1 Introduction and Rationale

Statistical methods leveraging massive amounts of digital data have been central to recent advances in computing. A key prerequisite for training and evaluating such methods is the creation of large curated datasets such as WordNet [61] and the Penn Treebank [56] in natural language processing, and ImageNet [19] in computer vision. These datasets have tens of thousands of citations collectively, indicating their far-reaching impact through rapid and pervasive use in research and industry. In particular, ImageNet recently has played a pivotal role in pushing the frontiers of computer vision by providing the data necessary for deep learning–based methods, leading to breakthroughs in object recognition and classification.

A similar revolution has yet to happen with **3D geometry data**. 3D representations are the most faithful digital encoding of physical objects, allowing us to store and manipulate high-level information (e.g., affordances, function) and low-level features (e.g., appearance, materials) about the object in a way that is more pure and complete than lower dimensional representations (such as images), or entirely symbolic ones (such as text-based knowledge graphs). Advances in processing 3D geometry can have huge impact on many fields including computer graphics/vision, robotics, mechanical engineering, computational biology, medicine, entertainment and e-commerce. Yet, existing algorithms for 3D geometry processing were developed and evaluated on small datasets, limiting their utility in solving real-world problems.

The lack of available 3D data is a key obstacle. Less than a decade ago, 3D data acquisition was mostly performed in specialized lab environments, and it was hard to obtain 3D data online. This situation has changed completely in recent years. Various large-scale online 3D repositories have appeared, including the Trimble 3D warehouse (2.5M shapes in total), Turbosquid (300K shapes) and Yobi3D (1M shapes). Moreover, the emergence of cheap, portable scanning devices such as Microsoft Kinect, Intel RealSense, and Google Tango has significantly reduced the cost of acquiring 3D geometry. As depth sensors are integrated into laptops, tablets, and cell phones, we can expect an explosion of available 3D geometry data online — just as has happened with images and videos. The time is ripe for a revolution in applying statistical methods to big 3D data. A large-scale, curated 3D dataset will allow us to develop algorithms to automatically organize, search and use this 3D data as well as connect it to other modalities. However, existing 3D repositories are generally unorganized and noisy. Furthermore, they are not intended for research so they do not provide consistent labels and, most of all, lack important geometric and semantic annotations. While some research-focused 3D datasets exist, they are small and tailored to specific research tasks.

The key goal of this proposal is to create the first large-scale 3D model infrastructure named **ShapeNet** *— a repository that organizes 3D raw geometry for research purposes and annotates it with rich semantic information*. We focus on 3D shapes and 3D scenes that can be found online, and propose to organize them into fine categories and compute common geometric and semantic attributes, exploiting both crowd-sourcing and algorithmic propagation tools. The proposed infrastructure advances the state-the-of-art on multiple fronts. The size of the dataset (estimated 3M in total) is two orders of magnitude larger than the combination of existing organized 3D datasets in total. The number of categories we consider (4K) is one order of magnitude more than the state-of-the-art. In particular, unlike existing datasets that focus on a specific task (e.g., classification and segmentation), we plan to create a comprehensive dataset with rich geometric and semantic attributes, including orientations and part decompositions of shapes, symmetries and associated transformations, geometric correspondences across shapes within each category (and even to real images of such objects), as well as fine-grained functional labels (e.g., the functions of parts in each shape and objects in each scene). We believe such a dataset is valuable because different tasks are correlated, and these geometric and semantic attributes will provide a valuable and unique resource for algorithm development in related scientific disciplines. A preliminary version of this 3D repository already exists and, as of January 2016, over 87 research teams from leading universities and institutes such as Berkeley, MIT, CMU, Cornell, Max Planck Center, and EPFL have been using ShapeNet in their research.

2 Background and Related Work

There has been substantial growth in the number of of 3D models available on-line over the last decade, with repositories like the Trimble 3D Warehouse providing millions of 3D polygonal models covering thousands of object and scene categories. Yet, there are few collections of 3D models that provide useful organizations and annotations. Meaningful textual descriptions are rarely provided for individual models, and online repositories are usually either unorganized or grouped into gross categories (e.g., furniture, architecture, etc. [26]). As a result, they have been poorly utilized in research and applications.

There have been previous efforts to build organized collections of 3D models (e.g., [21, 26]). However, they have provided quite small data sets, covered very few semantic categories, and included few structural and semantic annotations. Most of these previous collections have been developed for evaluating shape retrieval and classification algorithms. For example, data sets are created annually for the Shape Retrieval Contest (SHREC) that commonly contains sets of models organized in object categories. However, those data sets are very small — SHREC 2014 [49] contains a "large" dataset with around 9,000 models consisting of models from a variety of sources (Table 1): *Princeton Shape Benchmark (PSB)* [71], *SHREC 2012 Generic Shape Benchmark (SHREC12GTB)* [48], *Toyohashi Shape Benchmark (TSB)* [85], *Konstanz 3D Model Benchmark (CCCC)* [91], *Watertight Model Benchmark (WMB)* [89], *McGill 3D Shape Benchmark (MSB)* [100], *Bonn Architecture Benchmark (BAB)* [93], *Purdue Engineering Shape Benchmark (ESB)* [37] organized into 171 categories.

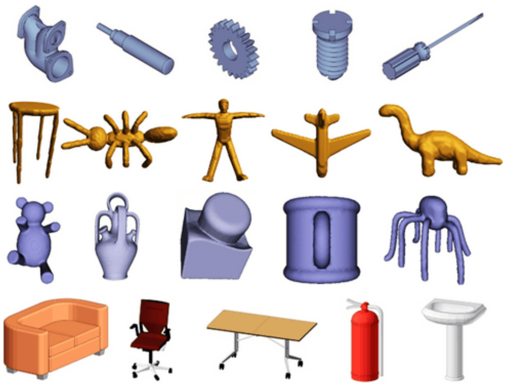


Figure 1: Some models from SHREC 2014.

The *Princeton Shape Benchmark* (provided by the PIs) is probably the most well-known and frequently used 3D shape collection to date (with over 1000 citations) [71]. It contains around 1,800 3D models grouped into 90 categories, but has no annotations beyond categorization. Other commonly-used datasets (many provided by the PIs) contain segmentations [16], correspondences [43, 42], hierarchies [54], symmetries [41], salient features [17], semantic segmentations and labels [98], alignments of 3D models with images [95], semantic ontologies [21], and other functional annotations — but again only for small datasets. For example, the *Benchmark for 3D Mesh Segmentation* consists of just 380 models in 19 object classes [16].

In contrast, there has been a flurry of activity on collecting, organizing, and labeling large datasets in computer vision and related fields. For example, *ImageNet* [19] provides a set of 14M images organized into 20K categories associated with "synsets" of WordNet [61]. *LabelMe* provides segmentations and label annotations of hundreds of thousands of objects in tens of thousands of images [67]. The *SUN* dataset

Benchmarks	Types	# models	# classes	Avg # models per class
SHREC14LSGTB	Generic	8,987	171	53
PSB	Generic	907+907 (train+test)	90+92 (train+test)	10/10 (train/test)
SHREC12GTB	Generic	1200	60	20
TSB	Generic	10,000	352	28
CCCC	Generic	473	55	9
WMB	Watertight (articulated)	400	20	20
MSB	Articulated	457	19	24
BAB	Architecture	2257	183+180 (function+form)	12+13 (function+form)
ESB	CAD	867	45	19

Table 1: Source datasets from SHREC 2014.

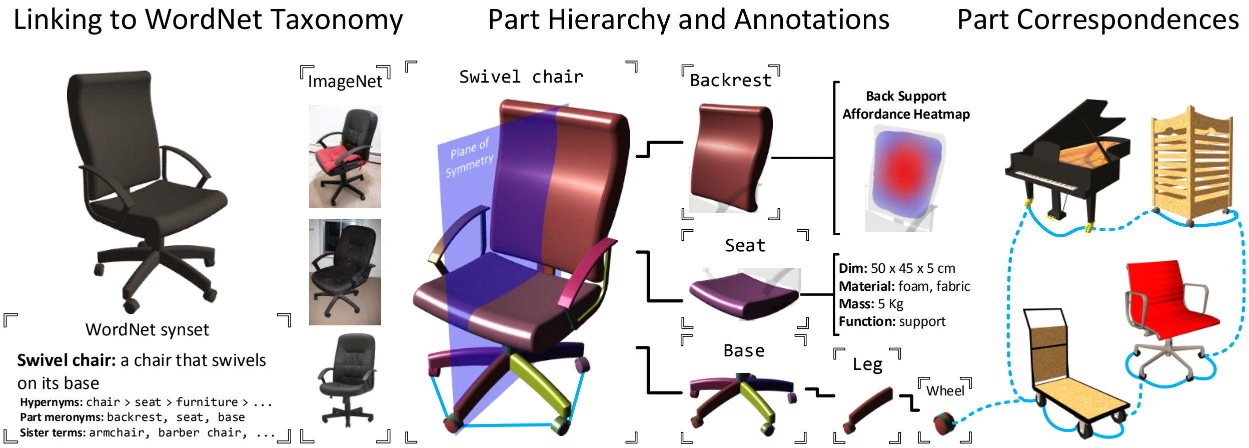

Linking to WordNet Taxonomy
Part Hierarchy and Annotations
Part Correspondences
ImageNet
Swivel chair
Plane of Symmetry
Backrest
Back Support Affordance Heatmap
Seat
Dim: 50 x 45 x 5 cm
Material: foam, fabric
Mass: 5 Kg
Function: support
Base
Leg
Wheel
WordNet synset
Swivel chair: a chair that swivels on its base
Hypernyms: chair > seat > furniture > ...
Part meronyms: backrest, seat, base
Sister terms: armchair, barber chair, ...

