Introduction and background

The iconic Stanford bunny, now 23 years old, has been melted, shattered, and deformed countless times. While mostly a fun subculture, “bunny torture” is also a legacy of an earlier time when few interesting and free 3D models existed. Testing on such standard models persists despite well-known limitations. As Greg Turk, originator of the bunny, advises, “I actually consider the bunny to be too good as a test model. It is fairly smooth, it has manifold connectivity, and it isn’t too complex” [Turk 2000].

Oversimplified testing provides a false sense of robustness and causes not only visual artifacts in computer graphics applications, but also fabrication and functionality artifacts when processing geometry intended for 3D printing. Fortunately, 3D models are now abundant. Modern consumer-level 3D printing technologies nurture new communities of professional and amateur 3D modelers, who share and sell 3D-printable models online (e.g., shapeways.com, sketchfab.com, thingiverse.com). This wealth of data also echoes the demand for state-of-the-art processing techniques and automation within 3D printing pipelines.

However, testing remains inadequate. Existing datasets contain only sanitized models (e.g., [Aim@Shape 2004; Levoy et al. 2005; Myles et al. 2014]) or draw from populations containing raw models not specifically intended for printing (rather, e.g., for shape classification [Shilane et al. 2004; Chang et al. 2015] or scene understanding [Nathan Silberman & Fergus 2012; Choi et al. 2016]).

In this paper, we will show that the characteristics and issues common to 3D printing models are distinct from models intended for visualization. As such, validating geometry processing techniques related to 3D printing requires a new representative dataset. This ideal dataset should encompass the different contextual and geometric characteristics of commonly printed shapes. Characteristics common to 3D printing models should appear with proportional distributions, and characteristics inconsistent with models intended for fabrication should be infrequent (e.g., the open boundaries of a video game character’s clothing).

We propose a dataset of 10,000 models culled from a popular shape repository for 3D printing enthusiasts, thingiverse.com. Hereon, we refer to our dataset as Thingi10K. Beyond collecting tags and class information available online, we analyze geometric characteristics of each model (e.g., manifoldness, lack of self-intersections, genus). We contrast these statistics against existing large datasets and investigate correlations within the data.

Existing datasets. Myles et al. collect 116 models from academic sources (Stanford Scanning Repository [Levoy et al. 2005] and Aim@Shape Repository [Aim@Shape 2004]) to test their parameterization algorithm [Myles et al. 2014]. These models correspond to best-case input due to their extreme cleanliness and general position assumption (i.e., no four points on a circle, no coplanar intersections, etc.). For 3D printing models in the wild, degeneracies, non-manifoldness and self-intersections are abundant, not special cases. Structured modeling and coordinate quantization tends to break rather than fulfill general position assumptions.

Figure 1: The Thingi10K dataset contains 10,000 models from featured “things” on thingiverse.com, a popular online repository.

Abstract

Empirically validating new 3D-printing related algorithms and implementations requires testing data representative of inputs encountered in the wild. An ideal benchmarking dataset should not only draw from the same distribution of shapes people print in terms of class (e.g., toys, mechanisms, jewelry), representation type (e.g., triangle soup meshes) and complexity (e.g., number of facets), but should also capture problems and artifacts endemic to 3D printing models (e.g., self-intersections, non-manifoldness). We observe that the contextual and geometric characteristics of 3D printing models differ significantly from those used for computer graphics applications, not to mention standard models (e.g., Stanford bunny, Armadillo, Fertility). We present a new dataset of 10,000 models collected from an online 3D printing model-sharing database. Via analysis of both geometric (e.g., triangle aspect ratios, manifoldness) and contextual (e.g., licenses, tags, classes) characteristics, we demonstrate that this dataset represents a more concise summary of real-world models used for 3D printing compared to existing datasets. To facilitate future research endeavors, we also present an online query interface to select subsets of the dataset according to project-specific characteristics. The complete dataset and per-model statistical data are freely available to the public.

