat that supports dense property inference, such as mass estimation.

Physical Property Estimation. Using the reconstructed 3D Gaussian field and inferred material properties, we perform per-point and dense physical property estimation. Each voxel is associated with a predicted property value (e.g., friction or density), enabling spatial queries for per-point properties or integration over the volume to obtain aggregate measures such as total mass.

4. Experiments and Results

4.1. Implementation Details

We employ the `splatfacto-big` variant of Nerfstudio [54] for 3D Gaussian Splatting, using default parameters except for a random scale of 2.0 and a random background color. Each scene is trained for 20,000 iterations on an NVIDIA RTX A5000 GPU.

For image segmentation, we use SAM [31] to obtain whole, part, and sub-part material decompositions. Material property estimation is performed using GPT-5 [52] as the vision–language model (VLM), conditioned on structured visual–text prompts derived from the segmented images.

4.2. Mass Estimation

Dataset. We employ the ABO dataset [7] for evaluating mass prediction, which includes a large set of consumer products listed on Amazon along with multi-view imagery, segmentation masks, physical measurements, and metadata. Specifically, we make use of the representative multi-view benchmark ABO-500 curated by [66], which selects a bal-

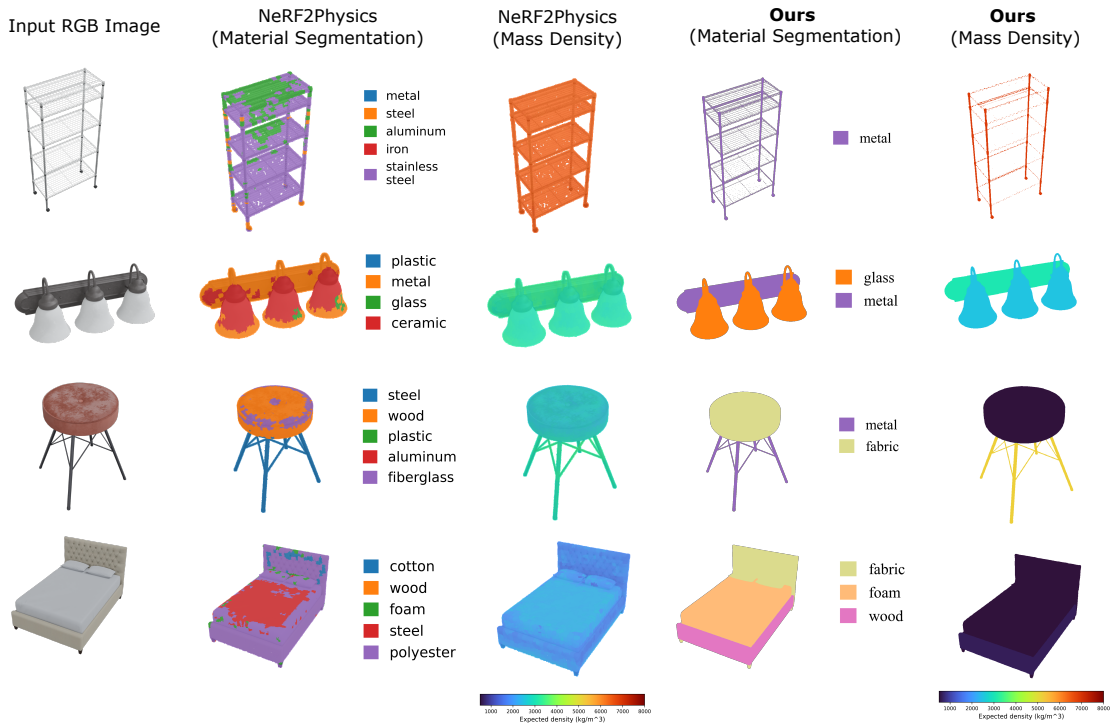


Figure 3. Qualitative comparison on the ABO-500 dataset. For each object, we show the input RGB image, material segmentation and mass-density predictions from NeRF2Physics, and the corresponding results from our method. *PhysGS* produces cleaner material segmentation with fewer artifacts compared to NeRF2Physics and more consistent part boundaries, and yields sharper, more plausible mass-density fields across diverse object categories.

anced subset of 500 items from the entire ABO dataset. It is divided into 300 training, 100 validation, and 100 testing instances.