Figure 2: Illustrated views of ShapeNet annotations at different levels for an example chair model. *Left:* links to the WordNet taxonomy provide definitions of objects, is-a and has-a relations, and a connection to images from ImageNet. *Middle:* hierarchical decomposition of shape into parts on which various attributes are defined: names, symmetries, dimensions, materials, masses, and affordances. *Right:* part-to-part and point-to-point connections are established at all levels within ShapeNet producing a dense and semantically rich network of correspondences.

provides 3M annotations of objects in 4K categories appearing in 131K images of 900 types of scenes. These large datasets and others (e.g., [45, 51]) have revitalized data-driven algorithms for recognition, detection, and editing of images, which in turn have revolutionized computer vision.

Similarly, large collections of annotated 3D data have had great influence on progress in other disciplines. For example, the *Protein Data Bank* [7] provides a database with over 100K protein 3D structures, each labeled with its source and links to structural and functional annotations [47]. This database is a common repository of *all* 3D protein structures solved to date and provides a shared infrastructure for the collection and transfer of knowledge about each entry. It has accelerated the development of data-driven algorithms in structural biology, has facilitated the creation of benchmarks, and has linked researchers and industry from around the world. We aim to provide a similar resource for 3D models of everyday objects.

3 ShapeNet: An Information-Rich 3D Model Repository

ShapeNet is a large, information-rich repository of 3D models. It contains millions of models collected from a variety of online sources and spanning a multitude of semantic categories. Unlike previous 3D model repositories, it provides extensive sets of annotations for every model and links between models in the repository and to other multimedia data outside the repository.

Like ImageNet, ShapeNet provides a hierarchical categorization of shapes according to WordNet synsets (see Figure 2). However, what differentiates ShapeNet from other repositories is the rich set of annotations provided for each shape and the correspondences provided between shapes. The annotations include geometric attributes such as upright and front orientation vectors, scale of object in real world units, shape symmetries (reflection plane, other rotational symmetries) and part decompositions, as well as semantic attributes such as materials and human-centric interaction (graspability, support, gaze, other functionalities). These attributes provide valuable resources for processing, understanding and visualizing 3D shapes in a way that is aware of the semantics of the shape. The correspondences provide links between semantically equivalent surface regions of different models within ShapeNet as well as links from surfaces in ShapeNet to images in ImageNet and other multimedia data sources. These correspondences are critical for information propagation and aggregation, shape exploration, shape retrieval, appearance modeling, and object recognition.

So far we have collected more than 3 million shapes from online 3D model repositories, and categorized 300 thousand of them against the WordNet taxonomy. We have created two subsets from these categorized models: *ShapeNetCore* and *ShapeNetSem*. ShapeNetCore provides consistent orientations for 55 common object categories (covering all 12 PASCAL 3D+ categories) with more than 50K unique 3D models. ShapeNetSem is a smaller, more densely annotated subset of 12K models spread over 270 categories. In addition to verified categories and alignments, it provides real-world dimensions, and estimates of object material density, volume and weight. The results of these preliminary efforts are available online at `http://www.shapenet.org` and described in more detail in a publicly released technical report [10].

In the following sections, we discuss how 3D models are collected for ShapeNet, what annotations will be added, how those annotations will be generated, how annotations will be updated as the dataset evolves over time, and what tools will be provided for the community to search, browse, and utilize existing data, as well as contribute new data.

4 Data Collection

The raw data for ShapeNet comes from 3D models that are available online or uploaded by members of the community. It will be an evolving repository with regular updates as more and more 3D models become available on the Web, as new 3D sensors become prevalent, and as more people contribute data.

To seed the data set, we have collected a large set of 3D polygonal models from two popular public repositories: Trimble 3D Warehouse [87] and Yobi3D [99]. The Trimble 3D Warehouse contains 2.4M user-designed 3D models and scenes. Yobi3D contains 350K additional models collected from a wide range of other online repositories. Together, they provide a diverse set of shapes from a broad set of object and scene categories — e.g., many organic shape categories (e.g., humans and mammals), which are rare in Warehouse3D, are plentiful in Yobi3D. In total, we have collected more than 3M models and categorized 300K of them into more than 4K categories.

In any effort like this there is always a tension between being inclusive of more models and model formats, vs. imposing certain restrictions which encourage homogeneity in the repository and facilitate the annotation acquisition and processing to be described. We will address this trade-off as we go along, according to the resources available for the project. Our initial intent is to collect 3D models from a wide variety of online repositories, described with multiple shape representations (polygonal meshes, point clouds, CSG, voxels, etc.), and stored in a range of industry-standard file formats (COLLADA, Wavefront OBJ, PLY, etc.). In particular, we intend to support both "designed" 3D models as well as "acquired" ones, including multi-view RGBD data, which will be a more common 3D data source in upcoming years, as RGBD cameras become prevalent. Whenever possible, ShapeNet will provide basic tools for converting between these shape representations and file formats. Future research projects may also enhance this functionality.

5 Annotation Types

ShapeNet is far more than a collection of 3D models: it also includes a rich set of annotations that provide information about those models, links between them, and links to other sources of data. These annotations are exactly what make ShapeNet uniquely valuable — the value of this dense network of interlinked attributes on shapes is illustrated in Figure 2.

This section describes what annotations we plan to include, why we have chosen to include them, and how we plan to obtain them. The next sections describe methodologies for representing, propagating, verifying, updating, and searching such annotations.

Category Annotations. The most important annotations provided with each 3D model are its categories. Categories provide semantic labels that are useful for indexing, grouping, and linking to related sources of data. As described in the previous section, we organize ShapeNet based on the WordNet [61] taxonomy, a widely-used, English lexical database that groups words into cognitive synonyms (synsets). Synsets are interlinked with various relations, such as hyper and hyponym, and part-whole relations. Due to the popularity of WordNet, we can leverage other resources linked to WordNet such as ImageNet, ConceptNet, Freebase, and Wikipedia. In particular, linking to ImageNet [19] will help information transport between images and shapes, in both directions.

Property Annotations. 3D objects have structure that strongly correlates with their semantics (form to function relationship). Every 3D model in ShapeNet will be provided with annotations describing properties of its shape/form, structure, materials, representations, and expected placements. These properties are useful for searching the collection for relevant models, for organizing collections of models, and for reasoning about how to compose models into meaningful subcollections and/or scenes.

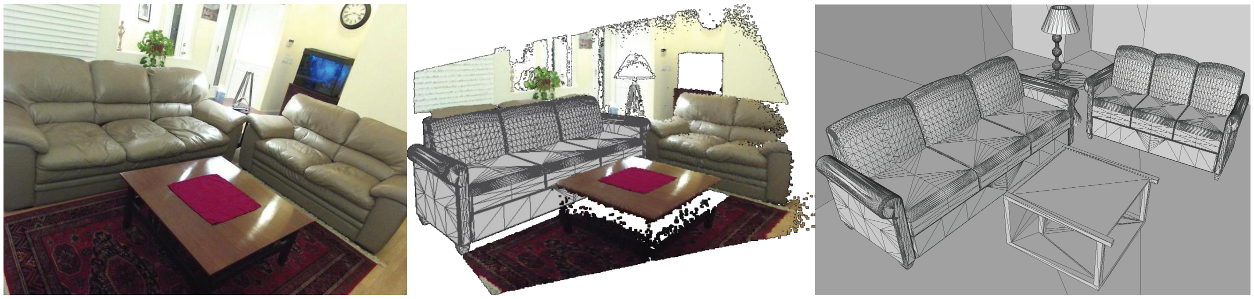
- **Internal structural/geometric properties.**
  - *Hierarchical part decompositions:* Each model is annotated not only with its top-level category labels, but also with a hierarchical decomposition of its surface into labeled parts. For example, the surface of a swivel chair might be decomposed into subsurfaces representing the back, seat, arms, and base. These subsurfaces may be divided further, e.g., the swivel base is decomposed into a main leg column and the individual feet. Each part is tagged with a reference to the WordNet synset representing the part, if one exists. These part decompositions serve as the building blocks for many applications such as shape modeling/synthesis, understanding shape structures, and computing abstractions for visualization. They also provide links between WordNet synsets. For example, a ShapeNet query to retrieve "wheels" can return not only the models representing a single wheel, but also all the parts of models that have been identified as wheels in the hierarchical decomposition of larger objects.
  - *Symmetries:* Many real-world objects exhibit symmetries, (e.g., planar reflectional symmetries, k-fold rotational symmetries, etc.), which are useful to maintain as the model is edited [31], leverage when finding correspondences [55], and/or guide physical interactions with other objects. Accordingly, ShapeNet will provide descriptions (e.g., a reflection plane) for a discrete set of prominent perfect, partial, and/or approximate symmetries for each model.

