Structural Design Explorer: CAD and Machine Learning Possibilities from SimJEB Dataset

Eamon Whalen, Bryan Ong, Ashley Hartwell
Massachusetts Institute of Technology (MIT)

1 Introduction

This project stems from research done by the Digital Structures group at MIT. Whalen et al. introduces SimJEB [1], a simulated jet engine bracket dataset consisting of a public collection of crowdsourced mechanical bracket designs and high-fidelity structural simulations designed specifically for surrogate modelling. The project aims to address one of the challenges of dataset presented for machine learning, particularly that there is no clear and intuitive way to explore the dataset of interest prior to downloading and processing it, which requires unnecessary time and labor that could be spent elsewhere. In addition, users not familiar with Computer Aided Design CAD software may not intuitively be able to work with the geometric dataset and quickly pull insights from it. Hence, we seek to design an interactive visualization that can aid machine learning experts interested in structural engineering and analysis to do the follows:

1. Extract design insights the large geometric and 3D structural simulated dataset
2. Discover relationships between shape and structural performance prior to download
3. Explore the geometric variety that exists in the crowdsourced design data

2 Related Work

To develop a machine learning model, a sufficiently large data set containing information about the properties or attributes of the objects being studied is required. There have been many publicly available data sets of designs or drawings such as ImageNet [2], MNIST [3], and Quickdraw! [4]. There have been others directly linked to CAD like SketchGraphs [5] which provides over 15 million real-world CAD models and the ABC [6] model dataset for geometric deep learning.

All of the referenced datasets are valuable in providing the data needed for further analyses to take place. However, if one were to view the platforms that house the datasets, one would be surprised to find that they provide little means of understanding relationships or concepts within the data. Further insights could only be obtained after downloading the data, processing it, and analyzing it individually. This entire system seems unnecessarily complicated and time consuming. Hence, our team sought to find ways of improving current ways in which online databases present their data for public use, particularly for machine learning datasets.

3 Methods

Adopting the nested model proposed by Munzner [7], the following section outlines the primary design decisions behind the visualization.

3.1 Operation and Data Type Abstraction

The objectives presented in introduction amount to an exploratory data visualization on the domain of 3D shapes. The shapes are represented as a collection of triangular meshes. Accompanying the 3D shapes is metadata describing the structural performance of each one, which is in tabular form. Each model also has a preview image. Thus the challenge can be abstracted as the presentation of triangular meshes such that a reader can extract patterns regarding their metadata and form.

3.2 Visual Encoding and Interaction Design

There are several ways of presenting complex 3D spatial data, including 3D models, multi-view images, and wireframes. One problem with static images is that the selection of fixed views of the models imposes a bias on the reader. Instead, an interactive 3D window seemed like the most effective way to enable unbiased and organic design exploration. A potential problem with the 3D window is that the reader is only exposed to a single model at a time. The use of preview images as tooltips was added so that readers can make quick comparisons before deciding which model to load into the 3D viewer.

For the metadata, all of the fields are quantitative except for the design category, which is nominal. Furthermore, many of the quantitative fields are strongly correlated, for example, mass and maximum displacement. The scatter plot was selected because it makes trends between pairs of quantitative variables easy to infer. Color was selected to encode the design category mainly because spatial channels were already taken and the design category was considered less important information than the other metadata.

3.3 Algorithm Design

The primary algorithmic challenge was efficiently loading a large collection of images and 3D models. The total memory footprint of the 381 triangular meshes is over 1GB; however, the typical user would likely not need to explore every mesh in the 3D viewer. Rather than load all of the models into memory at once, a data-as-needed approach was used to load models when the user requested them. A similar model was applied to the tooltip images. Note that the (upstream) interaction design decision to create tooltip preview images has the added benefit of simplifying model loading from an algorithmic standpoint: if users are shown a preview of the models then they will likely not have to load as many full 3D models to find what they are looking for.

4 Results

As the data presented from the research is found at the intersection of machine learning, structural design and mechanical engineering, a big purpose of the platform would be to introduce the background and functionality of the data presented. Our visualization starts with a home page that gives a brief blurb purpose and potential usage of the data shown in Figure 1.

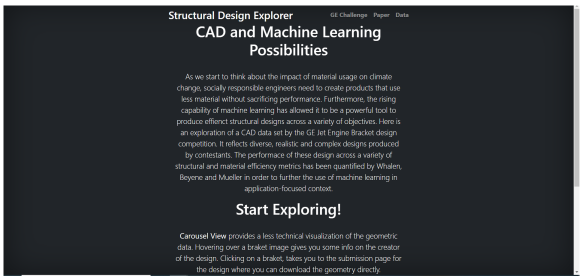


Figure 1: Home page of platform.

Found on the top left corner of the home page are links to the origin of the dataset, the research paper behind it [1], and the SimJEB dataset. The link will lead the user to external web pages housing the corresponding information as shown by Figure 2.

