Semantically-Enriched 3D Models for Common-sense Knowledge

Manolis Savva, Angel X.Chang, Pat Hanrahan
Computer Science Department, Stanford University
{msavva,angelx,hanrahan}@cs.stanford.edu

Abstract

We identify and connect a set of physical properties to 3D models to create a richly-annotated 3D model dataset with data on physical sizes, static support, attachment surfaces, material compositions, and weights. To collect these physical property priors, we leverage observations of 3D models within 3D scenes and information from images and text. By augmenting 3D models with these properties we create a semantically rich, multi-layered dataset of common indoor objects. We demonstrate the usefulness of these annotations for improving 3D scene synthesis systems, enabling faceted semantic queries into 3D model datasets, and reasoning about how objects can be manipulated by people using weight and static friction estimates.

1. Introduction

Despite much recent progress in 3D scene understanding, many simple questions about the structure of the visual world are hard to answer computationally: What is in a kitchen? Where on a couch can an iPad be placed? Can a person lift a refrigerator? How about a microwave? Answers to these questions are predicated on fundamental physical properties of the objects, their functionality within real-world environments, and common sense knowledge that connects the two.

At the same time, 3D content is becoming increasingly available. Online 3D model repositories continue to grow on a daily basis and a revolution in scanning methods is creating increasingly faithful 3D reconstructions of real environments. Yet, despite the geometric fidelity of 3D model representations, the semantics of real objects are unavailable. This makes it very hard to answer common sense questions and use the models in practical applications such as 3D scene synthesis, and object recognition in computer vision systems.

To address this lack of semantic information for 3D models we extract physical object properties from observations of the 3D models in a database of 3D scenes. The statistics of high-level structure in 3D scenes are easier to capture than in image space where many open vision problems have to be addressed: detection, segmentation, 3D layout estimation among others.

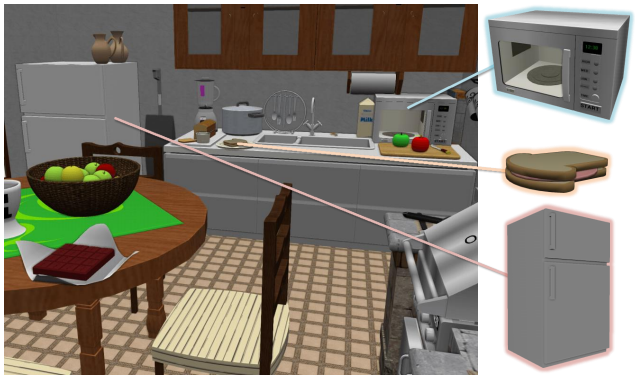

Figure 1. 3D scene of a kitchen containing 3D models of several common indoor objects. We use observations of objects in a corpus of 3D scenes and other information sources to create a semantically-enriched dataset of 3D models with properties such as physical sizes, natural orientations, and static support priors (e.g., sandwiches are placed on plates).

We focus on defining a set of fundamental properties of objects in the context of indoor 3D scenes, present simple approaches to extract and aggregate these properties, and finally demonstrate how these properties are useful in answering many common sense questions. In the process, we augment a dataset of 3D models with physical property annotations and provide it to the research community.[1]

Contributions We present how to connect several important physical properties of objects to 3D model representations by leveraging observations within 3D scenes. We augment a corpus of 3D models with several physical properties to create a semantically rich, multi-layered dataset of common indoor objects. We demonstrate the usefulness of these annotations for improving 3D scene synthesis systems, enabling faceted semantic queries into 3D model datasets, and reasoning about how objects can be manipulated by people using weight and static friction estimates.