Metrics. Following prior work on visual mass estimation [53], we evaluate using four complementary metrics that measure both absolute and relative error between the predicted mass $\hat{m}$ and the ground-truth mass m:

- Absolute Difference Error (ADE): $|m - \hat{m}|$,
- Absolute Log Difference Error (ALDE): $|\ln m - \ln \hat{m}|$,
- Absolute Percentage Error (APE): $\left|\frac{m-\hat{m}}{m}\right|$,
- Minimum Ratio Error (MnRE): $\min\left(\frac{m}{\hat{m}}, \frac{\hat{m}}{m}\right)$.

Baselines. We compare our system against several visual and multimodal baselines on the ABO-500 dataset:

- **Image2mass** [53]: a CNN that infers mass directly from RGB images and 3D bounding box dimensions.
- **2D CNN**: a lightweight regression model built upon a frozen ResNet50 [17] backbone, fine-tuned with additional layers for scalar mass prediction.
- **LLaVA** [37]: a vision-language model designed for instruction following.
- **NeRF2Physics** [66]: a NeRF-based approach that jointly estimates 3D geometry and per-point physical properties. It predicts mass by integrating predicted density across the reconstructed volume.

Qualitative Results. Figure 3 presents qualitative results on the ABO-500 dataset. For each object, we compare the material segmentation and mass-density predictions from NeRF2Physics with those produced by *PhysGS*. Our method yields substantially cleaner material segmentations with fewer spurious labels and more coherent part boundaries, while also producing sharper and more stable mass-density fields. These improvements are consistent across a wide range of object categories, demonstrating the advantage of combining vision-language priors with Bayesian inference over 3D Gaussian splats.

Quantitative Results. We evaluate the accuracy of our method on mass estimation using the ABO-500 test set (100 objects). As shown in Table 1, traditional 2D methods such as Image2mass [53] and 2D CNNs exhibit high mass estimation error due to their inability to capture 3D structure or material composition. VLM approaches (e.g. LLaVA [37]) show similar limitations, producing noisy predictions that vary across views. NeRF2Physics improves accuracy by exploiting neural radiance fields, and achieves the best ALDE (0.771) and MnRE (0.552) among existing baselines. *PhysGS* achieves the best performance on two key metrics: it reduces ADE from 8.730 to 8.254, corresponding to a 5.5% improvement, and reduces APE from 1.061 to 0.819, a substantial 22.8% improvement over

Table 1. Mass estimation on ABO-500 test set (100 objects). ADE is measured in kilograms. **Bold**: best model.

Method	ADE (↓)	ALDE (↓)	APE (↓)	MnRE (↑)
Image2mass [53]	12.496	1.792	0.976	0.341
2D CNN	15.431	1.609	14.459	0.362
LLaVA [37]	17.328	1.893	1.837	0.306
NeRF2Physics [66]	8.730	**0.771**	1.061	**0.552**
Ours	**8.254**	0.999	**0.819**	0.474

Table 2. Ablation study for mass estimation on ABO-500 val set (100 objects). ADE is measured in kilograms. BI refers to Bayesian Inference. **Bold**: best model.

Method	ADE (↓)	ALDE (↓)	APE (↓)	MnRE (↑)
NeRF2Physics [66]	9.786	**0.61**	0.931	**0.609**
Ours (w/o BI)	9.728	0.770	0.717	0.561
Ours (with BI)	**9.187**	0.827	**0.715**	0.539

NeRF2Physics.

Ablation Study. Table 2 shows that incorporating Bayesian inference yields clear gains over both NeRF2Physics and our non-Bayesian variant. Updating material and property beliefs across additional views reduces ADE by 5.6% and improves APE by 6.4% compared to the version without Bayesian updates. These improvements demonstrate that aggregating multi-view evidence to refine the posterior distribution leads to more accurate mass estimation, confirming the benefit of treating physical properties as latent variables that are iteratively updated rather than fixed from single-view predictions.