The Thingi10K dataset contains 10,000 models from featured “things” on thingiverse.com, a popular online repository.
collected contextual information and computed geometric properties of our dataset in detail against two existing datasets: MPZ14 and ShapeNetCore. We demonstrate that these represent two extreme cases in terms geometric quality while our dataset provides a mixture of geometric qualities reflecting real-world settings. All data and analysis of our dataset are freely available to the public. To facilitate exploration and future reuse, we provide an easy-to-use online query interface (see Figure 2). This interface augments the Thingiverse front-end with our geometric analysis of each model.

![Figure 2: Our online query interface selects subsets of Thingi10K.](image)

In addition to generic datasets, a variety of specialized datasets exist. For example, Lim et al. provide 219 IKEA 3D models for pose-estimation [Lim et al. 2013]. Recently, Choi et al. released a dataset of 10,000 scanned objects, with a subset of 383 successfully reconstructed 3D models [Choi et al. 2016]. The Shape Retrieval Contest releases multiple datasets each year to test retrieval algorithms including generic [Bronstein et al. 2010b; Li et al. 2012], non-rigid humans [Pickup et al. 2014], sketch-based shapes [Li et al. 2013; Li et al. 2014], shape correspondences [Bronstein et al. 2010a], facial expressions [Nair & Cavallaro 2008; Veltkamp et al. 2011], and range scans [Dutagaci et al. 2010]. Our Thingi10K dataset complements these sources by providing a specialized dataset for 3D printing objects.

We are not the first to utilize Thingiverse models for academic purposes. To test a rapid prototyping interface, Mueller et al. consider Thingiverse models, but report that meshing artifacts required manual cleanup before processing [Mueller et al. 2014]. Beyer et al. procedurally collect 2,250 models with specific tags from Thingiverse to test a decomposition algorithm [Beyer et al. 2015]. Buehler et al. manually sift through 25,000 models from search results on Thingiverse to identify 363 models as “assistive technologies” [Buehler et al. 2015]. Beyond testing a specific routine, these works do not analyze low-level geometric characteristics of the collected models. These collected datasets are also not publicly available.

**Contributions.** Unlike previous datasets, our Thingi10K dataset reflects the variety, complexity and (lack of) quality of 3D printing models. It is immediately useful for testing the performance of methods for structural analysis [Stava et al. 2012; Zhou et al. 2013; Umetani & Schmidt 2013], shape optimization [Prévost et al. 2013; Bächler et al. 2014; Musialski et al. 2015], or solid geometry operations [Zhou et al. 2016]. Due to its specialized nature and correlated contextual information, we suspect the dataset is also useful for machine learning and data mining algorithms. We compare the

![Figure 3: Percentile plots of vertex and component count.](image)

## 2 Methodology

Instead of our hiring professional modelers or scanning physical objects, we leverage the availability of 3D models hosted and shared online. Among all 3D shape repositories, we select Thingiverse for its large and active user community, its vast collection of print-validated designs, and its restriction to open-source licenses.

As one of the largest online shape repositories, Thingiverse hosts more than a million user-uploaded things, 3D designs consisting of one or more 3D models (i.e., one or more mesh files). As of October 2015, Thingiverse has more than 2 million active users, with 30-40 uploads each week and 1.7 million downloads per month [MakerBot 2015]. Thanks to this community, a design is typically not only modeled virtually but also fabricated by one or more users, which provides invaluable real-world validations.

Our Thingi10K dataset consists of 10,000 models (from 2011 things) systematically culled from Thingiverse via web crawling. Rather than randomly sample the entire repository, which may contain bogus models uploaded by inexperienced users or for testing purposes, we focus on things *featured* on Thingiverse. Featured things are entirely and independently selected by Thingiverse staff based on their design, beauty and manufacturability. In a sense, these 10,000 models represent a subset of the top-quality designs on Thingiverse. Thingi10K contains every 3D model of every thing featured by Thingiverse between Sept. 16, 2009 and Nov. 15, 2015.