[1] http://graphics.stanford.edu/projects/semgeo/

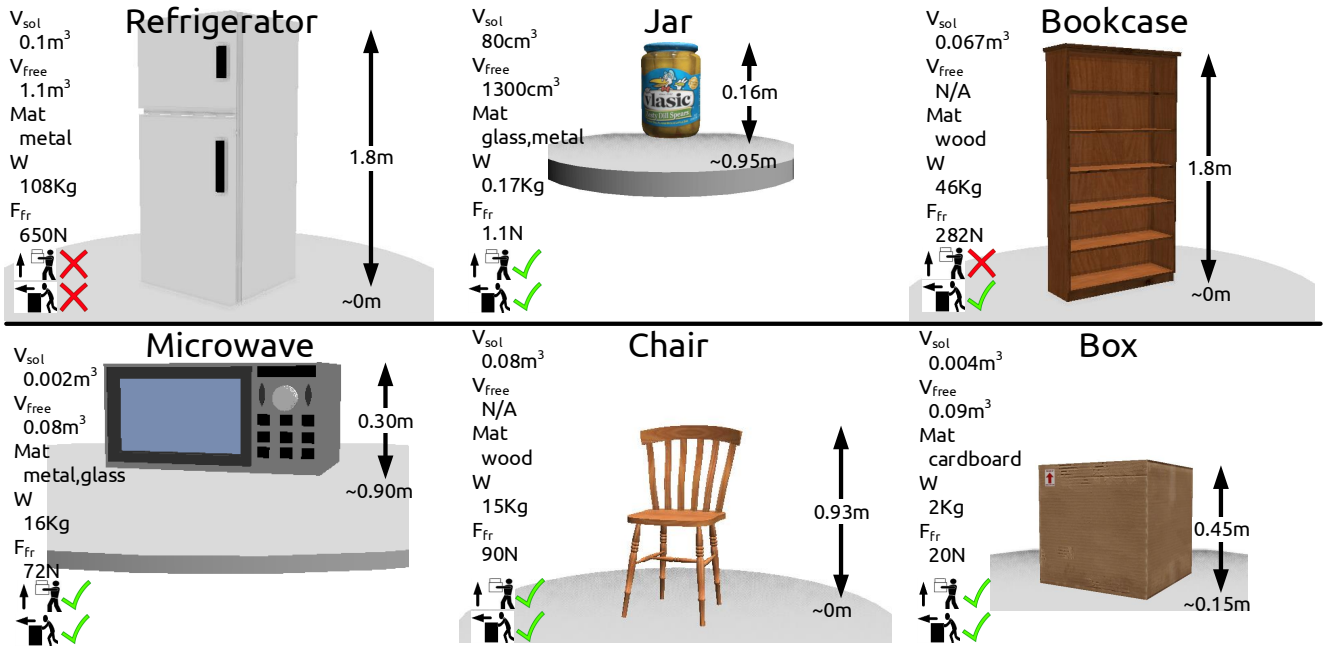

Figure 2. By jointly using estimates of the dimensions, static support surface height, material composition, and solidity of objects we estimate occupied and free container volume (V_{sol} and V_{free}), weight W, and static support friction F_{fr} forces of objects. This allows us to make predictions such as whether it would be easy for people to carry or push each object instance (indicated by symbols at bottom).

2. Related Work

Recently, there has been much interest in leveraging object affordance information for a variety of tasks such as object detection and recognition through associated human poses. One line of work uses hallucinated human poses to label objects in RGB-D data [11] and to plan placement of objects in novel scenes [12]. Another recent effort constructs a knowledge base of affordances for objects and demonstrates how it can be used for visual reasoning [31]. In recent robotics work, prediction of graspable and container parts of 3D models is used for planning robot grasping [23]. We similarly focus on augmenting a dataset of 3D models with properties that correlate with functionality. However, we leverage the context provided by observations of models within 3D scenes to collect static support and attachment priors.

Another line of work has focused on reasoning about the stability of volumetrically reconstructed 3D scenes [30] for scene understanding of RGB-D input data. We similarly reason about static support within 3D scenes but we focus on extracting support and attachment surface priors and using them to predict these surfaces on 3D models. More recent work has extracted the statistics of static support relations to enable novel interactive scene design interfaces (Clutterpalette [27]). We take a similar approach in extracting support priors, however we use 3D scenes instead of annotated images as input, allowing us to reason at a finer granularity about the geometry of the support surfaces of objects.

Much prior work in computer graphics has focused on low level geometric analysis tasks and has presented several 3D model datasets to be used as benchmarks. The most popular example is the *Princeton Shape Benchmark* [21]. However, such datasets typically only include object category labels for the 3D models. In computer vision, recent work has shown the benefit of a 3D model corpus for joint object detection and shape reconstruction from RGB-D data [22, 25]. The latter collected a large dataset of more than 120 thousand 3D models and manually verified the categories and orientations of a 10 category subset with 4899 3D models. Inspired by the demonstrated success of data-driven methods using 3D models for vision tasks, we create a 3D model corpus with rich physical property annotations containing 12490 models over 270 categories.

Most recently, the vision community is focusing on defining a Visual Turing test for deep understanding of visual structure and semantics in order to perform complex queries over image datasets for question answering and image retrieval tasks [9, 17]. We believe that richly annotated 3D representations of the world will become critical for making progress in these tasks. Recent work in scene understanding compellingly demonstrates the value of physically-grounded common sense knowledge [5, 6, 28, 29]. Our contribution of an approach to richly annotate 3D models with physical properties is a step towards providing a useful dataset for these opening research directions.