4.3. Friction and Hardness Estimation

Dataset. To evaluate our model's ability to infer dense, spatially varying physical properties within objects, we leverage the friction and hardness dataset containing paired image and real-world measurement data, curated by [66]. This dataset includes 15 household objects captured across 13 scenes, using multi-view RGB images paired with per-point measurements of kinetic friction coefficient and Shore A/D hardness.

Metrics. We report the same evaluation metrics for per-point friction and hardness estimation as above.

Baselines. As before, we compare our method to several visual and multimodal baselines.

- **GPT-4V** [64] and **GPT-5** [52]: Large VLMs capable of processing masked regions in its prompt.
- **CLIP** [47]: A vision–language baseline that uses global CLIP embeddings from the canonical view of the scene, rather than the fused multi-view patch features used in our method. This baseline evaluates how well static visual–semantic representations can generalize to physical property prediction.
- **NeRF2Physics** [66]: Same as in mass estimation baseline.



Figure 4. Qualitative results for the friction and hardness dataset. Given a single RGB view, *PhysGS* predicts dense friction coefficients and Shore hardness values for a variety of household objects. The resulting property fields are spatially smooth, physically plausible, and consistent across diverse materials and geometries.

Qualitative Results. Figure 4 presents qualitative results for friction and Shore hardness estimation on real objects. From a single RGB view, *PhysGS* produces dense per-point friction and hardness fields across materials such as rubber, leather, plastic, and metal. The predictions capture fine-grained material differences and exhibit clean boundaries between regions with distinct friction and hardness characteristics, reflecting the system's ability to localize subtle variations in contact and deformation properties.

Quantitative Results. Table 3 reports results for per-point Shore A/D hardness and kinetic friction estimation on our real-world dataset. Across all hardness metrics, PhysGS achieves substantial gains over existing approaches. Relative to the best baseline, our method reduces ADE by 61.2%, lowers ALDE by 34.4%, and decreases APE by 16.5%, while improving MnRE by 8.4%. For kinetic friction, PhysGS again outperforms the strongest baseline (NeRF2Physics) on the majority of metrics, reducing ADE by 15.5% and ALDE by 18.1%, and increasing MnRE by 9.4%.

4.4. Applications: Outdoor Scene Understanding

PhysGS can also estimate physical properties of outdoor environments, such as friction and stiffness, which are important for reasoning about natural, vegetation-rich terrain [5].

Table 3. Estimation of per-point Shore hardness (left) and kinetic friction coefficient (right) on the real-world dataset. **Bold**: best model.

Shore Hardness (31 points, 11 objects)					Kinetic Friction (6 points, 6 objects)				
Method	**ADE (↓)**	**ALDE (↓)**	**APE (↓)**	**MnRE (↑)**	**Method**	**ADE (↓)**	**ALDE (↓)**	**APE (↓)**	**MnRE (↑)**
GPT-4V [64]	32.752	0.330	0.304	0.758	GPT-4V [64]	0.209	0.430	0.549	0.692
GPT-5 [52]	16.044	0.222	0.277	0.816	GPT-5 [52]	0.259	0.519	0.646	0.615
CLIP [47]	32.857	0.294	0.266	0.774	CLIP [47]	0.222	0.455	0.602	0.654
NeRF2Physics [66]	34.295	0.315	0.276	0.765	NeRF2Physics [66]	0.155	0.321	**0.360**	0.736
Ours	**12.721**	**0.193**	**0.222**	**0.839**	**Ours**	**0.131**	**0.263**	0.365	**0.805**