- **Physical properties.**
  - *Materials:* The material properties of 3D models are important for many applications, including rendering, recognition, and physical simulation. Accordingly, parts of models in ShapeNet will be annotated accordingly, with references to the physical materials used in their construction, whenever those can be determined. For example, the entirety of an IKEA chair might be labeled with "pine wood," or a fancy dining room chair might have labels of "mahogany wood" on the legs and back and "velvet fabric" on the seat. As a starting point, we will consider only surface materials. However, construction materials could be added as the repository evolves.
  - *Weight, Strength, etc.:* Physical properties of 3D models are also very valuable for simulations and for reasoning about object uses, 3D printing and manufacturing, object participation in assemblies, etc.. For example, when composing a scene, it might be important to know how much an object weighs, how much weight in can bear, where its center of mass is, its brittleness, its deformation modes, etc. Such properties might be available from the model source or, in some cases, it might be possible to estimate thm from simulations, images or videos.
  - *Appearance:* Whenever available, we also record, color, texture, and other appearance information.

- **External interaction and placement properties.**
  - *Transformations:* Establishing a consistent canonical origin, scale, and orientation (e.g., upright and front) for every model is important for various tasks such as visualizing shapes [43], shape classification [34] and shape recognition [94]. Fortunately, most shapes in ShapeNet are by default placed in the upright orientations, and the front orientations are typically aligned with an axis. Therefore, we provide a simple interface for users to select the front orientations. This is sufficient for most of the shapes, and in rare cases we ask users to rotate shapes for annotating front orientations.
  - *Contact relationships:* Knowing how a model is typically in contact with other objects is useful for placing it into a scene and reasoning about how it is affected by changes to other models in a scene. Specifically, we augment each model by marking regions typically in contact with its supporting environment (e.g., tips of the legs of a desk) and regions suitable for supporting other objects (e.g., the flat surface on the top of a desk).
  - *Affordances:* Knowing how people usually interact with a 3D model (i.e., which parts of a human body come in contact with which points on the 3D surface) is useful for predicting functions and ergonomics of an object [27], reasoning about navigation through a scene [32, 83], and guiding algorithms that process the model to maintain properties important for human use [46, 38]. Accordingly, we augment each model with a discrete set of human poses and human-surface contact points representing likely interaction modalities.
  - *Canonical views:* Knowing how people typically view a model is useful for generating representative images (e.g., thumbnails), reasoning about surface saliency (e.g., what people typically see), and guiding model processing (e.g., simplification) [69]. Accordingly, we plan to provide a discrete set of "canonical views" for each model.

Correspondence Annotations.

- *Model-model correspondences:* Rather than being a repository of individuals, as a prominent feature (and to justify the "Net" in its name), ShapeNet records relations and provides correspondences between 3D models. These correspondences enable a wide range of applications, including propagating attributes across shapes, joint shape understanding [35] and shape exploration [43, 35]. Based on the characteristics of shapes, we consider three types of correspondences that are meaningful in the context: dense point-wise correspondences, part-wise correspondences, and sparse region-wise correspondences (including keypoints). Point-wise correspondences are suitable for shapes that are very similar to each other, e.g., between two human models. Part-wise correspondences exist among diverse shape collections that exhibit structural similarities, e.g., between two sedan vehicle models. Region-wise correspondences are used to capture relations across categories, e.g., contain contact regions of human hands/legs on various models (bicycles, chairs, keyboards). These correspondences can be established using a variety of extant techniques, using geometric shape descriptors, whole shape alignments [88] and joint correspondence optimization among a collection of shapes [43, 33, 35]. To control the quality of the resulting correspondences, we incorporate user-specified correspondences as constraints when optimizing correspondences.
- *Model-image correspondences:* It is also be valuable to include links from 3D models to other data that provide complimentary representations of the same objects or scenes. For example, we plan to establish correspondences between points on the surfaces of ShapeNet models and pixels of ImageNet photos and other large image collections. A preliminary version of this idea was introduced in PASCAL3D+ [95], which provided registrations between images and 3D models for 12 rigid categories in the PASCAL VOC 2012 dataset [20]. It was investigated further in the SUN RGB-D dataset, which provides alignments of ShapeNetCore models to more than 10,000 RGB-D images of indoor scenes



a) Input image b) 3D model alignment c) Scene reconstruction

Figure 3: Alignment of 3D models to images provides scene reconstructions and pixel-surface correspondences.

[74]. These pixel-accurate links provide opportunities for transfer of annotations (material properties, surface normals, part decompositions, etc.) and CAD-quality scene reconstructions (Figure 3).

In addition to establishing pixel-point correspondences, we are using links between images and 3D models to learn higher-level relationships. For example, we learn a similarity metric between images and 3D models by embedding them jointly into a common space where images of objects are near the 3D models they depict. We compute the embedding space first using the 3D models alone, as they capture object geometry in a more pure and complete form, and then we train a deep CNN to embed images of the models with variations in lighting, viewpoint, and occlusions into the same space — effectively learning how to ignore these nuisance factors when computing similarities of images to models (or to each other) [50]. Related ideas are being used to learn viewpoint predictors [78], depth estimators [77], and object detectors [75] from training sets of ShapeNet model-image pairs.

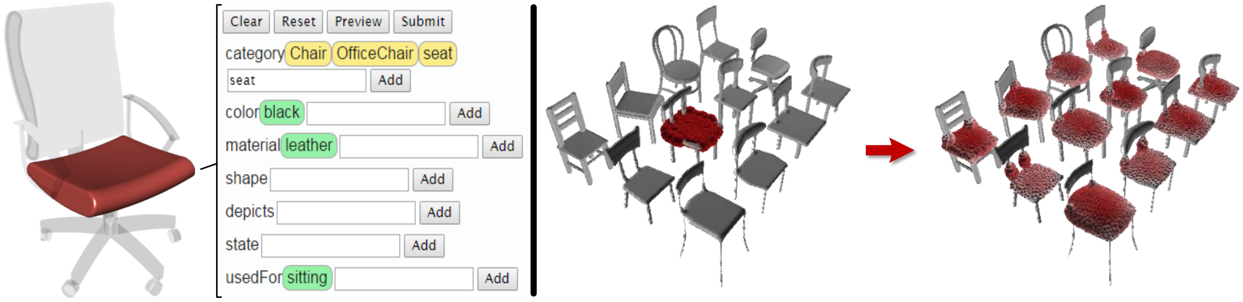
In general, the issue of compact and informative representation of all the above semantic attributes (parts, symmetries, attributes, affordances, correspondences, etc.) over shapes raises many interesting questions that we will need to address as part of this effort.

6 Annotation Acquisition, Propagation and Validation

A key challenge and source of innovation in ShapeNet is the methodology for acquiring and propagating annotations. Our goal is to provide all the annotations listed in the previous section with high accuracy. Although we cannot always guarantee that, we aim to estimate a quality/confidence metric for each annotation, as well as record its provenance — typically from a mix of human-generated data and algorithmic inference/propagation techniques. This will enable others to properly use and trust the information we provide.

6.1 Annotation Capture and Propagation

Human annotation of 3D shapes is very time-consuming and expensive. Previous data efforts to manually label collections with millions of entries have considered only category annotations, which requires just a single identifier per entry. In contrast, we intend to provide a rich set of annotations, including segmentations, which require tracing detailed segment boundaries [16, 67] and correspondences, which demand identifying semantically equivalent points or regions on *pairs of models* [43]. Obtaining this information manually for all of ShapeNet would take thousands of people-years, and therefore is impractical. Alternatively, we could use algorithms to predict annotations automatically, possibly utilizing a small set of training data provided by manual annotation to bootstrap the process.



category: Chair, OfficeChair, seat
seat
color: black
material: leather
shape
depicts
state
usedFor: sitting

Figure 4: Left: annotation interface for assigning attributes to shape parts. Right: algorithmic propagation of annotations from the single chair for which a human expert has provided an annotation, to other chairs in ShapeNet.

Most extant work along these lines has focused on the propagation of coarse category labels for shapes, using semi-supervised learning techniques. Coupled with geometric alignments, such methods can actually estimate where in the geometry is the support for a particular label (e.g., what makes a rocking chair a rocking chair) [34]. However, although such automatic algorithms for 3D shape analysis have improved vastly in the last decade, they still might not provide a satisfactory level of accuracy for ShapeNet users. For example, state-of-the-art methods for predictions of coarse model categories in the most recent SHREC challenge achieve approximately 85% accuracy [49]. To address this issue, we plan on using a hybrid approach. Our strategy is to use crowd-sourcing platforms like the Mechanical Turk to annotate part of the collection manually, use algorithms to propagate and predict annotations for remaining models, and then use crowd-sourcing again to verify or reject the predictions. This strategy leverages the fact that people can usually verify whether an annotation is correct much more quickly than they can generate it from scratch.

A novel challenge posed by ShapeNet is the richer annotation types that might describe functional parts of a 3D model, affordances, part labels, etc. — which go beyond coarse word annotations for the entire shape. A large class of these attributes can be thought of as functions painted on the surface of a shape that indicate the presence (or strength) of a key property. For example, the seat of a chair might be just be the 0-1 indicator function of where the seat portion is in a chair model. Or, for a shoe model, we may want to know the amount of pressure each point of the sole exerts on the foot during walking. Thus we need both user interfaces that allow users to "paint functions on surfaces" as well as algorithmic tools that propagate such functions between models (Figure 4). Note that anytime we can establish correspondences between 3D models (or modes and images), we have a possible "propagator" to transfer information between the models.