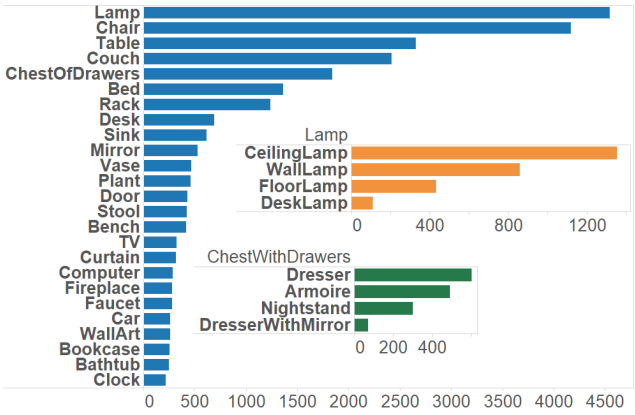

Figure 3. Distribution of 3D models in our corpus over categories at different taxonomy levels (inner distributions are over lamp and drawer furniture categories respectively). Our dataset is based on a 3D scene synthesis dataset from prior work [7] and consists of 12000 object models in total over about 200 basic categories.

3. Semantic Annotations for 3D Models

We aim to collect information that is useful for answering common sense questions about the visual structure of indoor environments. Some examples of such questions include:

- What objects are in a kitchen?
- Where can I look for apples in a living room?
- How big are apples?
- Which way does the TV face? Which way does the fridge open?
- How heavy is the fridge? Can you lift it? Push it?
- Can you put things inside a fridge? Inside a jar?
- Where do I look for a desk lamp? Which side of the desk lamp supports it? What about a wall lamp? A ceiling lamp?

Though the first few questions are possible to answer with high-level categorical knowledge and image co-occurrence statistics, answers to the latter questions rely on a knowledge of fundamental physical properties of objects: physical sizes, natural orientations, material composition, and object solidity. Our key insight is that many aspects of these properties are reflected strongly in the contextualized observations of 3D models within 3D scenes composed of object models (e.g., a living room with tables, chairs, couch, TV, etc.) We use a 3D scene dataset of common interior environments such as kitchens, living rooms and bedrooms from prior work on scene synthesis [7]. We annotate the 3D models used in these scenes with basic categories in a simple taxonomy (see Figure 3). Object models within scenes are a rich source of data which we can combine with other modalities to extract several important physical properties of the objects:

Absolute sizes. An attribute of real objects which is unfortunately frequently inconsistent in public repositories of 3D models is the absolute size of the objects. Absolute size is critical in the real world since it influences the usability of objects and even their identity (e.g., a model airplane vs a real airplane). The human cognitive system is also strongly geared towards recognizing and organizing objects by size [14]. We use a 3D model size estimation method designed to propagate physical size priors between models in 3D scenes [20] to obtain physical sizes for our corpus.

Natural Orientations. Objects are observed in the real world in typical configurations which reflect their context and the actions that they admit for people. Most artificial objects have a clear upright orientation dictated by the functions they admit to people (e.g., chairs for sitting). Similarly, objects such as monitors, clocks and whiteboards have a front side which is associated with the activities people perform with them. We annotate the natural upwards and front orientations for our object categories so that they can be used in reasoning about relative orientations in 3D scenes.

Static Support Priors. The most prevalent force which dictates the structure of our world is gravity. The impact of gravity can be felt continuously by people and influences the structure of objects in the world. We collect a set of priors over the types of surfaces in different objects that support other objects being placed on them, and correspondingly, typical attachment and support surfaces on an object for placing the object in static equilibrium on other objects. These priors are collected from observations of 3D models within the context of 3D scenes.

Materiality. Real objects are composed of materials with properties that influence the appearance, density and texture of parts of the object and consequently their functionality (e.g., many chair seats are made from fabrics that are soft and comfortable to sit on). Such physical properties have a big impact on the semantics of objects but are frequently absent from 3D model representations. We establish priors on the materials that different objects are composed from by aggregating material annotations in 2D images [2] and corresponding them to 3D model object categories.