## 3 Analysis

The 10,000-model dataset comes from 2,011 unique things designed by 1,083 unique users, covering a large variety. Nearly all models are stored as .stl files (9,956); the rest are .obj (42), .ply (1), and .off (1). We analyze both geometric and contextual information of our dataset to illustrate its representational quality and diversity.
3.1 Geometry information

We analyze a variety of mesh complexity and quality measures on our dataset of 3D printing models and compare with two existing datasets: MPZ14 (116 models) and ShapeNetCore (2000 models uniformly sampled from 57,459).

3.1.1 Complexity

Complexity of 3D model does not directly correlate with 3D printing cost. We evaluated three different measures to quantify the complexity of our dataset: number of vertices, number of disconnected components and genus.

Figure 3 provides the percentile plot of both vertex and component count over each dataset. The vertex count plot indicates that the MPZ14 dataset favors moderately high resolution models and excludes extremely low or high resolution models. On the other hand, the distribution over our dataset and ShapeNetCore is similar, with our dataset covering a larger range. Figure 4 illustrates the highest resolution model of each dataset.

Many geometry processing algorithms assume input will be processed one component at a time, so it is not a surprise that MPZ14 contains exclusively single-component models. This assumption is not valid in the context of 3D printing, where multiple components could overlap to form a larger shape. Analysing each component separately may lead to incorrect results. Within our dataset 29% of models have more than one component. ShapeNetCore is 83% multi-component, but close inspection finds many models are composed of incoherent patches or isolated faces (see Figure 5). In contrast, 3D printing models with high numbers of components in Thingi10K are typically by design, with the base shape naturally decomposing into smaller components.

The genus distribution of our Thingi10K dataset is similar to MPZ14, but our dataset covers a larger range of genus, with the highest genus over 60 times larger than in MPZ14 (see Figure 6). To avoid confusion, we limit the genus comparison to single-component, closed and manifold meshes. Zero of the ShapeNetCore models meet this criteria.

3.1.2 Mesh quality

Mesh qualities of a dataset play a major role in determining its usability and representation of models in the wild. For example, degenerate or sliver triangles will cause poor accuracy in non-robust finite element simulations, and fragile volumetric meshing routines will fail in the presence of self-intersections. It is crucial to understand the mesh quality of real-world input data in order to design robust and practical algorithms. Existing datasets often focus on high-level properties and provide little insight on their mesh qualities. Our analysis aims to fill this gap.

We analyze 13 mesh quality measurements:

- Closed: Every edge is adjacent to 2 or more faces.
- Orientable
- No degenerated faces
- No coplanar intersection
- No self-intersection
- Solid
- No isolated vertices
- No duplicate faces
- No self-intersection
- No coplanar intersection
- No degenerated faces
- Edge manifold
- Vertex manifold
- PWN
- Orientable
- Closed

Figure 7: MPZ14 models are “too clean,” whereas ShapeNetCore are unrealistically corrupted in the context of 3D printing.
No isolated vertices: All vertices are adjacent to at least one face.
No duplicated faces: There does not exist a pair of faces sharing the same set of vertices.
Vertex-manifold: The one-ring neighborhood of every vertex is a topological disc.
Edge-manifold: Every non-boundary edge must be incident to exactly two faces.
No degeneracy: All faces must have non-collinear vertices. Degeneracy can be checked with exact predicates [Shewchuk 1997].
No self-intersection: The intersection of any two faces is either empty, a shared vertex, or a shared edge. Exact predicates are necessary to ensure correctness.
No coplanar intersections: No two faces are coplanar and overlapping. This is a strictly weaker condition than “no self-intersection.”
Piecewise-constant winding number (PWN): The winding number field at any non-mesh point is piece-wise constant ([Zhou et al. 2016]).
Solid: The input mesh must be a valid boundary of a subspace of $\mathbb{R}^3$. Specifically, it must be PWN, self-intersection free and induce a $\{0,1\}$ winding number field.
Aspect ratio: The aspect ratio of a triangle is the ratio of its circumradius to the diameter of its incircle.
Intrinsically Delaunay: All edges must have non-negative cotangent weights [Fisher et al. 2007].