Fortunately, our functional formulation [64] of maps between shapes allows for soft attributes and for inconsistencies among users. In part decomposition, for example, we can ask multiple users to label the part decomposition of each individual shape, as was done on the Princeton Segmentation Benchmark [16]. Evaluation of shape segmentations can be done using the protocols described therein.

In order to generate functional annotations at scale in ShapeNet we plan to use a two-pronged strategy.

- *Attribute transportation networks in dense shape collections:* As we get more and more data, we are effectively sampling more and more densely the underlying shape space. This makes it possible to find better correspondences between nearby shapes, allowing more reliable information transport between them. Furthermore, by interconnecting many such pairs into a multiply connected network, we can cross-check the validity of information transport by moving along multiple paths. We can hope that correct information is supported by evidence from multiple directions and will aggregate coherently, while attribute noise will cancel out — but in all cases we get information, as as well as a quality measure for the information. We have used such network analysis techniques to derive shape parts in shape collections, in a completely unsupervised manner — the parts emerge as the functions most consistently transported by the network [35]. These methods can easily incorporate and benefit from supervised information.

- *Active learning:* We will generate annotation queries for users based on the quality of the information in the network — in other words, we will generate attribute annotation queries for those shapes for which additional information would most increase quality for that attribute in the network. We will also in the same spirit select the most informative views of those shapes to display to the annotators. Finally, we will learn which transporters are most appropriate for which attributes, by cross-checking not only multiple transportation paths but also multiple transport operators.

As the above makes clear we view the operations of user-assisted annotation and annotation prorogation as tightly coupled and intertwined, each assisting the other.

6.2 Annotation Provenance and Quality

A critical aspect of this hybrid approach is that every annotation produced in ShapeNet will be associated with a record of how it was created (its provenance) and how confident the creator (person or algorithm) is about its accuracy. Indeed, a model might have multiple annotations generated for the same property, and they could even be conflicting. For example, a model may have an orientation annotation indicating that it was created by an particular algorithm, which estimated it to be correct with 90% confidence. It may also have a conflicting orientation annotation produced by another algorithm with a different confidence level. Either of these annotations could be marked as confirmed by a person, or there could even be another conflicting orientation annotation created manually by a person. The general idea is that every annotation provides complete information about its provenance and accuracy, and tools are provided to aggregate annotations and to resolve conflicts as briefly outlined above. Such divergence of opinions must be tolerated, especially if in the future ShapeNet tries to store more subjective attributes (e.g., is this shape “beautiful”?). Ultimately this will require us to model not only shapes but also humans reasoning about shapes.

The exact methodology for producing and combining manual annotations, algorithmically produced annotations, and crowd-sourced verifications may be unique for every type of annotation, because the effort required for manual annotation, the accuracies achieved by automatic algorithms, and the interfaces for manual verification may differ. However, the general principles will be the same for all of them.

7 Use, Evolution and Maintenance

We aim to make ShapeNet easily accessible and evolvable. With the explosion of online 3D data, we expect the repository to evolve continuously in the quality, quantity, and diversity of shapes and annotations. Recognizing the success of community-driven ecosystems such as Wikipedia, we aim to cultivate a dynamic user community and encourage user contributions. To support this, we will provide a documented, open source API and software tools for contributing to ShapeNet. This will encourage continuous addition of 3D models and annotations, and also allow for improved quality assurance as ShapeNet grows. We will use a combination of algorithmic tools and community-building efforts in service of this goal:

Search: For ShapeNet to be generally useful, it must be searchable. For geometry-based search, initially ShapeNet will incorporate the Princeton 3D search engine as described in [23] based on spherical harmonic descriptors which are translation and rotation invariant. This allows a simple high-dimensional nearest neighbor search to be used for retrieval of similar models. In addition, simpler methods, such as those based on the D2 shape descriptor [63], can be used as preliminary filters, by clustering shapes into more homogeneous classes. View-based methods may also be explored, such as the lightfield descriptor of [15]. Such methods, unfortunately, are not rotation-independent and the matching process becomes significantly more expensive. Some variants, however, include query by user 2D sketch, which can be a useful capability. We will also provide simple textual search using traditional TF-IDF inverted indices. Perhaps most interesting will the indexing of the functional attributes described in Sections 5 and 6. If we precluster shapes into more homogeneous collections, then the functional spaces and maps machinery initiated by [64] will allow us to express all these functional attributes a simply vectors in a common shared functional space, thus allowing efficient organization and search.

Exploration: We wish to compute visualizations (2D or 3D embeddings) of modest shape collections in ways that make low-dimensional structure in the collection visible to the user. Specifically, our goal is to extract and parameterize shape variability within the collection, discovering the principal axes of variation. Note that variability can be both continuous (e.g., a chair has thicker legs than another similar chair) and discrete (a building has an additional floor as compared to another similar building). Based on functional maps, the shape difference machinery developed by the PIs [68] can be used to estimate how shapes differ, to highlight the areas of difference, and to provide useful visualizations of the collection. Shape differences can also be used for localized search, by indexing each shape via its shape differences form certain landmark shapes. Without additional preprocessing, shape difference additionally allow regional search and visualization — by focusing only on a specific area of interest within a class of shapes (e.g., we are only interested in the variation in the shapes of the backs of chairs, not the entire chairs).

User Access Control: We will make ShapeNet models and annotations browsable and visualizable to the general public. However, only registered users with confirmed research purposes will be granted download and upload privileges to ShapeNet. This will ensure that the ShapeNet infrastructure and collected data is not abused for commercial purposes, or for efforts unrelated to research.

Gathering User Feedback and Suggestions: Through online forums and regular polls on a mailing list for all registered users, we will encourage feedback and suggestions for improving ShapeNet. This will not only be helpful for improving ShapeNet, but also for guiding new research. We will also encourage sharing of information between users on the ShapeNet forum. The PIs will assign administrators who will compile user issues and address them on a regular basis. We also plan to organize workshops and panel discussions so that users and the ShapeNet construction team can meet and share information in person.

Incorporating User Data: Along with uploaded models and annotations, we will record in the provenance metadata the uploader's ID and relevant information for proper citation. This provenance metadata will be displayed to other users, further encouraging proper attribution. We will periodically process new annotations through our quality verification system, as described in Section 3. After processing, verification results will be recorded and the annotations will be published on the central ShapeNet website.

Community Toolbox: We expect that ShapeNet will have a broad impact within the research community, both in fields that traditionally deal with 3D data, such as vision and graphics, and in other fields such as paleontology and archeology. To help users access the data and upload new data, we will provide an open source, fully documented API and a set of relevant software tools. These tools will cover basic functions such as downloading, reading, visualizing shapes and annotations, uploading new shapes or annotations, and defining new annotations.

Benchmarks and Challenges: The unprecedented scale and annotation fidelity of ShapeNet will make it a strong benchmark dataset for existing and future algorithms. We plan to organize regular challenges and contests for tasks such as categorization, pose alignment, and model correspondences. These challenges will be publicized widely and the results will be contributed back to ShapeNet helping it evolve on a regular basis. We will encourage participants to submit source code so that their algorithms can be executed on all future content in ShapeNet. This strategy will strongly engage ShapeNet users by improving the visibility of their research contributions, and will simultaneously lead to growth in the ShapeNet dataset.

8 Research Enabled by ShapeNet

We have tracked ShapeNet usage via Google Scholar. Even during its very short existence, the current repository has been used in several completed research works from other research groups, including flow prediction using synthetic data [59], object detection [57] and view estimation [79] — and many more are ongoing. We are confident that the ShapeNet effort will enable research in a variety of fields within Computer Science dealing with 3D data, including of course computer graphics, computer vision, and robotics, but not confined to these.

Semantic Webs Based on 3D Shapes A promising direction for future research at the convergence of large-scale vision and graphics is to establish direct instance-level, part-level, and (when possible) point-level correrspondences between 3D shapes, images, and other online data. Linkage through 3D models can establish a powerful transport mechanism for knowledge to flow between data types. For example, product catalogues contain images as well as textual information about the sizes, materials, and physical properties of products which can both be corresponded to 3D models of the products.

Rendering 3D Shapes for Vision Tasks Rich 3D information is embodied in 3D shapes. By rendering shapes into images through particular viewpoints we project shape structure into image space, giving rise to patterns of object boundaries and occlusions. We can then vary lighting condition, material property and many other factors, infinite images can be rendered with ground-truth annotation costing negligible human efforts. Those images with annotation form large-scale training dataset for learning robust classifiers and other predictors in vision tasks. Recent work [78] explores along this line leveraging ShapeNet data and achieves state-of-the-art performance on the 3D viewpoint estimation task. There can be a big potential for many core vision tasks with this approach, including object detection and segmentation.

Priors for RGB-D-based 3D Reconstruction The rapid improvements in RGB-D technology and the ubiquity of RGB-D sensors has led to much research in 3D reconstruction. Newly published work has shown that learning a set of structural priors from collections of models such as ShapeNet can assist in completing RGB-D scan data and improving reconstruction quality [62]. The part-level annotations within ShapeNet will enable a finer-grained understanding of object structure which can be explored by learning priors for object parts and their relations. These priors can take the form of learned "shape grammars" that capture the hierarchical composition of shapes and how parts in combination give rise to affordances and functionalities.