Solidity. Physical objects occupy 3D space and have solid volume—an aspect which is only implicit in surface representations such as triangle meshes. Combined with the materiality of objects, a distinction between the solid regions that a 3D model represents and any empty space it contains is important for determining weight, potential for containment, and simulating physics. Since common geometric

representations are surface-based, extracting solidity priors directly with geometric analysis is challenging. Our insight is that solidity is reflected strongly by language describing objects (e.g., the bowl is *in* the microwave). We estimate the empty volume within 3D models by using priors extracted from linguistic information that implies container-like objects.

In the following section we discuss our approach for extracting these properties and connecting them to the 3D model representations.

4. Constructing a Semantically-Enriched 3D Model Dataset

Our general approach is to use simple algorithmic approaches that attempt to connect informative priors on each of the physical properties we presented. As part of a larger pipeline we plan to augment these algorithmic predictions through manual annotation and verification by people using crowdsourcing.

4.1. Categorization

We define a manual taxonomy of categories for our dataset of 3D models. Since we focus on indoor scene design, our taxonomy mainly consists of furniture, common household objects, and electronics. Using a taxonomy is important, as it allows for generalization from fine-to-coarse grained categories (see Figure 3). We break up basic categories into subcategories mainly by geometric variation and functionality. For example, the *lamp* basic category is subcategorized into *table lamp*, *desk lamp*, *floor lamp*, *wall lamp*, and *ceiling lamp*. The key distinction is the typical location and the type of static support surface for the lamp. For the contrast between table and desk lamps the difference is between radially symmetric and focused spotlights for desk tasks.

4.2. Absolute Sizes

Another critical attribute of objects is their physical size. Unfortunately, most commonly available 3D model formats have incorrect or missing physical scale information. Prior work has looked at propagating priors on 3D model physical sizes through observations of the models in scenes, and predicting the size for new model instances [20]. We use this approach on all models observed within our 3D scene corpus to establish category-level size priors and then propagate these priors to all models within each category.

4.3. Natural Orientations

Consistent alignments within each category of objects are extremely useful in a variety of applications. There has been some prior work in predicting the upright orientations of 3D models [8]. However, since most models retrieved from web repositories already have a consistent upright orientation, we just manually verify each model. During this verification, we also specify a *front* side, in addition to the *upright* orientation, to provide a ground truth *natural orientation* for each object. Though most object categories have a common upright orientation, some categories may not have a well-defined front side (e.g., bowls, round tables). In these cases, the front side is assumed to be given by the original orientation in which the 3D model was designed. The presence of rotational symmetries can indicate such cases, so an interesting avenue for future work is to use geometric analysis to predict whether a semantic front exists for a given model and if it does, identify it.

Figure 4. Some examples of consistently oriented categories of models: chairs, monitors, desk lamps, and cars.

The specification of both up and front directions establishes a common reference frame for all 3D models (see Figure 4). This common reference frame is valuable for performing pose alignment of 3D models to images [1] and for synthesizing 3D scenes with naturally oriented objects [4].

4.4. Static Support

The surfaces on which objects are statically supported determine many other object attributes, and critically the likely placements of objects within scenes. In order to establish a set of simple Bayesian priors for static support surfaces and object attachment points, we first segment our 3D models using the SuperFace algorithm [13] to obtain a set of mostly planar surfaces. Given a 3D scene dataset we now extract priors on the support surface attributes and object attachment surfaces/points by observing how the surfaces of each model instance support other model instances in each scene.

We use a scene dataset from prior work on 3D scene synthesis [7], containing about 130 indoor scenes. This dataset includes a support tree hierarchy for each scene from

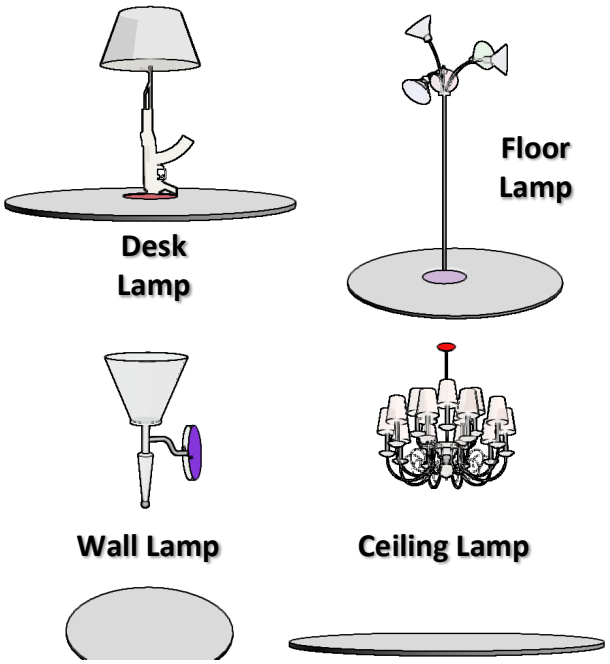

Figure 5. Predictions of the highest likelihood attachment surfaces for several types of lamp fixtures shown as colored surface regions on the 3D models.

