Physical Property Understanding from Language-Embedded Feature Fields

Supplementary Material

Abstract

In the following supplementary material, we provide additional experimental details (Sec. A) and additional results (Sec. B). We invite readers to watch the supplementary video (`supp.mp4`) for further visualization.

A. Additional Experimental Details

A.1. Point Cloud Extraction Details

We use the default point cloud exporter in Nerfstudio [?], which allows one to define a bounding box for the points to keep. For ABO-500, we use a bounding box of size $1 \times 1 \times 1$ centered at $(0, 0, 0)$. For our in-house dataset, we use a bounding box of size $1.5 \times 1.5 \times 1.5$ centered at $(0, 0, -0.75)$. Points are filtered out if the average distance to their 20 nearest neighbors is over 10 standard deviations away from the mean.

A.2. Prompting Details

We provide the prompts used for captioning (Fig. 1) and material proposal with mass density (Fig. 2), friction coefficient (Fig. 3), and hardness (Fig. 4) values. The prompt used for thickness estimation is provided in Fig. 5. We also provide the prompts used for LLaVA mass estimation (Fig. 6), GPT-4V friction estimation (Fig. 7), and GPT-4V hardness estimation (Fig. 8).

> Question: Give a detailed description of the object. Answer:

Figure 1. Prompt used for captioning with BLIP-2.

A.3. CNN Baseline Details

Our CNN baseline takes one RGB image and outputs the predicted mass. We use an ImageNet-pretrained ResNet50 backbone with the last classification layer removed. A set of fully connected layers with ReLU activations follows the ResNet50. Within the fully connected layers, feature dimensions are as follows: 2048, 32, 16, and 1. We attach a LogSigmoid layer to ensure the output is non-negative. Our input image is normalized based on ImageNet statistics before feeding into our model. We train for 20 epochs using an Adam optimizer with a learning rate of $\alpha = 0.001$.

> **System:** You will be provided with captions that each describe an image of an object. The captions will be delimited with quotes ("). Based on the caption, give me 5 materials that the object might be made of, along with the mass densities (in kg/m^3) of each of those materials. You may provide a range of values for the mass density instead of a single value. Try to consider all the possible parts of the object. Do not include coatings like "paint" in your answer.
>
> Format Requirement:
> You must provide your answer as a list of 5 (material: mass density) pairs, each separated by a semi-colon (;). Do not include any other text in your answer, as it will be parsed by a code script later. Your answer must look like:
> (material 1: low-high kg/m^3);(material 2: low-high kg/m^3);(material 3: low-high kg/m^3);(material 4: low-high kg/m^3);(material 5: low-high kg/m^3)

Figure 2. Prompt used for proposing materials and providing their mass density values.

> **System:** You will be provided with captions that each describe an image. The captions will be delimited with quotes ("). Based on the caption, give me 3 materials that the surfaces in the image might be made of, along with the kinetic friction coefficient of each material when sliding against a fabric surface. You may provide a range of values for the friction coefficient instead of a single value. Try to consider all the possible surfaces.
>
> Format Requirement:
> You must provide your answer as a list of 3 (material: friction coefficient) pairs, each separated by a semi-colon (;). Do not include any other text in your answer, as it will be parsed by a code script later. Your answer must look like:
> (material 1: low-high);(material 2: low-high);(material 3: low-high)
> Try to provide as narrow of a range as possible for the friction coefficient.

Figure 3. Prompt used for proposing materials and providing their friction coefficients.

B. Additional Results

B.1. Young's Modulus and Thermal Conductivity

Our method can be used to predict physical properties in an open-vocabulary manner. We show that it can be used to predict Young's modulus and thermal conductivity on objects from ABO-500 (val) in Fig. 9. For visualization purposes, we force the candidate materials to be the same as those proposed for mass density estimation. The exact prompts used for these results can be found in Fig. 10 (Young's modulus) and Fig. 11 (thermal conductivity). We use GPT-3.5 Turbo

System: You will be provided with captions that each describe an image of an object. The captions will be delimited with quotes ("). Based on the caption, give me 3 materials that the object might be made of, along with the hardness of each of those materials. Choose whether to use Shore A hardness or Shore D hardness depending on the material. You may provide a range of values for hardness instead of a single value. Try to consider all the possible parts of the object.