These mesh quality measures are not by no means complete. Additional quality measures ([Shewchuk 2002; Attene 2013]) can be easily adopted.

MPZ14 has perfect mesh quality according to seven different measures. In particular, all models are manifold, oriented and degeneracy-free. Because many geometry processing algorithms do not require the input model to be closed or self-intersection free, data from MPZ14 are perfect as proof-of-concept examples. However, their high quality is due to the fact that models were selected not on merits of their shape, functionality or aesthetics, but rather because they meet certain quality criteria or have been sanitized.

On the other hand, ShapeNetCore has very poor mesh quality according to 6 measures in Figure 7. Its maximum and average triangle aspect ratios are visibly worse than Thingi10K and MPZ14. This is partially due to the fact that these data are collected directly from the internet, where models were not necessarily designed for fabrication purposes. Many existing learning algorithms sidestep the quality issues by transforming boundary representations to depth images or bounding box hierarchies [Hu 2012]. Performing geometry processing algorithms directly on these models is very hard due to poor mesh quality.

In contrast, our dataset offers a curated collection of 3D meshes with a large range of mesh qualities. It contains a significant number of high quality models as well as a non-negligible proportion of models with common mesh quality problems. Due to its large quantity, our dataset is ideal for stress-testing purposes where one can easily select a subset of the data that matches any combination of mesh criteria (Section 4). Because all data are sampled from real-world models designed to be 3D printed, our dataset provides an unbiased view of the mesh qualities used in practice. Our analysis could be used to gauge the restrictions posed by various assumptions on mesh quality. For example, an algorithm assuming self-intersection-free input would automatically exclude 45% of inputs, which may not be acceptable in a real-world settings.

3.2 Contextual information

Each thing in our dataset is annotated by its original designer. Thingiverse supports three types of annotations: category, subcategory and tags. The first two must be selected from a predefined list of categories, and the last one is a set of free-form texts created by the user. A total of 4892 distinct tags are used in our dataset. Figure 10 illustrates the most frequently used tags.

Unlike ShapeNet [Chang et al. 2015], which focuses on providing categorical annotations specific to object classification purposes, our dataset comes with a rich and diverse set of original tags ranging from the semantics of a 3D model to the printer/material used for fabrication. For example, Figure 9 shows all models with tags math, sculpture and scan.

When combined with geometric analysis, our annotations reveal interesting insights unavailable from previous works. For example, a simple frequency analysis indicates OpenSCAD is the most popular modeling tool used by Thingiverse users. Our dataset shows that 98% of OpenSCAD models are closed, while only 91% of SketchUp models and 85% of TinkerCAD models are closed.

Furthermore, due its fabrication-focused nature, many uploaded meshes are “print-ready” in the sense that their orientation and decompositions are designed for optimal printing outcome (See figure 11). Recent papers have tried to solve problems such as decomposing a large model to fit in the print volume and finding ideal print orientations [Chen et al. 2015]. The Thingi10K models, by their intrinsic nature of being successful prints, represent ground truth data.
auto-generated python script to batch download results for custom search terms.

Users of our online query interface can view all contextual and geometry model details (Figure 13). In particular, we respect the copyright of each model. On the model detail page, we clearly indicate the original author and open source licence of each model. We also provide links to the original Thingiverse pages where the raw data can be obtained.

To demonstrate the power of our online query interface, Figure 14 shows some interesting search results and the query used.

5 Conclusion

In this work, we present a large-scale annotated 3D dataset based on models used in 3D printing applications. Our dataset consists of 10,000 meshes crawled systematically from Thingiverse. We analyze both the contextual and geometric information of our dataset and compare with two existing 3D model datasets. Our analysis shows our data covers a large range of categories and provides a balanced representation of real-world data in terms of mesh complexity and quality. The entire dataset and our analysis are freely available to the public, and we provide a query interface to facilitate the exploration and dissection of our dataset.