which we extract child-parent pairs of statically supported and supporting objects. For each such pair, we identify the surfaces on the supporting parent by a proximity threshold to the midpoint of each bounding box face around the child object. Given an identified pair of parent support surface and child bounding box plane, we also retrieve the attachment surfaces of the child object that are within a small threshold ($1\,\mathrm{cm}$) of the attachment plane.

We aggregate the above detected support pairs onto the parent and child object categories to establish a set of priors on the supporting surfaces and attachment surfaces:

$$P_{surf}(s|C_c) = \frac{\text{count}(C_c \text{ on surface with } s)}{\text{count}(C_c)}$$

where C_p and C_c refer to the parent and child object categories, and s is a surface descriptor. We then use these priors to evaluate the likelihood of support and most likely supporting surface attributes in new instances of objects in unlabeled scenes through a simple featurization of the supporting and attachment surfaces s.

We first featurize the supported object attachment surfaces by bounding box side: top, bottom, front, back, left, or right. For instance, posters are attached on their back side to walls, rugs are attached on their bottom side to floors. Then, we featurize the parent supporting surface depending on the direction of the surface normal (up, down, horizontally) and whether the surface is interior (facing into the bounding box of the supporting object) or exterior (facing out of the bounding box). For instance, a room has a floor which is an upwards interior supporting surface, roof (upwards exterior), ceiling (downwards interior), and inside walls (horizontally interior). Given this featurization, we now learn from observations in scenes the distribution of supporting surface and attachment surface type for each category of object. With these learned Bayesian priors, we can now predict the static attachment probability for a model's surface (Figure 5), and the support probability for each surface of a candidate parent object within a 3D scene (Figure 6).

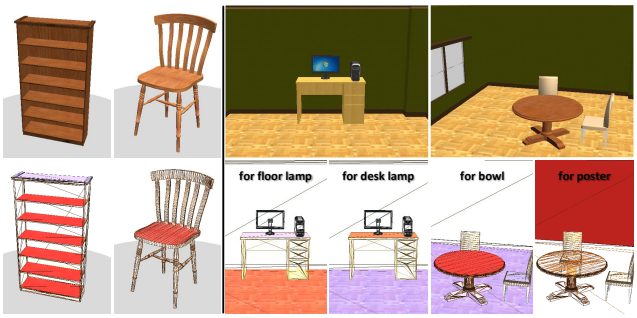

Figure 6. Left: predicted high likelihood support surfaces on a bookcase model and a chair model (red indicates surface with high probability of statically supporting other objects, magenta is low probability). Right: Likelihoods of static support for some object categories on surfaces in two different rooms.

To handle data sparsity we utilize our category taxonomy. If there are fewer than $k = 5$ support observations for a given category, we back off to a parent category in the taxonomy for more informative priors. If there are no observations available we use the geometry of the model instance to make a decision as follows. For attachment surfaces, if the object has roughly equal dimensions in 3D we assume the attachment surface is the bottom. If the object is flat, we assume either the back or bottom are attachment surfaces, choosing the one which is anti-parallel to the upright orientation (e.g., iPad bottom side). If the object is thin and long we choose a side along the long axis (e.g., side of a pen).

4.5. Materiality

We obtain an estimate of the material distribution for each object category, by counting how frequently a given material is annotated on instances of that category within the OpenSurfaces dataset [2]. We note that this is a naive approach which does not take into account that specific object instances may exhibit significant variation in the material composition (e.g. some mugs are entirely ceramic whereas others are entirely metallic). Instead, we only aggregate a distribution at the category level. Despite this, the data offers a useful first order estimate to establish common sense priors. See Figure 7 for the computed material distributions