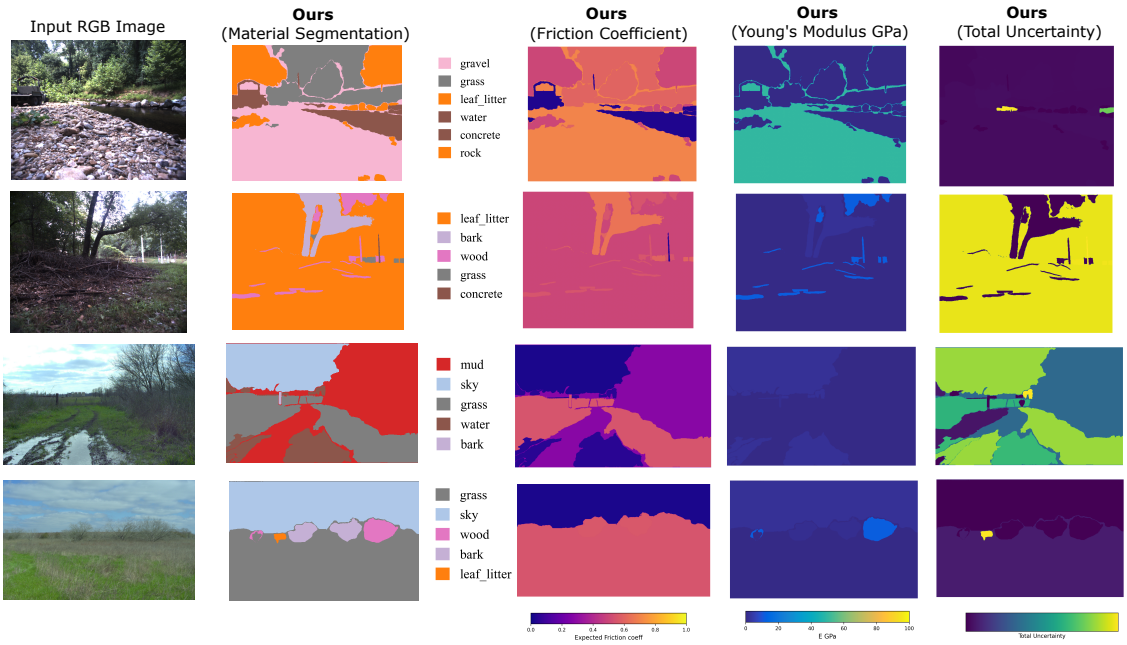


Figure 5. Outdoor scene results on real environments. From a single RGB view, *PhysGS* predicts material segmentation, friction coefficients, Young's modulus, and total uncertainty (*aleatoric* + *epistemic*). The method captures broad material variations across natural terrain and vegetation while producing pixel-wise physical property estimates with associated confidence. Higher total uncertainty in Rows 2 and 3 corresponds to scenes with dense clutter and visually ambiguous regions, where SAM provides less precise part-level masks (e.g., separating leaf litter from wood or mud from grass). Rows 1 and 4 exhibit lower uncertainty due to clearer material boundaries and more uniform regions.

Our method also provides per-pixel uncertainty (aleatoric + epistemic) for these estimates. We demonstrate this capability on the RUGD [61] and RELLIS-3D [23] datasets (Figure 5), which contain challenging outdoor scenes where accurate physical property prediction and uncertainty assessment are essential for interpreting complex terrain.

5. Conclusion, Limitations and Future Work

We presented *PhysGS*, a Bayesian-inferred 3D Gaussian Splatting framework for estimating dense physical properties from RGB images and vision-language priors. Across indoor and outdoor real-world datasets, *PhysGS* achieves consistent gains over existing approaches. On ABO-500, our method improves mass estimation by 5.5% in ADE and 22.8% in APE. *PhysGS* also reduces Shore hardness error by up to 61.2% and kinetic friction error by up to 18.1% relative to the strongest baselines. Outdoor experiments on RUGD and RELLIS-3D further show that the method generalizes to complex natural environments, capturing material segmentation, friction, stiffness, and uncertainty. A primary limitation of *PhysGS* is its sensitivity to segmentation quality. When part-level masks merge visually similar materials or fail to isolate fine-grained regions, the resulting physical property estimates inherit this ambiguity, reducing material separation and increasing predictive uncertainty. Future work may incorporate VLM-guided segmentation refinement to reject low-quality masks and preserve fine-grained material structure. Another limitation is the lack of calibrated VLM confidence estimates. Future work could improve reliability through improved calibration and using complementary sensing modalities such as tactile or force-based measurements.