Our dataset could be used as input for stress-testing purposes as well as ground truth for learning algorithms. As for future work, we plan to update and increase the size of the dataset over time to reflect the fast-evolving nature of the 3D printing community. Specifically, we would like to include all featured things from Thingiverse and add support for users to suggest additional models for inclusion. We hope our dataset and the accompanying analysis provide an informative summary of 3D printing models and clarify the requirements for geometry processing algorithms to be robust.
Acknowledgments

We thank M. Campen, C. Tymms, and J. Panetta for early feedback and proofreading. Funded in part by NSF grants CMMI-11-29917, IIS-14-09286, and IIS-17257.

References

AIM@SHAPE, 2004. Aim@shape digital shape workbench v5.0 - shape repository. http://visionair.ge.imati.cnr.it.

ATTENE, M. 2013. Surface mesh qualities.

BACHER, M., WHITING, E., BICKEL, B., AND SORKINE-HORNUNG, O. 2014. Spin-it: Optimizing moment of inertia for spinnable objects. ACM Trans. Graph.

BEYER, D., GUREVICH, S., MUELLER, S., CHEN, H.-T., AND BAUDISCH, P. 2015. Platener: Low-fidelity fabrication of 3d objects by substituting 3d print with laser-cut plates. In Proc. SIGCHI.

BROSTEIN, A., BROSTEIN, M., CASTELLANI, U., DUBROVINA, A., GUIBAS, L., HORAUD, R., KIMMEL, R., KNOSOW, D., VON LAVANTE, E., MATEUS, D., ET AL. 2010. Shrec 2010: robust correspondence benchmark. In EW3DOR.

BROSTEIN, A., BROSTEIN, M., CASTELLANI, U., FALCI-DIENO, B., FUSIELLO, A., GODIL, A., GUIBAS, L., KOKKINOS, I., LIAN, Z., OVSJANIJKOV, M., ET AL. 2010. Shrec 2010: robust large-scale shape retrieval benchmark. Proc. 3DOR.

BUENLEL, E., BRANHAM, S., ALI, A., CHANG, J. J., HOFFMANN, M. K., HURST, A., AND KANE, S. K. 2015. Sharing is caring: Assistive technology designs on thingiverse. In Proc. SIGCHI.

CHANG, A. X., FUNKHouser, T., GUIBAS, L., HANKAHAN, P., HUANG, Q., LI, Z., SAVARESE, S., SAVVA, M., SONG, S., SU, H., XIAO, J., YI, L., AND YU, F. 2015. ShapeNet: An Information-Rich 3D Model Repository. Tech. Rep. arXiv:1512.03012 [cs.GR].

CHEN, X., ZHANG, H., LIN, J., HU, R., LU, L., HUANG, Q., BENES, B., COHEN-OR, D., AND CHEN, B. 2015. Dapper: Decompose-and-pack for 3d printing. ACM Trans. Graph.

CHOI, S., ZHOU, Q.-Y., MILLER, S., AND KOLTUN, V. 2016. A large dataset of object scans. arXiv:1602.02481.

DUTAGACI, H., GODIL, A., CHEUNG, C. P., FURUYA, T., HILLENBRAND, U., AND OHBUCHI, R. 2010. Shrec’10 track: Range scan retrieval. In 3DOR.

FISHER, M., SPRINGBORN, B., SCHRODER, P., AND BOBENKO, A. I. 2007. An algorithm for the construction of intrinsic delaunay triangulations with applications to digital geometry processing. Computing 81, 2-3, 199–213.

HU, W. 2012. Learning 3d object templates by hierarchical quantization of geometry and appearance spaces. In CVPR.

LEVOY, M., GERTH, J., CURLESS, B., AND PULL, K., 2005. The stanford 3d scanning repository. http://graphics.stanford.edu/data/3Dscanrep/.

LI, B., GODIL, A., AONO, M., BAI, X., FURUYA, T., LI, L., LOPEZ-SASTRE, R. J., JOHAN, H., OHBUCHI, R., REDONDO-CABRERA, C., ET AL. 2012. Shrec’12 track: Generic 3d shape retrieval. In 3DOR.