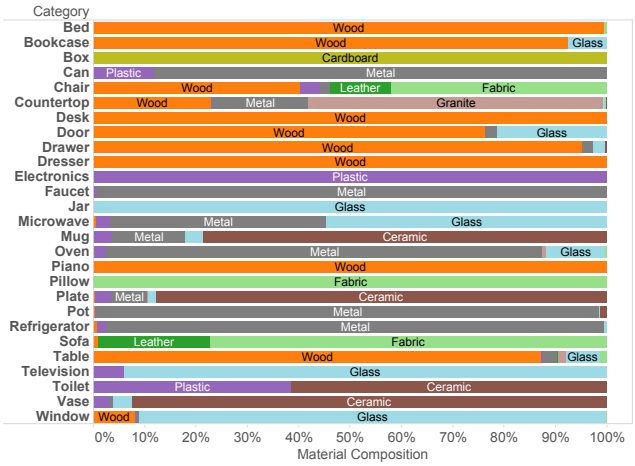

Figure 7. Material composition priors for some common categories of objects extracted from OpenSurfaces dataset.

solid	bed, bookcase, chair, mug, plate, table
container	box, can, jar, microwave, oven, refrigerator, vase, window

Table 1. Solidity predictions for common object categories extracted by comparing probabilities of references "in X" and "on X" from large-scale language models trained on web text [3]. Note that "window" is an interesting failure case due to the common expression "in the window" which does not imply containment.

over some common categories of objects.

In order to leverage these material composition distributions for computing object weights, we also collect overall material density values from the NIST STAR material composition database.[2] We assume that the metal in indoor appliances is mostly aluminum, and that wood is oak wood with average density.[3]

4.6. Solidity

Is an object internally mostly solid, or is it container-like and designed with free space for containing other objects? To determine whether 3D models represent solid objects or mostly empty container-like objects we look at linguistic cues indicating objects that are typically used as containers. To get these linguistic cues we use recently developed language models [10] that were learned from billions of webpages [3]. This pre-learned language model gives the probability of a sequence of words occurring together.

We establish the probability of the utterances "in X" and "on X", where X is an object category. We assume that container-like objects will more frequently occur in sentences with "in X" than "on X" thus giving us a simple test for how likely an object X is to contain other objects. We approximate "in X" as the average log probability of "in a(n) X" and "in the X", and similarly for "on X". We then use the difference between these likelihoods to make the binary prediction for whether a certain object category is solid or container-like. Table 1 shows predictions obtained using this approach for several common object categories. This approach will not perform well when statements such as "in X" or "on X" are rare (e.g., rabbit) but otherwise gives correct predictions for many common categories of objects.

With these prediction for a given 3D model, we can now estimate the total solid volume by either voxelizing the surface 3D mesh representation or densely filling the same 3D voxelization. We obtain both surface and solid voxelizations of 3D meshes using the voxelization approach implemented by Binvox [19] with a resolution of 128x128x128 on the 3D model centered at the origin and rescaled to fit within a unit cube. Combined with the physical dimension estimates, we can thus compute the total occupied volume of each 3D model.

5. Demonstrative Applications

We demonstrate the usefulness of the physical attributes that we have collected by applying them to faceted semantic querying of our 3D model corpus, and to scene synthesis. In addition to the applications we describe here, we believe a semantically-enriched dataset such as the one we described can be useful for many vision and robotics applications. For instance, prior work in vision has used 3D models for detection [16, 22] and fine-grained classification [15].

5.1. Semantic Queries

The set of object attributes that we defined can be used to enable faceted querying into the 3D model corpus. Beyond the straight-forward keyword search over the category taxonomy we can now refine our queries with constraints on the dimensions of the object, the number and the attributes of static support surfaces, the material composition, and the total weight. To illustrate this form of faceted search, we compute the surface support and physical size statistics of the bookcase models in our corpus. Figure 8 shows a faceted query example where we can retrieve bookcases fitting high-level descriptions of the approximate number of books and the overall height compared to other bookcases.

5.2. Scene Synthesis

The object attributes that we collected are critical for enabling automatic scene layout and scene synthesis applications explored by much prior work [18, 26, 7]. The layout of a scene is highly constrained by the priors of static support. In other words, once we determine what objects we would like to appear in a scene, knowing how they would support each other is a big part of producing a realistic scene (e.g., plates are typically on dining tables). Static support priors

[2] http://physics.nist.gov/cgi-bin/Star/compos.pl

[3] http://www.engineeringtoolbox.com/wood-density-d_40.html

Figure 8. The physical properties we collected allow us to perform high-level faceted queries into our 3D model corpus, demonstrated here by searching for combinations of "tall" (above 80th percentile height), "short" (below 20th), "can fit 20 or 100 books", assuming each book requires 100 cm^2 of horizontal shelf space.