Format Requirement:
You must provide your answer as a list of 3 (material: hardness, Shore A/D) tuples, each separated by a semi-colon (;). Do not include any other text in your answer, as it will be parsed by a code script later. Your answer must look like:
(material 1: low-high, <Shore A or Shore D>);(material 2: low-high, <Shore A or Shore D>);(material 3: low-high, <Shore A or Shore D>)
Make sure to use Shore A or Shore D hardness, not Mohs hardness.

Figure 4. Prompt used for proposing materials and providing their Shore hardness values.

and a kernel regression temperature of $T = 0.01$.

B.2. Failure Cases

We show example failure cases of NeRF2Physics in Fig. 12, covering the two main failure modes. In the case of the wooden box, BLIP-2 mistakes the object for a brick, causing GPT to output erroneous materials such as sand and concrete. With no correct materials in the dictionary, the CLIP-based regression is unable to produce accurate predictions. One direction for future work could be to implement a more robust view selection strategy to avoid such recognition failures.

In the case of the black cart, the caption and materials are correct, but the CLIP-based regression mistakes the bulk of the cart as steel instead of plastic. This occurs because the local appearances of black-painted steel and black-painted plastic can look identical, and the patch-based CLIP features do not contain enough global information to accurately distinguish between them.

System: You will be provided with captions that each describe an image of an object, along with a set of possible materials used to make the object. For each material, estimate the thickness (in cm) of that material in the object. You may provide a range of values for the thickness instead of a single value.

Format Requirement:
You must provide your answer as a list of 5 (material: thickness) pairs, each separated by a semi-colon (;). Do not include any other text in your answer, as it will be parsed by a code script later. Your answer must look like:
(material 1: low-high cm);(material 2: low-high cm);(material 3: low-high cm);(material 4: low-high cm);(material 5: low-high cm)

User: Caption: "a lamp with a white shade" Materials: "fabric, plastic, metal, ceramic, glass"
Assistant: (fabric: 0.1-0.2 cm);(plastic: 0.3-1.0 cm);(metal: 0.1-0.2 cm);(ceramic: 0.2-0.5 cm);(glass: 0.3-0.8 cm)
User: Caption: "a grey ottoman" Materials: "wood, fabric, foam, metal, plastic"
Assistant: (wood: 2.0-4.0 cm);(fabric: 0.2-0.5 cm);(foam: 5.0-15.0 cm);(metal: 0.1-0.2 cm);(plastic: 0.5-1.0 cm)
User: Caption: "a white frame" Materials: "plastic, wood, aluminum, steel, glass"
Assistant: (plastic: 0.1-0.3 cm);(wood: 1.0-1.5 cm);(aluminum: 0.1-0.3 cm);(steel: 0.1-0.2 cm);(glass: 0.2-0.5 cm)
User: Caption: "a metal rack with three shelves" Materials: "steel, aluminum, wood, plastic, iron"
Assistant: (steel: 0.1-0.2 cm);(aluminum: 0.1-0.3 cm);(wood: 1.0-2.0 cm);(plastic: 0.5-1.0 cm);(iron: 0.5-1.0 cm)

Figure 5. Prompt used for estimating the thickness of proposed materials. Since this is a somewhat confusing task, we provide a few in-context examples to help the LLM understand what we mean by thickness.

System: Estimate the mass of the object in kilograms. Provide your answer as only a decimal number.

Figure 6. Prompt used for estimating mass with LLaVA.

System: You will be given an image, followed by a mask specifying a point on the image. Estimate the coefficient of kinetic friction between a fabric surface and the surface at that point in the image.

Format Requirement:
You must provide either a single number or a range (e.g. "0.6-0.8") as your answer. Give your best guess. Do not include any other text in your answer, as it will be parsed by a code script later.

Figure 7. Prompt used for estimating friction with GPT-4V.