References

[1] Aude Billard and Danica Kragic. Trends and challenges in robot manipulation. *Science*, 364(6446):eaat8414, 2019. 1

[2] Katherine L Bouman, Bei Xiao, Peter Battaglia, and William T Freeman. Estimating the material properties of fabric from video. In *Proceedings of the IEEE international conference on computer vision*, pages 1984–1991, 2013. 2

[3] Xiaoyi Cai, Siddharth Ancha, Lakshay Sharma, Philip R Osteen, Bernadette Bucher, Stephen Phillips, Jiuguang Wang, Michael Everett, Nicholas Roy, and Jonathan P How. Evora: Deep evidential traversability learning for risk-aware off-road autonomy. *IEEE Transactions on Robotics*, 2024. 2, 3

[4] Xiaoyi Cai, James Queeney, Tong Xu, Aniket Datar, Chenhui Pan, Max Miller, Ashton Flather, Philip R Osteen, Nicholas Roy, Xuesu Xiao, et al. Pietra: Physics-informed evidential learning for traversing out-of-distribution terrain. *IEEE Robotics and Automation Letters*, 2025. 3

[5] Jiaqi Chen, Jonas Frey, Ruyi Zhou, Takahiro Miki, Georg Martius, and Marco Hutter. Identifying terrain physical parameters from vision-towards physical-parameter-aware locomotion and navigation. *IEEE Robotics and Automation Letters*, 2024. 1, 2, 3, 7

[6] Samarth Chopra, Fernando Cladera, Varun Murali, and Vijay Kumar. Agrinerf: Neural radiance fields for agriculture in challenging lighting conditions. *arXiv preprint arXiv:2409.15487*, 2024. 2

[7] Jasmine Collins, Shubham Goel, Kenan Deng, Achleshwar Luthra, Leon Xu, Erhan Gundogdu, Xi Zhang, Tomas F Yago Vicente, Thomas Dideriksen, Himanshu Arora, et al. Abo: Dataset and benchmarks for real-world 3d object understanding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 21126–21136, 2022. 5

[8] Alberto Elfes. Using occupancy grids for mobile robot perception and navigation. *Computer*, 22(6):46–57, 2002. 2

[9] Gian Erni, Jonas Frey, Takahiro Miki, Matias Mattamala, and Marco Hutter. Mem: Multi-modal elevation mapping for robotics and learning. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 11011–11018. IEEE, 2023. 3

[10] Parker Ewen, Adam Li, Yuxin Chen, Steven Hong, and Ram Vasudevan. These maps are made for walking: Real-time terrain property estimation for mobile robots. *IEEE Robotics and Automation Letters*, 7(3):7083–7090, 2022. 1, 2, 3

[11] Parker Ewen, Hao Chen, Yuzhen Chen, Anran Li, Anup Bagali, Gitesh Gunjal, and Ram Vasudevan. You've got to feel it to believe it: Multi-modal bayesian inference for semantic and property prediction. *arXiv preprint arXiv:2402.05872*, 2024. 2, 3, 4

[12] David D Fan, Kyohei Otsu, Yuki Kubo, Anushri Dixit, Joel Burdick, and Ali-Akbar Agha-Mohammadi. Step: Stochastic traversability evaluation and planning for risk-aware off-road navigation. *arXiv preprint arXiv:2103.02828*, 2021. 3

[13] Jonas Frey, Manthan Patel, Deegan Atha, Julian Nubert, David Fan, Ali Agha, Curtis Padgett, Patrick Spieler, Marco Hutter, and Shehryar Khattak. Roadrunner-learning traversability estimation for autonomous off-road driving. *IEEE Transactions on Field Robotics*, 2024. 3

[14] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016. 3

[15] Lily Goli, Cody Reading, Silvia Sellán, Alec Jacobson, and Andrea Tagliasacchi. Bayes' rays: Uncertainty quantification for neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20061–20070, 2024. 3