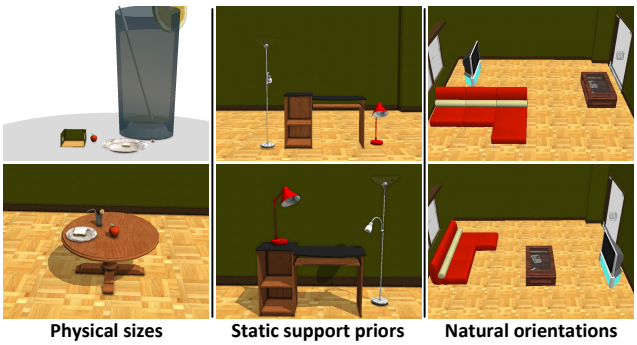

Figure 9. Comparison of scene synthesis without (top) and with (bottom) annotations of physical sizes, static support surface priors, and natural orientations. Scenes generated with the system of Chang et al. [4] constrained to use the same set of models with and without each of the priors.

allow us to transform a set of objects into a static support tree reflecting the structure of real-world environments.

The physical sizes of 3D models are also integral to recreating a realistic 3D scene, as Figure 9 illustrates. Without priors on the absolute sizes of categories of objects and specific size values for object instances a scene synthesis algorithm can easily produce implausible configurations (Figure 9 left). Similarly, typical object orientations for each object instance's upright and front sides are invaluable (Figure 9 right). Without this information, clocks and monitors can face in the wrong orientation rendering them unusable in the synthesized scenes and lowering the plausibility of the created environment.

5.3. Materiality for Physics

Given the aggregated material distributions for each category and an estimated solid volume for a 3D model we can compute a rough approximate of the total weight for that object instance. By retrieving coefficients of static friction[4] for the object material and combining them with the predicted weight we can also compute the total force necessary to horizontally displace the object. Combined with tabulated values of the average maximum human lift and push strengths [24] we can now predict whether the object can be lifted or pushed horizontally by a person of average strength. Figure 2 illustrates some of these predictions.

Though this rough approximation makes a series of naïve simplifying assumptions, it still demonstrates the benefit of physical property annotations on 3D models for reasoning about how people might physically interact with common objects.

6. Future Work and Discussion

We defined and collected several key properties of 3D models which can be used to answer common sense questions. We provided a dataset of 3D models that have been enriched with these properties. Finally, we demonstrated how such a richly-annotated 3D model corpus can be useful in the setting of 3D scene synthesis, in faceted semantic queries, and in predicting how people can interact with the objects.

This is a small step towards the goal of a large-scale, richly-annotated 3D model dataset. Following on this work, we plan to use crowdsourcing to create a broader range and larger volume of verified annotations. These annotations can be used as ground truth data that will enable quantitative evaluation of algorithmic predictions. We also hope that this dataset will enable future research on the propagation of semantic attributes to larger scale model datasets.

While we have highlighted some important physical attributes, there are many other annotations that are useful. For instance, part segmentation and part level annotation (e.g., name, attributes, functionalities) are extremely important for a finer-granularity understanding of object materiality and functionality.

We hope this work will inspire the community to think about how richly annotated 3D models can be used in a variety of problems that deal with common sense knowledge.

References

[1] M. Aubry, D. Maturana, A. A. Efros, B. C. Russell, and J. Sivic. Seeing 3D chairs: exemplar part-based 2D-3D alignment using a large dataset of CAD models. In *CVPR*, 2014. 4

[4] http://www.engineeringtoolbox.com/friction-coefficients-d_778.html

[2] S. Bell, P. Upchurch, N. Snavely, and K. Bala. OpenSurfaces: A richly annotated catalog of surface appearance. *ACM Trans. on Graphics (SIGGRAPH)*, 32(4), 2013. 3, 5

[3] C. Buck, K. Heafield, and B. van Ooyen. N-gram counts and language models from the common crawl. LREC, 2014. 6

[4] A. X. Chang, M. Savva, and C. D. Manning. Learning spatial knowledge for text to 3D scene generation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014*, 2014. 4, 7

[5] W. Choi, Y.-W. Chao, C. Pantofaru, and S. Savarese. Understanding indoor scenes using 3d geometric phrases. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 33–40. IEEE, 2013. 2

[6] L. Del Pero, J. Bowdish, B. Kermgard, E. Hartley, and K. Barnard. Understanding bayesian rooms using composite 3d object models. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 153–160. IEEE, 2013. 2

[7] M. Fisher, D. Ritchie, M. Savva, T. Funkhouser, and P. Hanrahan. Example-based synthesis of 3D object arrangements. *ACM Transactions on Graphics (TOG)*, 2012. 3, 4, 6