Robotics By encoding and understanding the structure and semantics of common objects in the world we enable a broad range of data-driven, statistical approaches for robotic scene understanding, planning and manipulation. Along this direction, simulation within 3D scenes constructed with annotated 3D models has been shown to be a promising approach [36]. Patterns in the visual, geometric and physical properties of objects can aid robot recognition and manipulation. For example, object size information is useful for speeding up detection by eliminating object category hypotheses. Object part information is essential for planning robot-object interaction [84]. Object material information is useful for estimating object weights, stability and possible deformations [101].

Data-driven 3D Content Creation The rich semantics that ShapeNet will provide for 3D models will enable research into better 3D model and 3D scene design tools. Recent work in graphics has focused on algorithms for automated probabilistic shape synthesis [13, 39] and for more efficient shape design UIs [12]. The same principles can be applied to 3D scene design where contextually salient attributes of the models such as size, weight, and canonical placement orientations are critical in more efficient scene assembly. In particular, recent work in text to 3D scene generation [11] has shown how important these attributes are for language-based scene generation. In this sense, ShapeNet will enable the next generation of research that will democratize the 3D content creation process which is currently dominated by experts. We look forward to the day when 3D design tools become as user-friendly and intuitive as word processors.

Connecting Language and Shapes Work at the intersection of NLP and robotics has recently focused on the problem of grounding words and concepts to concrete representations of the entities and reasoning with the semantics of a physically grounded world [58, 66, 86, 92]. Distributed, continuous representations of words have been shown to be powerful features for many NLP tasks so it stands to reason that a deeper connection with the continuous representational space of 3D shapes. This is in parallel to the recent trend of bringing together NLP and vision methods to perform image captioning and image retrieval [40, 90]. The critical starting point for all this work is a parallel corpus of images and textual annotations provided by the ImageNet corpus. We envision ShapeNet as a similar starting point for bridging the spaces of shape and language. In particular, a possible outcome of organizing ShapeNet around synsets of WordNet is that it will be possible to augment WordNet with statistics of synset relationships commonly found in 3D scenes. Two of the co-PIs work closely with Christiane Fellbaum (the lead curator for WordNet) at Princeton University and are discussing ways in which ShapeNet annotations can feed back into the publicly released WordNet dataset. For example, we expect that it will be useful to augment pairs of synsets with frequencies of spatial relationships, support relationships, part cardinalities, and other prepositional relationships, which might be useful as a knowledge base for future NLP and topic modeling algorithms.

Semantic Shapes in Education Corresponding shapes to language and symbolic representations of high-level semantics has another important application in addition to core NLP tasks. Much of learning and education is supported by visuals in conjuction with language. For example, most language learning courses start with annotated diagrams of objects, people, and environments. The recent surge of computational learning systems can benefit immensely from algorithms that can better correlate 3D shapes, parts of shapes, and 3D scenes to languages both native and foreign. Research in Computer-Assisted Language Learning (CALL) [1] and domain-specific instructional material generation would find much use for better language-based 3D model retrieval and 3D model view determination (e.g., googling "3D model of an airplane showing the location of flaps and ailerons"). ShapeNet has the potential to accelerate research in this cross-modal problem space, and will help to bring closer the day when IKEA products can come with links to web-based 3D model instructions illustrating part-by-part assembly.

We are especially excited by this role ShapeNet can play as "connector" between 3D data, images, videos and language. For example, ShapeNet can improve the links of images to words, because images of related objects from very different aspects can themselves be linked through 3D models, so image word annotations can be transferred in ways impossible when operating purely in the image domain. Conversely, the extant tools for keyword-based image search and our ability to connect images and 3D models, transitively enable keyword-based 3D model search, as already illustrated in [50].

9 Broader Impacts of the Proposed Work

Impact On Other Research Disciplines: The methodologies developed for acquisition, annotation, and maintenance of 3D data in ShapeNet could have impact on other disciplines that acquire and organize 3D data sets, including paleontology, archaeology, molecular biology, mechanical engineering, and medicine. For example, it is increasingly common for researchers in paleontology, archaeology, and other fields to acquire 3D scans of artifacts for the purposes of documentation, analysis, comparison, and visualization. Yet, the currently available tools for working with 3D surface data are relatively primitive in those domains. Algorithms developed for ShapeNet might help. For example, in past work by the PIs, algorithms developed originally for computer graphics [53] have provided the basis for automatic surface matching in paleontology that is able to achieve species classification performance comparable to human experts [8]. Similarly, shape matching algorithms developed by the PIs have shown human-level performance on assembling fragments of archaeological artifacts [24].

Impact on Industrial Partners: ShapeNet has been born out of a line of research that has included collaborations with several companies, including Adobe, Autodesk, Google, and Intel. For example, Intel provided seed funding for the early stages of this collaboration between Princeton and Stanford as part of the Intel Science and Technology Center for Visual Computing (ISTC-VC). Google funded early research on 3D shape-based retrieval and scene understanding through their Faculty Research award program. Adobe supported the development of core algorithms for analyzing and visualizing correspondences in collections of 3D models through their internship and University gifts program [43]. These industrial collaborations will continue and strengthen through the development and evolution of ShapeNet.

Impact on Society: Understanding 3D data is central to many tasks that are currently performed by computers, or will be in the near future: self-driving cars, robotic assembly lines, security surveillance, online shopping, augmented reality, and 3D printing are just a few examples. ShapeNet will provide a large, evolving data set that can be used for training and benchmarking algorithms in these domains.

Education/Dissemination Plan: The project will have impact on education through tight integration with classes at the participating universities, courses at conferences, workshops organized by the PIs, and educational resources on the web. Within the past two years, the PIs have taught advanced university classes on "Data-Driven Geometry Processing" (Huang, Stanford 2014), "3D Representation and Recognition" (Saverese, Stanford 2015), "Advanced Computer Graphics" (Funkhouser, Princeton 2014), and "Geometric Modeling and Processing" (Guibas, Stanford 2014); they have organized workshops on "Scene Understanding" (Xiao, CVPR 2014), "RGB-D" (Xiao, RSS 2014), "Structured Learning for Scene Understanding" (Savarese, Stanford 2014), "Advances in Imaging and Graphics" (Funkhouser, Princeton 2014), "Functoriality in Geometric Data" (Guibas, 2015); and they have taught a conference course on "Structure-Aware Shape Processing" (Huang, SIGGRAPH ASIA 2013 and SIGGRAPH 2014). A ShapeNet-based contest is planned for SHREC 2016 and a meeting on ShapeNet related topics is already scheduled for 2017 in Dagstuhl (Guibas). These educational and dissemination activities have been accompanied by slides, exercises, bibliographies, datasets, sample source code, and project ideas distributed freely on the web. We expect these activities to continue and even increase as part of the proposed project. For example, courses on "Geometric Modeling and Processing" and "Geometric and Topological Data Analsysis" are planned for next year at Stanford, and a course on "Shape Analysis" is planned for the spring at Princeton, both of which can leverage ShapeNet Data. Such data can also be useful for undergraduate theses or summer undergraduate research projects.

Outreach: This project will also provide mentoring and research opportunities for a diverse set of students, including members of groups typically under-represented in computer science. The PIs have a strong track record on promoting diversity in research and mentoring such students.

10 The Team and Timeline

The proposed team of researchers has a past record of impactful research, interdisciplinary collaboration, community outreach, open dissemination, and tight collaboration. It combines three full professors (Guibas, Hanrahan, and Funkhouser) who are widely recognized as leaders in the field (e.g., they have 4 ACM awards and over 70K citations) with three younger faculty (Huang, Savarese, and Xiao) who are shaping the future of 3D data (e.g., expanding into RGB-D, collecting massive data sets, etc.). The team is broad not only in age, but also in background: two PIs are from computer graphics, two from computer vision, one from computational geometry, and one from applied math. Despite this diversity, they have a long history of collaboration: Huang is a recent student of Guibas, Funkhouser's last Ph.D. student was a postdoc with Guibas, Hanrahan hosted Funkhouser for a recent sabbatical, Xiao and Funkhouser co-advise students at Princeton, etc. This combination of depth, diversity, and collaboration will be a great asset for the proposed

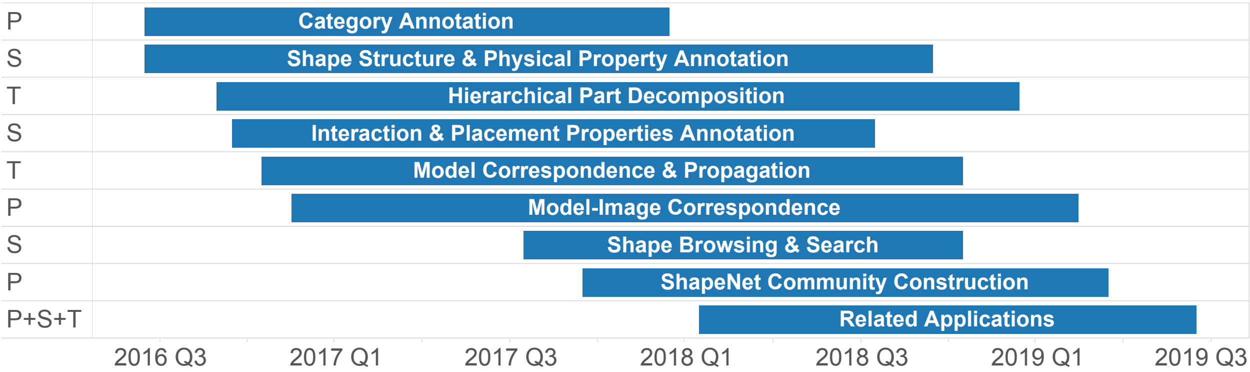


Figure 5: Three-year timeline, indicating the lead institution for each task (S = Stanford, P = Princeton, T = TTI).

project. A brief view at the project timeline is shown above. The project has the strong support of the PI departments at all three institutions (Stanford, Princeton, TTI).