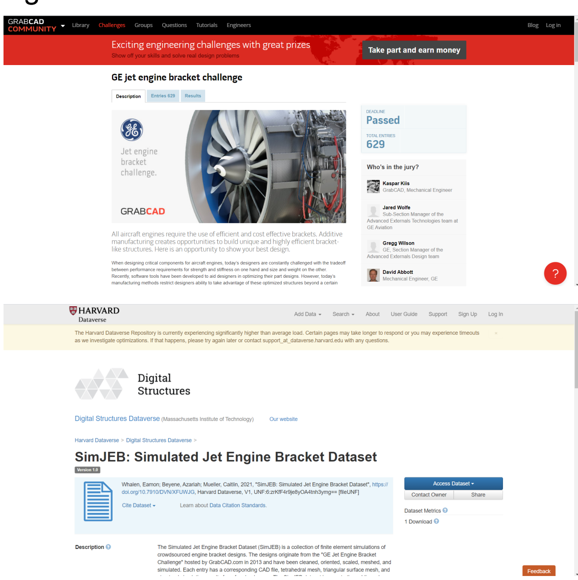


FIgure 2: GE Challenge containing original dataset and Harvard Dataverse containing SimJEB dataset.

To allow users to both extract design insights of the large dataset, as well as explore the geometric variety that exists in the crowdsourced design dataset, we decided to give users two different options of viewing types: Carousel View and Data View. The Carousel View provides a less technical overview of the geometric data available to allow users to familiarize themselves with the type of models present. The Carousel View is also presented to give credit to individual designers of the presented bracket designs shown in the data.

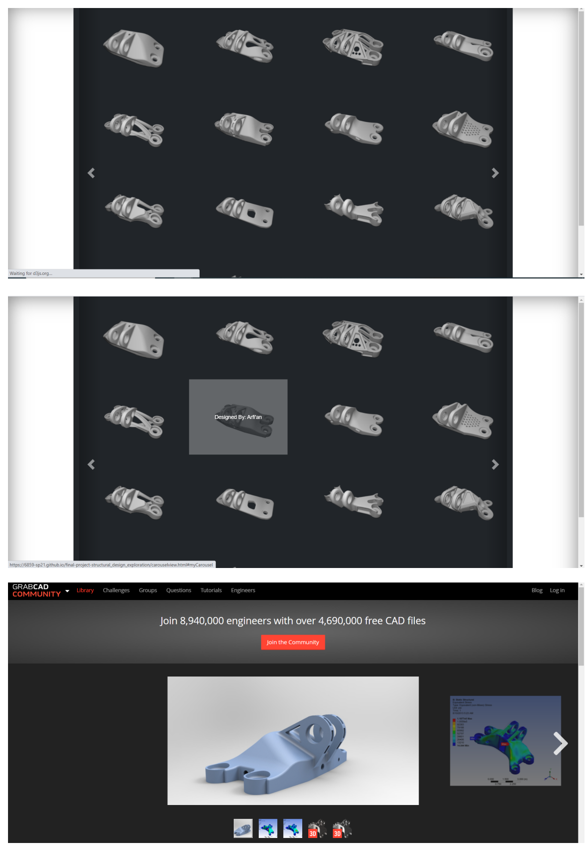
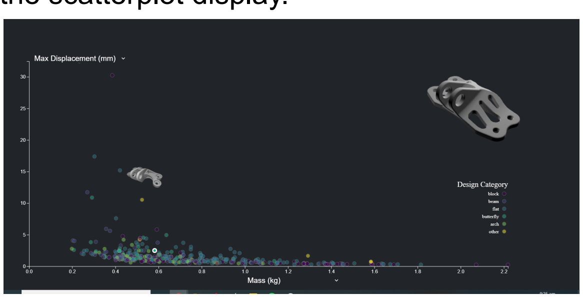


Figure 3: Carousel View (top) with hover-over functionality (middle) that links to individual designs (bottom).

Shown above, Figure 3 depicts the ability of users to quickly screen through all 381 bracket designs using the carousel function. Hovering over individual designs would show users the original creators of the bracket shown, and clicking on the image tile would bring them to the original design submission page on the GE Challenge website.

The Data View provides the user the ability to delve deeper into structural simulation results after familiarizing themselves with the data using the Carousel View. These results were found by performing finite element analysis on the brackets in the set. This view contains interactive plots that allows users to see how specific brackets perform as shown in Figure 4. The plots are encoded using discrete color according to design categories based on the brackets typology. The clickable legend not only serves as a key for the different typologies in the set, but also serves as a filtering mechanism to ease exploration. Users simply click on the label in the legend and the tool allows users to highlight the points associated with the selected categories, while graying out the non selected points. This can help ease obscuration with similar performing designs across different categories, but also allows the user to only look at one group of interest at a time, or easily distinguish categorical clustering in the scatterplot display.