[16] Jun Guo, Xiaojian Ma, Yue Fan, Huaping Liu, and Qing Li. Semantic gaussians: Open-vocabulary scene understanding with 3d gaussian splatting. *arXiv preprint arXiv:2403.15624*, 2024. 2

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 6

[18] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. *arXiv preprint arXiv:1903.12261*, 2019. 2

[19] Matthew D Hoffman, Tuan Anh Le, Pavel Sountsov, Christopher Suter, Ben Lee, Vikash K Mansinghka, and Rif A Saurous. Probnerf: Uncertainty-aware inference of 3d shapes from 2d images. In *International Conference on Artificial Intelligence and Statistics*, pages 10425–10444. PMLR, 2023. 3

[20] Jiarui Hu, Xianhao Chen, Boyin Feng, Guanglin Li, Liangjing Yang, Hujun Bao, Guofeng Zhang, and Zhaopeng Cui. Cg-slam: Efficient dense rgb-d slam in a consistent uncertainty-aware 3d gaussian field. In *European Conference on Computer Vision*, pages 93–112. Springer, 2024. 3

[21] Xu Hu, Yuxi Wang, Lue Fan, Junsong Fan, Junran Peng, Zhen Lei, Qing Li, and Zhaoxiang Zhang. Semantic anything in 3d gaussians. *CoRR*, 2024. 2

[22] Krishna Murthy Jatavallabhula, Alihusein Kuwajerwala, Qiao Gu, Mohd Omama, Tao Chen, Alaa Maalouf, Shuang Li, Ganesh Iyer, Soroush Saryazdi, Nikhil Keetha, et al. Conceptfusion: Open-set multimodal 3d mapping. *arXiv preprint arXiv:2302.07241*, 2023. 2

[23] Peng Jiang, Philip Osteen, Maggie Wigness, and Srikanth Saripalli. Rellis-3d dataset: Data, benchmarks and analysis. In *2021 IEEE international conference on robotics and automation (ICRA)*, pages 1110–1116. IEEE, 2021. 8

[24] Wen Jiang, Boshu Lei, and Kostas Daniilidis. Fisherrf: Active view selection and mapping with radiance fields using fisher information. In *European Conference on Computer Vision*, pages 422–440. Springer, 2024. 3

[25] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems*, 30, 2017. 2, 3

[26] Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7482–7491, 2018. 2, 3

[27] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 2

[28] Justin Kerr, Chung Min Kim, Ken Goldberg, Angjoo Kanazawa, and Matthew Tancik. Lerf: Language embedded radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 19729–19739, 2023. 2

[29] Junyoung Kim, Minsik Jeon, Jihong Min, Kiho Kwak, and Junwon Seo. E2-bki: Evidential ellipsoidal bayesian kernel inference for uncertainty-aware gaussian semantic mapping. *arXiv preprint arXiv:2509.11964*, 2025. 3

[30] Mijeong Kim, Jongwoo Lim, and Bohyung Han. 4d gaussian splatting in the wild with uncertainty-aware regularization. *Advances in Neural Information Processing Systems*, 37:129209–129226, 2024. 3

[31] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023. 5

[32] Yo Kobayashi, Jun Okamoto, and Masakatsu G Fujie. Physical properties of the liver and the development of an intelligent manipulator for needle insertion. In *Proceedings of the 2005 IEEE international conference on robotics and automation*, pages 1632–1639. IEEE, 2005. 1

[33] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017. 3

[34] Sibaek Lee, Kyeongsu Kang, Seongbo Ha, and Hyeonwoo Yu. Bayesian nerf: Quantifying uncertainty with volume density for neural implicit fields. *IEEE Robotics and Automation Letters*, 2025. 3

[35] Ruiqi Li and Yiu-ming Cheung. Variational multi-scale representation for estimating uncertainty in 3d gaussian splatting. *Advances in Neural Information Processing Systems*, 37:87934–87958, 2024. 3