11 Results from Prior NSF Support

Professor Guibas has been supported by earlier related NSF grants in geometric algorithms and computational geometry, shape analysis and geometry processing, sensor networks, and computer vision. A completed NSF grant closely related to the topics of this proposal is FODAVA grant DMS-0808515, *Global Structure Discovery on Sampled Spaces*, $450,000, July 1 , 2008 to June 30, 2011 (joint with Prof. G. Carlsson of Stanford).

Intellectual Merit: Under this grant Guibas developed new techniques for symmetry detection, both extrinsic and intrinsic [6, 73], for Voronoi-based feature extraction [60], for the computation of novel multiscale signatures of shape neighborhoods based on heat diffusion [81, 65], and for shape analysis using topological persistence techniques [14, 72].

Broader Impact: Publication [81] (heat kernel signatures) has already received over 580 citations and is widely used in geometry processing. D. Morozov, who was a postdoctoral fellow in the program, is now employed at the Lawrence Berkeley National Laboratories. Two students completed their Ph.D. with significant support from this grant (M. Ovsjanikov, Q. Huang) and now have faculty positions (Ecole Polytechnique in Paris, Toyota Technological Institute in Chicago).

Professor Funkhouser has completed several NSF grants on 3D shape analysis. The most related is FODAVA grant CCF-0937139, *Interactive Discovery and Semantic Labeling of Patterns in Spatial Data,* $500,00K, September 2009 to August 2012.

Intellectual Merit: Under recent grants, Funkhouser and his students developed new algorithms for constructing probabilistic models from collections of shapes [22, 42, 54], detecting salient regions of 3D shapes [70, 17], finding correspondences between surfaces [44, 53, 55], exploring collections of shapes [43], segmenting collections of 3D surfaces [28, 29], analyzing surface structures [52, 80], and detecting symmetries [31, 41]. Papers published on these topics in the past five years have received more than a thousand citations.

Broader Impact: Previous grants have led to publications, software, and datasets with high impact on the research community and other disciplines. Within the last five years, benchmark datasets have been released for evaluating surface segmentation algorithms [16], detecting symmetries in 3D surface models [41], finding intrinsic surface correspondences [44], exploring collections of shapes [43, 42], studying how people perceive shape in line drawings [18], and recognizing objects in point clouds [30, 9].

Professor Hanrahan has been supported by earlier related NSF grants in computer graphics, vision, and data analysis. A recent NSF grant related to the topics of this proposal is FODAVA grant DMS-

0937123, *Scalable Visualization and Model Building*, $450,000, July 1, 2008 to June 30, 2011 (joint with Prof. W. S. Cleveland of Purdue).

Intellectual Merit: Prof. Hanrahan was very involved in articulating the research agenda for the NSF FODAVA Program. Some of the achievements of this work include fast vector virtual machines and just-in-time compilers optimized for array calculations, the divide and recombine framework for data analysis, and programming language techniques for building domain-specific languages. These intellectual contributions have been documented in publications in the programming language, graphics, and visualization community. Four PhDs have resulted from the work at Stanford, Ma .Fisher, S. Lin, Z. DeVito, and J. Talbot. The work by Lin won the Best Paper Award at EuroVis 2013 and a Best Paper Honorable Mention at CHI 2013.

Broader Impact: These projects both involve creating software infrastructure for building domain-specific languages, and the development of DSLs for visualization, data analysis, and scientific computing. More specifically, several software systems have been developed that were released as open source. Two examples which are being widely used are Terra, a new environment for building DSLs, and Ripose, a new JIT compiler for the R programming language.

Professor Savarese has been supported by numerous NSF grants in computer vision. The most related one to the topics of this proposal is the NSF CAREER 1054127 — Toward Discovering the 3D Geometrical and Semantic Structure of Objects and Scenes (January 2011 to January 2015) ($500,000).

Intellectual merit: The main objective of this project is a theoretical framework for jointly understanding the 3D spatial and semantic structure of complex scenes from images. This project has resulted so far in several publications in top-tier computer vision conferences and journals. Under this grant Savarese developed new techniques for 3D object detection and tracking from images [96, 97, 4], a new formulation for joint recognition and reconstruction [5, 25, 3, 2, 2], new models for scene segmentation from RGB and RGB-D imagery [82, 76], and a new benchmark for evaluating 3D object pose and shape recovery the 3DPASCAL+ dataset [95].

Broader Impact: The novel paradigm for joint object detection and scene reconstruction has started a new area of research in computer vision related to semantic reconstruction. Two students completed their PhD with significant support from this grant (Y. Bao, M. Sun) and are now research scientist at Magic Leap and on faculty at National Tsing Hua University, respectively.

Professor Xiao has recently received his first grant from the NSF through the NSF/Intel Partnership program on Visual and Experiential Computing (VEC) for “VEC: Small: Collaborative Research: Scene Understanding from RGB-D Images” (IIS-1562763, $960,000, 2015-2018). It is a joint project with Professors Funkhouser (also from Princeton), A. Efros, J. Malik (from UC Berkeley).

Intellectual merit: Xiao has developed a 3D deep learning algorithm for detecting objects from RGB-D images that significantly outperforms the state-of-the-art, becoming the best 3D object detectors for many applications in robotics, vision, and graphics. This algorithm is currently trained on a few hundred RGB-D images captured from the real world, and the small size of training set limits the performance of learning powerful features. The CAD models collected from this proposal would be a natural choice for increasing the size of training set by several orders of magnitudes. Since the size of training data is crucial for 3D deep learning algorithms, we expect to see a large performance benefit using ShapeNet.

Broader Impact: The project has already had an impact on RGB-D object recognition and scene understanding. It has drawn attention from the community to put more emphasis on “3D model object detection” in RGB-D images, to support many common applications that require reasoning grounded in the real 3D world, such as robot manipulation and semantic mapping. Our novel paradigm for 3D recognition has also created “3D deep learning” as a new area of study within vision and robotics.

Professor Huang has been supported by a related NSF grant DMS-1521583, *Joint Analysis of Correlated Data*, $250,000, September 15, 2015 to August 31, 2018 (joint w. Prof. L. Guibas). The main objective of this project is to study how to enable collaborative investigations based on large-scale shared annotations on data. The output of this project will provide new tools for organizing models on ShapeNet and facilitating their use. This brand new project has resulted in a RECOMB’2016 paper on multiple alignment of graphs (e.g., protein-protein interaction networks).

References

[1] Nike Arnold and Lara Ducate. *Present and future promises of CALL: From theory and research to new directions in language teaching*. Computer Assisted Language Instruction Consortium, 2011.

[2] Sid Yingze Bao, Mohit Bagra, Yu-Wei Chao, and Silvio Savarese. Semantic structure from motion with points, regions, and objects. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2012.

[3] Sid Yingze Bao, Axel Furlan, Li Fei-Fei, and Silvio Savarese. Understanding the 3D layout of a cluttered room from multiple images. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2014.

[4] Sid Yingze Bao, Yu Xiang, and Silvio Savarese. Object co-detection. In *Proc. of European Conference of Computer Vision*, 2012.

[5] Yingze Bao, Manmohan chandraker, Yuanqing Lin, and Silvio Savarese. Dense object reconstruction using semantic priors. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2013.

[6] Mirela Ben-Chen, Adrian Butscher, Justin Solomon, and Leonidas Guibas. On discrete killing vector fields and patterns on surfaces. *Computer Graphics Forum*, 29(5):1701–1711, 2010.

[7] H.M. Berman, J. Westbrook, J. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. The protein data bank. *Nucleic Acids Research*, 28:235–242, 2000.

[8] Doug M. Boyer, Yaron Lipman, Elizabeth St. Clair, Jesus Puente, Biren A. Patel, Thomas Funkhouser, Jukka Jernvall, and Ingrid Daubechies. Algorithms to automatically quantify the geometric similarity of anatomical surfaces. *PNAS*, October 2011.

[9] Aleksey Boyko and Thomas Funkhouser. Cheaper by the dozen: Group annotation of 3D data. In *ACM Symposium on User Interface Software and Technology (UIST)*, October 2014.

[10] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. ShapeNet: An information-rich 3D model repository. Technical Report arXiv:1512.03012 [cs.GR], Stanford University — Princeton University — Toyota Technological Institute at Chicago, 2015.

[11] Angel X Chang, Manolis Savva, and Christopher 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.