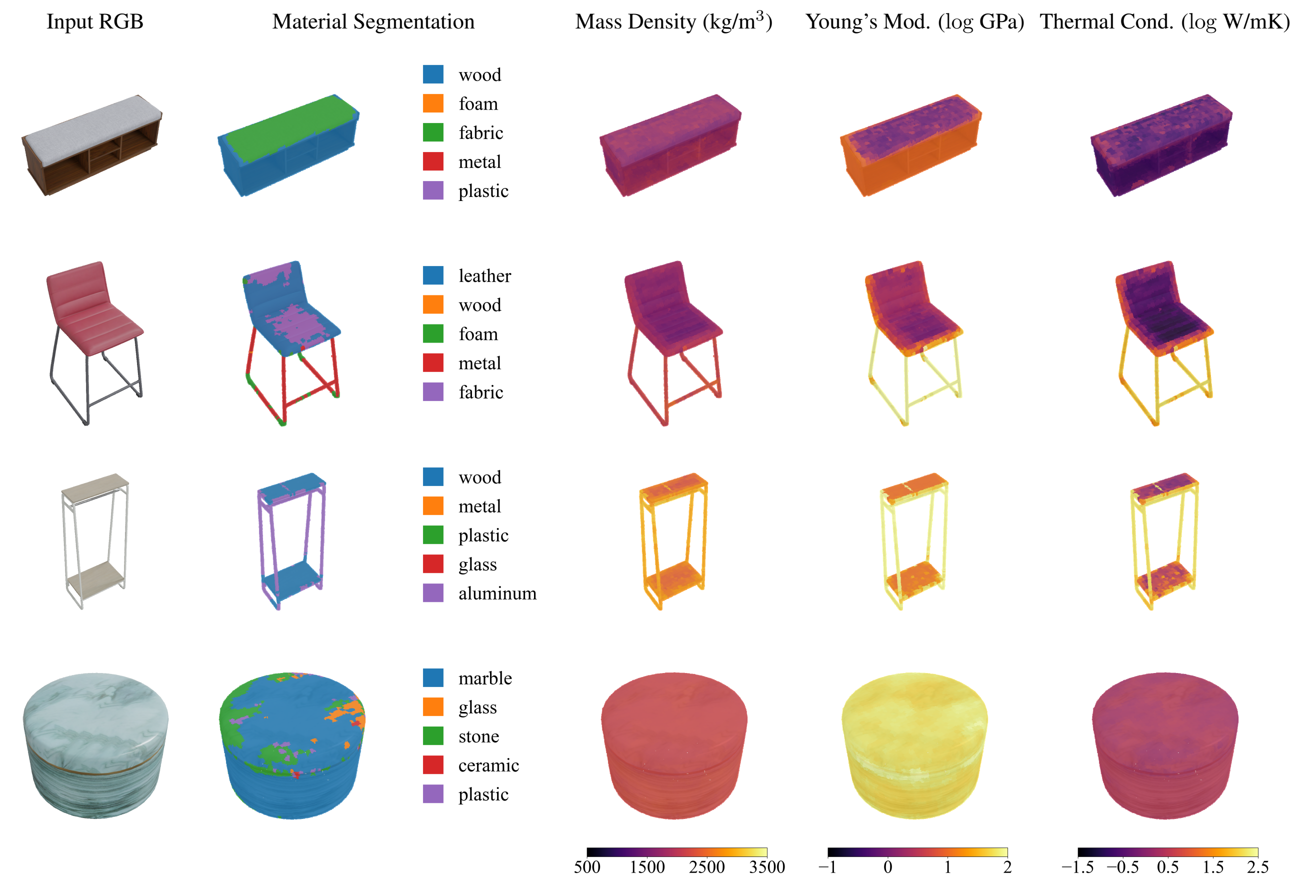

Figure 9. **Example predictions of different physical properties.** We visualize more zero-shot predictions of mass density of objects from ABO-500, along with predictions of Young's modulus fields and thermal conductivity fields. Our method produces accurate predictions across a wide variety of objects and materials.