[36] Jing Liang, Kasun Weerakoon, Tianrui Guan, Nare Karapetyan, and Dinesh Manocha. Adaptiveon: Adaptive outdoor local navigation method for stable and reliable actions. *IEEE Robotics and Automation Letters*, 8(2):648–655, 2022. 1, 2

[37] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *Advances in neural information processing systems*, 36:34892–34916, 2023. 6, 7

[38] Haozhe Lou, Mingtong Zhang, Haoran Geng, Hanyang Zhou, Sicheng He, Zhiyuan Gao, Siheng Zhao, Jiageng Mao, Pieter Abbeel, Jitendra Malik, et al. Dream: Differentiable real-to-sim-to-real engine for learning robotic manipulation. In *3rd RSS Workshop on Dexterous Manipulation: Learning and Control with Diverse Data*. 1

[39] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 2

[40] Helen Oleynikova, Alexander Millane, Zachary Taylor, Enric Galceran, Juan Nieto, and Roland Siegwart. Signed distance fields: A natural representation for both mapping and planning. In *RSS 2016 workshop: geometry and beyond-representations, physics, and scene understanding for robotics*. University of Michigan, 2016. 2

[41] Sadao Omata, Yoshinobu Murayama, and Christos E Constantinou. Real time robotic tactile sensor system for the determination of the physical properties of biomaterials. *Sensors and Actuators A: Physical*, 112(2-3):278–285, 2004. 1

[42] Changmin Park, Beomjoon Lee, Haechan Jung, Haejin Jung, and Changjoo Nam. Understanding physical properties of unseen deformable objects by leveraging large language models and robot actions. *arXiv preprint arXiv:2506.03760*, 2025. 1, 2

[43] Manthan Patel, Jonas Frey, Deegan Atha, Patrick Spieler, Marco Hutter, and Shehryar Khattak. Roadrunner m&m-learning multi-range multi-resolution traversability maps for autonomous off-road navigation. *IEEE Robotics and Automation Letters*, 2024. 3

[44] Philippe Pierre Pebay. Formulas for robust, one-pass parallel computation of covariances and arbitrary-order statistical moments. Technical report, Sandia National Laboratories (SNL), Albuquerque, NM, and Livermore, CA . . . , 2008. 4

[45] Songyou Peng, Kyle Genova, Chiyu Jiang, Andrea Tagliasacchi, Marc Pollefeys, Thomas Funkhouser, et al. Openscene: 3d scene understanding with open vocabularies. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 815–824, 2023. 2

[46] Minghan Qin, Wanhua Li, Jiawei Zhou, Haoqian Wang, and Hanspeter Pfister. Langsplat: 3d language gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20051–20060, 2024. 2

[47] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR, 2021. 2, 7, 8

[48] Erik Sandström, Ganlin Zhang, Keisuke Tateno, Michael Oechsle, Michael Niemeyer, Youmin Zhang, Manthan Patel, Luc Van Gool, Martin Oswald, and Federico Tombari. Splat-slam: Globally optimized rgb-only slam with 3d gaussians. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 1680–1691, 2025. 3

[49] Luca Savant, Diego Valsesia, and Enrico Magli. Modeling uncertainty for gaussian splatting. *arXiv preprint arXiv:2403.18476*, 2024. 3

[50] Gershom Seneviratne, Kasun Weerakoon, Mohamed Elnoor, Vignesh Rajgopal, Harshavarthan Varatharajan, Mohamed Khalid M Jaffar, Jason Pusey, and Dinesh Manocha. Crossgait: Cross-attention-based multimodal representation fusion for parametric gait adaptation in complex terrains. *arXiv preprint arXiv:2409.17262*, 2024. 1

[51] Yinghao Shuai, Ran Yu, Yuantao Chen, Zijian Jiang, Xiaowei Song, Nan Wang, Jv Zheng, Jianzhu Ma, Meng

Yang, Zhicheng Wang, et al. Pugs: Zero-shot physical understanding with gaussian splatting. *arXiv preprint arXiv:2502.12231*, 2025. 2