[12] Siddhartha Chaudhuri, Evangelos Kalogerakis, Stephen Giguere, and Thomas Funkhouser. AttribIt: content creation with semantic attributes. In *ACM Symposium on User Interface Software and Technology (UIST)*, October 2013.

[13] Siddhartha Chaudhuri, Evangelos Kalogerakis, Leonidas Guibas, and Vladlen Koltun. Probabilistic reasoning for assembly-based 3D modeling. In *ACM Transactions on Graphics (TOG)*, volume 30, page 35. ACM, 2011.

[14] F. Chazal, D. Cohen-Steiner, L. J. Guibas, F. Mémoli, and S. Y. Oudot. Gromov-Hausdorff stable signatures for shapes using persistence. *Computer Graphics Forum (proc. SGP 2009)*, pages 1393–1403, 2009.

[15] Ding-Yun Chen, Xiao-Pei Tian, Yu-Te Shen, and Ming Ouhyoung. On visual similarity based 3D model retrieval. *Computer graphics forum*, 22(3):223–232, 2003.

[16] Xiaobai Chen, Aleksey Golovinskiy, and Thomas Funkhouser. A benchmark for 3D mesh segmentation. *ACM Trans. Graph.*, 28(3):73:1–73:12, July 2009.

[17] Xiaobai Chen, Abulhair Saparov, Bill Pang, and Thomas Funkhouser. Schelling points on 3D surface meshes. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, August 2012.

[18] Forrester Cole, Kevin Sanik, Doug DeCarlo, Adam Finkelstein, Thomas Funkhouser, Szymon Rusinkiewicz, and Manish Singh. How well do line drawings depict shape? *ACM Transactions on Graphics (Proc. SIGGRAPH)*, August 2009.

[19] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition (CVPR), 2009*, 2009.

[20] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.

[21] B. Falcidieno. Aim@shape. `http://www.aimatshape.net/ontologies/shapes/`, 2005.

[22] Matthew Fisher, Daniel Ritchie, Manolis Savva, Thomas Funkhouser, and Pat Hanrahan. Example-based synthesis of 3D object arrangements. *ACM Transactions on Graphics (TOG)*, 31(6):135, 2012.

[23] Thomas Funkhouser, Patrick Min, Michael Kazhdan, Joyce Chen, Alex Halderman, David Dobkin, and David Jacobs. A search engine for 3D models. *ACM Transactions on Graphics (TOG)*, 22(1):83–105, 2003.

[24] Thomas Funkhouser, Hijung Shin, Corey Toler-Franklin, Antonio Garca Castaeda, Benedict Brown, David Dobkin, Szymon Rusinkiewicz, and Tim Weyrich. Learning how to match fresco fragments. *ACM Journal of Computing and Cultural Heritage*, 4(2), November 2010.

[25] Axel Furlan, David Miller, Domenico G. Sorrenti, Li Fei-Fei, and Silvio Savarese. Free your camera: 3D indoor scene understanding from arbitrary camera motion. In *BMVC*, 2013.

[26] Paul-Louis George. Gamma. `http://www.rocq.inria.fr/gamma/download/download.php`, 2007.

[27] James J Gibson. The theory of affordances. *Hilldale, USA*, 1977.

[28] Aleksey Golovinskiy and Thomas Funkhouser. Randomized cuts for 3D mesh analysis. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, December 2008.

[29] Aleksey Golovinskiy and Thomas Funkhouser. Min-cut based segmentation of point clouds. In *IEEE Workshop on Search in 3D and Video (S3DV) at ICCV*, sep 2009.

[30] Aleksey Golovinskiy, Vladimir G. Kim, and Thomas Funkhouser. Shape-based recognition of 3D point clouds in urban environments. *International Conference on Computer Vision (ICCV)*, September 2009.

[31] Aleksey Golovinskiy, Joshua Podolak, and Thomas Funkhouser. Symmetry-aware mesh processing. *Mathematics of Surfaces 2009 (invited paper)*, LNCS 5654, September 2009.

[32] Harry Heft. The ecological approach to navigation: A gibsonian perspective. In *The construction of cognitive maps*, pages 105–132. Springer, 1996.

[33] Qixing Huang and Leonidas Guibas. Consistent shape maps via semidefinite programming. *Computer Graphics Forum, Proc. Eurographics Symposium on Geometry Processing (SGP)*, 32(5):177–186, 2013.

[34] Qixing Huang, Hao Su, and Leonidas Guibas. Fine-grained semi-supervised labeling of large shape collections. *ACM Transactions on Graphics*, 32:190:1–190:10, 2013.

[35] Qixing Huang, Fan Wang, and Leonidas Guibas. Functional map networks for analyzing and exploring large shape collections. *ACM Trans. Graph.*, 33(4):36:1–36:11, July 2014.

[36] Ashesh Jain, Debarghya Das, and Ashutosh Saxena. PlanIt: A crowdsourcing approach for learning to plan paths from large scale preference feedback. In *ICRA*, 2015.

[37] Subramaniam Jayanti, Yagnanarayanan Kalyanaraman, Natraj Iyer, and Karthik Ramani. Developing an engineering shape benchmark for CAD models. *Computer-Aided Design*, 2006.

[38] Yun Jiang, Hema S. Koppula, and Ashutosh Saxena. Hallucinated humans as the hidden context for labeling 3D scenes. *CVPR*, 2013.

[39] Evangelos Kalogerakis, Siddhartha Chaudhuri, Daphne Koller, and Vladlen Koltun. A probabilistic model for component-based shape synthesis. *ACM Transactions on Graphics (TOG)*, 31:55, 2012.

[40] Andrej Karpathy and Li Fei-Fei. Deep visual-semantic alignments for generating image descriptions. *CVPR*, 2015.

[41] Vladimir Kim, Yaron Lipman, Xiaobai Chen, and Thomas Funkhouser. Mobius transformations for global intrinsic symmetry analysis. *Computer Graphics Forum (Symposium on Geometry Processing)*, July 2010.

[42] Vladimir G. Kim, Wilmot Li, Niloy J. Mitra, Siddhartha Chaudhuri, Stephen DiVerdi, and Thomas Funkhouser. Learning part-based templates from large collections of 3D shapes. *ACM Trans. Graph.*, 32(4):70:1–70:12, July 2013.

[43] Vladimir G. Kim, Wilmot Li, Niloy J. Mitra, Stephen DiVerdi, and Thomas Funkhouser. Exploring collections of 3D models using fuzzy correspondences. *ACM Trans. Graph.*, 31(4):54:1–54:11, July 2012.

[44] Vladimir G. Kim, Yaron Lipman, and Thomas Funkhouser. Blended intrinsic maps. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, August 2011.

[45] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3D object representations for fine-grained categorization. In *4th International IEEE Workshop on 3D Representation and Recognition (3dRR-13)*, Sydney, Australia, 2013.

[46] Chien-Fu Kuo and Chih-Hsing Chu. An online ergonomic evaluator for 3d product design. *Computers in industry*, 56(5):479–492, 2005.

[47] R.A. Laskowski, E.G. Hutchinson, A.D. Michie, A.C. Wallace, M.L. Jones, and J.M. Thornton. Pdbsum: A web-based database of summaries and analyses of all pdb structures. *Trends Biochem. Sci.*, 22:488–490, 1997.

[48] B. Li, A. Godil, M. Aono, X. Bai, T. Furuya, L. Li, R. López-Sastre, H. Johan, R. Ohbuchi, C. Redondo-Cabrera, A. Tatsuma, T. Yanagimachi, and S. Zhang. SHREC'12 track: Generic 3D shape retrieval. In *Proceedings of the 5th Eurographics Workshop on 3D Object Retrieval*, 2012.

[49] B. Li, Y. Lu, C. Li, A. Godil, T. Schreck, M. Aono, Q. Chen, N.K. Chowdhury, B. Fang, T. Furuya, T. Johan, R. Kosaka, H. Koyanagi, R. Ohbuchi, and A. Tasuma. SHREC14 track: Large scale comprehensive 3D shape retrieval. In *Eurographics Workshop on 3D Object Retrieval*, 2014.

[50] Yangyan Li, Hao Su, Charles Ruizhongtai Qi, Noa Fish, Daniel Cohen-Or, and Leonidas J Guibas. Joint embeddings of shapes and images via cnn image purification. *ACM Transactions on Graphics (TOG)*, 34(6):234, 2015.

[51] Joerg Liebelt and Cordelia 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.

[52] Yaron Lipman, Xiaobai Chen, Ingrid Daubechies, and Thomas Funkhouser. Symmetry factored embedding and distance. *ACM Trans Graphics (Proc SIGGRAPH)*, 2010.

[53] Yaron Lipman and Thomas Funkhouser. Mobius voting for surface correspondence. *ACM Transactions on Graphics (SIGGRAPH 2009)*, 28(3), August 2009.

[54] Tianqiang Liu, Siddhartha Chaudhuri, Vladimir G. Kim, Qi-Xing Huang, Niloy J. Mitra, and Thomas Funkhouser. Creating consistent scene graphs using a probabilistic grammar. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, December 2014.

[55] Tianqiang Liu, Vladimir G. Kim, and Thomas Funkhouser. Finding surface correspondences using symmetry axis curves. *Computer Graphics Forum (Proc. Symposium on Geometry Processing)*, July 2012.