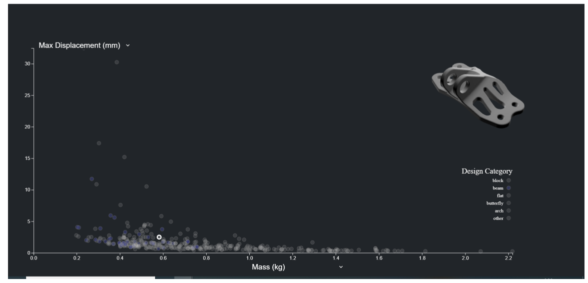


Figure 4: Data View page (top) with interactive legend functionality (bottom)

The tooltip function on the plot shows snapshots of designs that fall on the graph. This connects the numerical simulation results and attributes displayed on the chart to a visual representation. One can commit to the design by clicking on the point, which will load a 3D rendering of the design on the top right corner for further inspection. The inspection window as shown by Figure 5 allows for rotating about all 3 axes via a click and drag function of the mouse, and zooming in and out of the CAD model for a closer glance at fine design details of the bracket. It should be noted that these mouse functionalities are the same as those found in CAD programs such as Solidworks, Fusion 360, Rhino and Hyperworks to allow users to understand the design better.

Figure 5: 3D rendered models.

Lastly, users have some agency over which relationships they want to explore between the designs in the dataset structural performance results. This is done by changing using the drop down menus on the X and Y axes to change the data values being displayed. When one or both axes are changed, the selected point remains highlighted with a white border so that the user can easily gauge how their chosen design performs across other metrics. Shown in Figure 6, the axes changed from Max Displacement against Mass to Max Stress against Max Displacement, highlighting different relationships between the results that might be insightful to the user. While this is not exhaustive of all the properties corresponding to each bracket in the dataset, it is more than sufficient for the average engineer to gain insight in the data

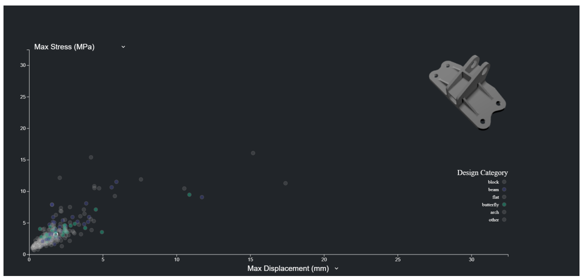


Figure 6: Interactive axis containing simulation results of data.

5 Discussion

This tool gives users from a variety of disciplines the ability to examine mechanical simulation data of engine brackets attached to their geometric representations without having to download the data and build flexible visualizations themselves or run expensive and computationally intensive software. This will cut user data processing time of this set significantly. The two views enable users to choose how deeply they want to examine the data set based on their individual needs. Finally the home page houses information for this published work so that when users of the structural design exploration tool are ready to dig in they can download the paper and the data on their own and build robust surrogate models for their own design practice and work

6 Future Work

There are several key features that may prove useful to users as they continue to take advantage of this tool. The first could be for the users to be able to define their own axis bounds as you could in another plotting software. Another feature that may be of interest to users of this tool may be the ability to link between the carousel view and the data view directly. It also could be interesting to allow users the ability to expand the 3D bracket view to a full page and perhaps have some annotations of key features and attributes such as a engineering drawing. This would require a significant amount more of post-processing of the dataset. Lastly we imagine that users may want to export a plot like this directly for presentation, so perhaps there could be an export to png button, vs having users, screenshot or user other browser functions to accomplish this task.

References

[1] E. Whalen, A. Beyene, and C. Mueller, “SimJEB: Simulated Jet Engine Bracket Dataset,” ArXiv210503534 Cs, May 2021, Accessed: May 12, 2021. [Online]. Available: http://arxiv.org/abs/2105.03534
[2] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in 2009 IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.
[3] Y. LECUN, “THE MNIST DATABASE of handwritten digits,” http://yann.lecun.com/exdb/mnist/, Accessed: Apr. 01, 2021. [Online].

Available:
https://ci.nii.ac.jp/naid/10027939599/
[4] D. Ha and D. Eck, “A neural representation of sketch drawings,” *ArXiv Prepr. ArXiv170403477*, 2017.
[5] A. Seff, Y. Ovadia, W. Zhou, and R. P. Adams, “SketchGraphs: A Large-Scale Dataset for Modeling Relational Geometry in Computer-Aided Design,” ArXiv Prepr. ArXiv200708506, 2020.
[6] S. Koch et al., “ABC: A Big CAD Model Dataset For Geometric Deep Learning,” ArXiv181206216 Cs, Apr. 2019, Accessed: May 05, 2021. [Online]. Available: http://arxiv.org/abs/1812.06216
[7] Munzner, Tamara. "A nested model for visualization design and validation." IEEE transactions on visualization and computer graphics 15.6 (2009): 921-928.