[52] Aaditya Singh, Adam Fry, Adam Perelman, Adam Tart, Adi Ganesh, Ahmed El-Kishky, Aidan McLaughlin, Aiden Low, AJ Ostrow, Akhila Ananthram, et al. Openai gpt-5 system card. *arXiv preprint arXiv:2601.03267*, 2025. 5, 7, 8

[53] Trevor Standley, Ozan Sener, Dawn Chen, and Silvio Savarese. image2mass: Estimating the mass of an object from its image. In *Conference on Robot Learning*, pages 324–333. PMLR, 2017. 6, 7

[54] Matthew Tancik, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, et al. Nerfstudio: A modular framework for neural radiance field development. In *ACM SIGGRAPH 2023 conference proceedings*, pages 1–12, 2023. 5

[55] Sebastian Thrun. Probabilistic algorithms in robotics. *Ai Magazine*, 21(4):93–93, 2000. 2

[56] Sebastian Thrun. Probabilistic robotics. *Communications of the ACM*, 45(3):52–57, 2002.

[57] Jonathan Tremblay, Aayush Prakash, David Acuna, Mark Brophy, Varun Jampani, Cem Anil, Thang To, Eric Cameracci, Shaad Boochoon, and Stan Birchfield. Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 969–977, 2018. 2

[58] Yi Wang, Jiafei Duan, Dieter Fox, and Siddhartha Srinivasa. Newton: Are large language models capable of physical reasoning? In *Findings of the association for computational linguistics: EMNLP 2023*, pages 9743–9758, 2023. 2

[59] Kasun Weerakoon, Adarsh Jagan Sathyamoorthy, Jing Liang, Tianrui Guan, Utsav Patel, and Dinesh Manocha. Graspe: Graph based multimodal fusion for robot navigation in outdoor environments. *IEEE Robotics and Automation Letters*, 8(12):8090–8097, 2023. 1, 2

[60] DHD West. Updating mean and variance estimates: An improved method. *Communications of the ACM*, 22(9):532–535, 1979. 4

[61] Maggie Wigness, Sungmin Eum, John G Rogers, David Han, and Heesung Kwon. A rugd dataset for autonomous navigation and visual perception in unstructured outdoor environments. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5000–5007. IEEE, 2019. 8

[62] William Xie, Maria Valentini, Jensen Lavering, and Nikolaus Correll. Deligrasp: Inferring object properties with llms for adaptive grasp policies. *arXiv preprint arXiv:2403.07832*, 2024. 2

[63] Xinli Xu, Wenhang Ge, Dicong Qiu, ZhiFei Chen, Dongyu Yan, Zhuoyun Liu, Haoyu Zhao, Hanfeng Zhao, Shunsi Zhang, Junwei Liang, et al. Gaussianproperty: Integrating physical properties to 3d gaussians with lmms. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7231–7240, 2025. 1, 2, 5, 4

[64] Zhengyuan Yang, Linjie Li, Kevin Lin, Jianfeng Wang, Chung-Ching Lin, Zicheng Liu, and Lijuan Wang. The dawn of lmms: Preliminary explorations with gpt-4v (ision). *arXiv preprint arXiv:2309.17421*, 2023. 7, 8

[65] Aiko Yoshizawa, Jun Okamoto, Hiroshi Yamakawa, and Masakatsu G Fujie. Robot surgery based on the physical properties of the brain-physical brain model for planning and navigation of a surgical robot. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pages 904–911. IEEE, 2005. 1

[66] Albert J Zhai, Yuan Shen, Emily Y Chen, Gloria X Wang, Xinlei Wang, Sheng Wang, Kaiyu Guan, and Shenlong Wang. Physical property understanding from language-embedded feature fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 28296–28305, 2024. 1, 2, 5, 6, 7, 8

[67] Yizhen Zheng, Huan Yee Koh, Jiaxin Ju, Anh TN Nguyen, Lauren T May, Geoffrey I Webb, and Shirui Pan. Large language models for scientific discovery in molecular property prediction. *Nature Machine Intelligence*, pages 1–11, 2025. 2