LI, B., LU, Y., GODIL, A., SCHRECK, T., AONO, M., JOHAN, H., SAAVEDRA, J. M., AND TASHIRO, S. 2013. Shrec’13 track: large scale sketch-based 3d shape retrieval. In EW3DOR.

LI, B., LU, Y., LI, C., GODIL, A., SCHRECK, T., AONO, M., BURTSCHER, M., FU, H., FURUYA, T., JOHAN, H., ET AL. 2014. Shrec’14 track: extended large scale sketch-based 3d shape retrieval. In EW3DOR.

LIM, J. J., PIRSIAVASH, H., AND TORRALBA, A. 2013. Parsing IKEA Objects: Fine Pose Estimation. ICCV.

MAKERBOT, 2015. Celebrating a maker milestone: 1 million uploads on makerbot’s thingiverse. http://www.makerbot.com/blog/2015/10/29.

MUELLER, S., MOHR, T., GUENTHER, K., FROHINHOFFEN, J., AND BAUDISCH, P. 2014. faBrickation: Fast 3d printing of functional objects by integrating construction kit building blocks. In Proc. SIGCHI.

MUSIALSKI, P., AUZINGER, T., BIRSAK, M., WIMMER, M., AND KOBBIET, L. 2015. Reduced-order shape optimization using offset surfaces. ACM Trans. Graph.

MYLES, A., PIETRONI, N., AND ZORIN, D. 2014. Robust field-aligned global parametrization. ACM Trans. Graph.

NAIR, P., AND CAVALLARO, A. 2008. Shrec’08 entry: registration and retrieval of 3d faces using a point distribution model.

Figure 13: Both contextual and geometric information of each model are available on its model detail page.
Figure 14: Our web interface returns subsets of the Thingi10K dataset via text queries.
NATHAN SILBERMAN, DEREK HOIEM, P. K., AND FERGUS, R.
2012. Indoor segmentation and support inference from rgbd images. In ECCV.

PICKUP, D., SUN, X., ROSIN, P. L., MARTIN, R., CHENG, Z., LIAN, Z., AONO, M., HAMZA, A. B., BRONSTEIN, A., BRONSTEIN, M., ET AL. 2014. ShRec’14 track: Shape retrieval of non-rigid 3d human models. In EW3DOR.

PÉVOST, R., WHITING, E., LEFEBVRE, S., AND SORKINE-HORNUNG, O. 2013. Make it stand: Balancing shapes for 3d fabrication. ACM Trans. Graph.

SHECHUK, J. R. 1997. Adaptive precision floating-point arithmetic and fast robust geometric predicates. Discrete & Computational Geometry 18, 3, 305–363.

SHECHUK, J. 2002. What is a good linear finite element? interpolation, conditioning, anisotropy, and quality measures (preprint).

SHILANE, P., MIN, P., KAZHDAN, M., AND FUNKHOUSE, T. 2004. The princeton shape benchmark. In Shape modeling applications, 2004. Proceedings, IEEE, 167–178.

STAVA, O., VANEK, J., BENES, B., CARR, N., AND MÉCH, R. 2012. Stress relief: Improving structural strength of 3d printable objects. ACM Trans. Graph.

TURK, G. 2000. The stanford bunny
http://www.cc.gatech.edu/~turk/bunny/bunny.html.

UMETANI, N., AND SCHMIDT, R. 2013. Cross-sectional structural analysis for 3d printing optimization. In SIGAsia Technical Briefs.

VELTKAMP, R. C., VAN JOLE, S., DRIRA, H., AMOR, B. B., DAOUDI, M., LI, H., CHEN, L., CLAES, P., SMEETS, D., HERMANS, J., ET AL. 2011. Shrec’11 track: 3d face models retrieval. In 3DOR.

ZHOU, Q., PANETTA, J., AND ZORIN, D. 2013. Worst-case structural analysis. ACM Trans. Graph.

ZHOU, Q., GRINSPUN, E., ZORIN, D., AND JACOBSON, A. 2016. Mesh arrangements for solid geometry. ACM Trans. Graph.