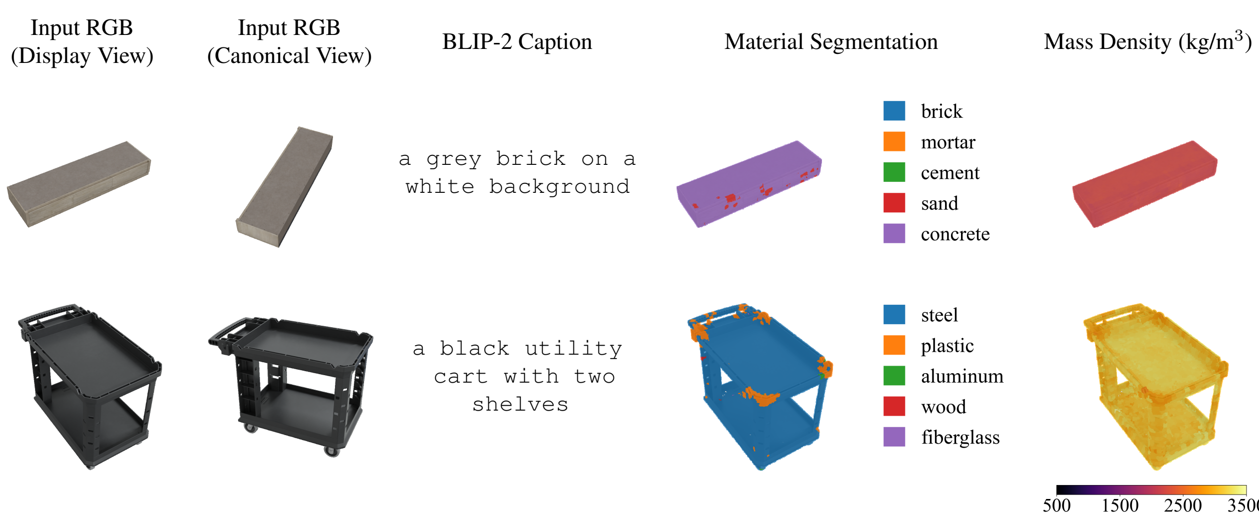

Figure 12. **Example failure cases.** We visualize the main failure modes of NeRF2Physics in the above examples. The first example demonstrates object recognition failure at the captioning stage, and the second example demonstrates material recognition failure at the CLIP-based retrieval stage.

System: You will be given an image, followed by a mask specifying a point on the image. Estimate the hardness of the object in the image at the given point. Choose whether to use Shore A hardness or Shore D hardness depending on the material.

Format Requirement:
You must provide a pair of either a single number or a range (e.g. "0.6-0.8") and whether it is in Shore A or Shore D as your answer. Give your best estimate, and make sure to use Shore A or Shore D hardness, not Mohs hardness. Do not include any other text in your answer. Your answer must look like:
(number, <Shore A or Shore D>)

Figure 8. Prompt used for estimating hardness with GPT-4V.

System: You will be provided with captions that each describe an image of an object, along with a set of possible materials used to make the object. For each material, estimate the Young's modulus (in GPa) of that material in the object. You may provide a range of values for the Young's modulus instead of a single value.

Format Requirement:
You must provide your answer as a list of 5 (material: Young's modulus) pairs, each separated by a semi-colon (;). Do not include any other text in your answer, as it will be parsed by a code script later. Your answer must look like:
(material 1: low-high GPa);(material 2: low-high GPa);(material 3: low-high GPa);(material 4: low-high GPa);(material 5: low-high GPa)

Figure 10. Prompt used for estimating Young's modulus.

System: You will be provided with captions that each describe an image of an object, along with a set of possible materials used to make the object. For each material, estimate the thermal conductivity (in W/mK) of that material in the object. You may provide a range of values for the thermal conductivity instead of a single value.

Format Requirement:
You must provide your answer as a list of 5 (material: thermal conductivity) pairs, each separated by a semi-colon (;). Do not include any other text in your answer, as it will be parsed by a code script later. Your answer must look like:
(material 1: low-high W/mK);(material 2: low-high W/mK);(material 3: low-high W/mK);(material 4: low-high W/mK);(material 5: low-high W/mK)

Figure 11. Prompt used for estimating thermal conductivity.