[56] Mitchell P Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. Building a large annotated corpus of english: The penn treebank. *Computational linguistics*, 19, 1993.

[57] Francisco Massa, Bryan Russell, and Mathieu Aubry. Deep exemplar 2d-3d detection by adapting from real to rendered views. *arXiv preprint arXiv:1512.02497*, 2015.

[58] Cynthia Matuszek*, Nicholas FitzGerald*, Luke Zettlemoyer, Liefeng Bo, and Dieter Fox. A Joint Model of Language and Perception for Grounded Attribute Learning. In *Proc. of the 2012 International Conference on Machine Learning*, Edinburgh, Scotland, June 2012.

[59] Nikolaus Mayer, Eddy Ilg, Philip Häusser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. *arXiv preprint arXiv:1512.02134*, 2015.

[60] Quentin Mérigot, Maks Ovsjanikov, and Leonidas Guibas. Robust Voronoi-based curvature and feature estimation. In *SIAM/ACM Joint Conference on Geometric and Physical Modeling*, 2009.

[61] George A. Miller. WordNet: a lexical database for English. *CACM*, 1995.

[62] Roland Angst Minhyuk Sung, Vladimir G. Kim and Leonidas Guibas. Data-driven structural priors for shape completion. *ACM Transactions on Graphics (TOG)*, 2015.

[63] Robert Osada, Thomas Funkhouser, Bernard Chazelle, and David Dobkin. Shape distributions. *ACM Trans. Graph.*, 21(4):807–832, October 2002.

[64] Maks Ovsjanikov, Mirela Ben-Chen, Justin Solomon, Adrian Butscher, and Leonidas Guibas. Functional maps: A flexible representation of maps between shapes. *ACM Transactions on Graphics*, 31(4), 2012.

[65] Maks Ovsjanikov, Quentin Mérigot, Facundo Mémoli, and Leonidas Guibas. One point isometric matching with the heat kernel. *CGF*, 29(5):1555–1564, 2010.

[66] Brandon C Roy, Soroush Vosoughi, and Deb Roy. Grounding language models in spatiotemporal context. In *International Speech Communication Association*, 2014.

[67] B.C. Russell and A. Torralba. Building a database of 3D scenes from user annotations. In *Computer Vision and Pattern Recognition (CVPR), 2009 IEEE Conference on*, 2009.

[68] Raif M. Rustamov, Maks Ovsjanikov, Omri Azencot, Mirela Ben-Chen, Frédéric Chazal, and Leonidas Guibas. Map-based exploration of intrinsic shape differences and variability. *ACM Trans. Graph.*, 32(4):72:1–72:12, July 2013.

[69] Adrian Secord, Jingwan Lu, Adam Finkelstein, Manish Singh, and Andrew Nealen. Perceptual models of viewpoint preference. *ACM Transactions on Graphics*, 30(5), October 2011.

[70] Philip Shilane and Thomas Funkhouser. Distinctive regions of 3D surfaces. *ACM Transactions on Graphics*, 26(2), June 2007.

[71] Philip Shilane, Patrick Min, Michael Kazhdan, and Thomas Funkhouser. The Princeton shape benchmark. In *Shape Modeling Applications, 2004. Proceedings*. IEEE, 2004.

[72] P. Skraba, M. Ovsjanikov, F. Chazal, and L. Guibas. Persistence-based segmentation of deformable shapes. In *CVPR Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment*, page to appear, June 2010.

[73] Justin Solomon, Mirela Ben-Chen, Adrian Butscher, and Leonidas Guibas. Discovery of intrinsic primitives on triangle meshes. In *Proc. Eurographics 2011*, 2011.

[74] Shuran Song, Samuel Lichtenberg, and Jianxiong Xiao. Sun rgb-d: A rgb-d scene understanding benchmark suite. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2015.

[75] Shuran Song and Jianxiong Xiao. Sliding shapes for 3D object detection in depth images. In *Proc. of European Conference of Computer Vision*, 2014.

[76] Byung soo Kim, Pushmeet Kohli, and Silvio Savarese. 3D scene understanding by voxel-crf. In *Proceedings of the International Conference on Computer Vision*, 2013.

[77] Hao Su, Qixing Huang, Niloy Mitra, Yangyan Li, and Leonidas Guibas. Estimating image depth using shape collections. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 2014.

[78] Hao Su, Charles R Qi, Yangyan Li, and Leonidas Guibas. Render for CNN: Viewpoint estimation in images using CNNs trained with rendered 3D model views. *International Conference on Computer Vision (ICCV)*, 2015.

[79] Baochen Sun, Xingchao Peng, and Kate Saenko. Generating large scale image datasets from 3d cad models. In *Workshop on the Future of Datasets in Vision in Computer Vision and Pattern Recognition*, 2015.

[80] Jian Sun, Xiaobai Chen, and Thomas Funkhouser. Fuzzy geodesics and consistent sparse correspondences for deformable shapes. *Computer Graphics Forum (Symposium on Geometry Processing)*, July 2010.

[81] Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. A concise and provably informative multi-scale signature based on heat diffusion. *Computer Graphics Forum (Proceedings of the Eurographics Symposium on Geometry Processing)*, 28(5):1383–1392, 2009.

[82] Min Sun, Byung-soo Kim, Pushmeet Kohli, and Silvio Savarese. Relating things and stuff via object property interactions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(7), July 2014.

[83] Tien-Lung Sun, Wen-Yang Feng, and Chin-Jung Chao. Dynamic generation of human-populated vr models for workspace ergonomic evaluation. In *Digital Human Modeling*, pages 979–987. Springer, 2007.

[84] Jaeyong Sung, Seok H Jin, and Ashutosh Saxena. Robobarista: Object part based transfer of manipulation trajectories from crowd-sourcing in 3D pointclouds. In *International Symposium on Robotics Research (ISRR)*, 2015.

[85] Atsushi Tatsuma, Hitoshi Koyanagi, and Masaki Aono. A large-scale shape benchmark for 3D object retrieval: Toyohashi shape benchmark. In *Proceedings of APSIPA2012 (Asia Pacific Signal and Information Processing Association)*, 2012.

[86] Stefanie Tellex, Pratiksha Thaker, Joshua Joseph, and Nicholas Roy. Learning perceptually grounded word meanings from unaligned parallel data. *Machine Learning*, 94(2):151–167, 2014.

[87] Trimble. Trimble 3D warehouse. `https://3dwarehouse.sketchup.com/`, 2012. Accessed: 2015-12-13.

[88] Oliver van Kaick, Hao Zhang, Ghassan Hamarneh, and Daniel Cohen-Or. A survey on shape correspondence. *Comput. Graph. Forum*, 30(6):1681–1707, 2011.

[89] RC Veltkamp and FB ter Harr. SHREC 2007 3D shape retrieval contest. Technical report, Utrecht University, Dept of Info and Comp. Sci, Technical Report UU-CS-2007-015, 2007.

[90] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. *CVPR*, 2015.

[91] Dejan V Vranić. 3D model retrieval. *University of Leipzig, Germany, PhD thesis*, 2004.

[92] Matthew R Walter, Sachithra Hemachandra, Bianca Homberg, Stefanie Tellex, and Seth Teller. A framework for learning semantic maps from grounded natural language descriptions. *The International Journal of Robotics Research*, 33(9):1167–1190, 2014.

[93] Raoul Wessel, Ina Blümel, and Reinhard Klein. A 3D shape benchmark for retrieval and automatic classification of architectural data. In *Eurographics 2009 Workshop on 3D Object Retrieval*, pages 53–56. The Eurographics Association, 2009.

[94] Zhirong Wu, Shuran Song, Aditya Khosla, Xiaoou Tang, and Jianxiong Xiao. 3D shapenets for 2.5D object recognition and next-best-view prediction. *CoRR*, abs/1406.5670, 2014.

[95] Yu Xiang, Roozbeh Mottaghi, and Silvio Savarese. Beyond PASCAL: A benchmark for 3D object detection in the wild. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2014.

[96] Yu Xiang and Silvio Savarese. Object detection by 3D aspectlets and occlusion reasoning. In *4th International IEEE Workshop on 3D Representation and Recognition*, 2013.

[97] Yu Xiang, Changkyu Song, Roozbeh Mottaghi, and Silvio Savarese. Monocular multiview object tracking with 3D aspect parts. In *European Conference on Computer Vision (ECCV)*, 2014.

[98] Jianxiong Xiao, Andrew Owens, and Antonio Torralba. SUN3D: A database of big spaces reconstructed using sfm and object labels. In *IEEE International Conference on Computer Vision, ICCV 2013, Sydney, Australia, December 1-8, 2013*, pages 1625–1632, 2013.

[99] Yobi3D. Yobi3D - free 3D model search engine. `https://www.yobi3d.com/`. Accessed: 2015-12-13.

[100] Juan Zhang, Kaleem Siddiqi, Diego Macrini, Ali Shokoufandeh, and Sven Dickinson. Retrieving articulated 3D models using medial surfaces and their graph spectra. In *Energy minimization methods in computer vision and pattern recognition*, 2005.

[101] Bo Zheng, Yibiao Zhao, Joey C Yu, Katsushi Ikeuchi, and Song-Chun Zhu. Detecting potential falling objects by inferring human action and natural disturbance. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 3417–3424. IEEE, 2014.