[8] H. Fu, D. Cohen-Or, G. Dror, and A. Sheffer. Upright orientation of man-made objects. *ACM Transactions on Graphics*, 2008. 4

[9] D. Geman, S. Geman, N. Hallonquist, and L. Younes. Visual turing test for computer vision systems. *Proceedings of the National Academy of Sciences*, 2015. 2

[10] K. Heafield. Kenlm: Faster and smaller language model queries. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 187–197. Association for Computational Linguistics, 2011. 6

[11] Y. Jiang, H. Koppula, and A. Saxena. Hallucinated humans as the hidden context for labeling 3d scenes. In *CVPR*, 2013. 2

[12] Y. Jiang, M. Lim, C. Zheng, and A. Saxena. Learning to place new objects in a scene. *The International Journal of Robotics Research*, 2012. 2

[13] A. D. Kalvin and R. H. Taylor. Superfaces: Polygonal mesh simplification with bounded error. *Computer Graphics and Applications, IEEE*, 1996. 4

[14] T. Konkle and A. Oliva. A real-world size organization of object responses in occipitotemporal cortex. 2012. 3

[15] J. Krause, M. Stark, J. Deng, and L. Fei-Fei. 3D object representations for fine-grained categorization. In *4th International IEEE Workshop on 3D Representation and Recognition (3dRR-13)*, Sydney, Australia, 2013. 6

[16] J. Liebelt and C. Schmid. Multi-view object class detection with a 3D geometric model. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1688–1695. IEEE, 2010. 6

[17] M. Malinowski and M. Fritz. Towards a visual turing challenge. *arXiv preprint arXiv:1410.8027*, 2014. 2

[18] P. Merrell, E. Schkufza, Z. Li, M. Agrawala, and V. Koltun. Interactive furniture layout using interior design guidelines. *ACM Transactions on Graphics (TOG)*, 30(4):87, 2011. 6

[19] F. S. Nooruddin and G. Turk. Simplification and repair of polygonal models using volumetric techniques. *Visualization and Computer Graphics, IEEE Transactions on*, 2003. 6

[20] M. Savva, A. X. Chang, G. Bernstein, C. D. Manning, and P. Hanrahan. On being the right scale: Sizing large collections of 3D models. *Stanford University Technical Report CSTR 2014-03*, 2014. 3, 4

[21] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser. The Princeton shape benchmark. In *Shape Modeling Applications, 2004. Proceedings*. IEEE, 2004. 2

[22] S. Song and J. Xiao. Sliding shapes for 3d object detection in depth images. In *ECCV*. 2014. 2, 6

[23] M. Tenorth, S. Profanter, F. Balint-Benczedi, and M. Beetz. Decomposing cad models of objects of daily use and reasoning about their functional parts. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 5943–5949. IEEE, 2013. 2

[24] A. R. Tilley, J. Anning, and R. Welles. *The Measure of Man and Woman: Human Factors in Design. Revised Edition*. Wiley & Sons, 2002. 7

[25] Z. Wu, S. Song, A. Khosla, X. Tang, and J. Xiao. 3D ShapeNets for 2.5D object recognition and next-best-view prediction. *CVPR*, 2015. 2

[26] L.-F. Yu, S. K. Yeung, C.-K. Tang, D. Terzopoulos, T. F. Chan, and S. Osher. Make it home: automatic optimization of furniture arrangement. *ACM Trans. Graph.*, 30(4):86, 2011. 6

[27] L.-F. Yu, S.-K. Yeung, and D. Terzopoulos. The clutterpalette: An interactive tool for detailing indoor scenes. *Visualization and Computer Graphics, IEEE Transactions on*, 2015. 2

[28] Y. Zhao and S.-C. Zhu. Scene parsing by integrating function, geometry and appearance models. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3119–3126. IEEE, 2013. 2

[29] B. Zheng, Y. Zhao, J. C. Yu, K. Ikeuchi, and S.-C. Zhu. Beyond point clouds: Scene understanding by reasoning geometry and physics. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3127–3134. IEEE, 2013. 2

[30] B. Zheng, Y. Zhao, J. C. Yu, K. Ikeuchi, and S.-C. Zhu. Detecting potential falling objects by inferring human action and natural disturbance. In *ICRA*, 2014. 2

[31] Y. Zhu, A. Fathi, and L. Fei-Fei. Reasoning about object affordances in a knowledge base representation. In *ECCV*. 2014. 2