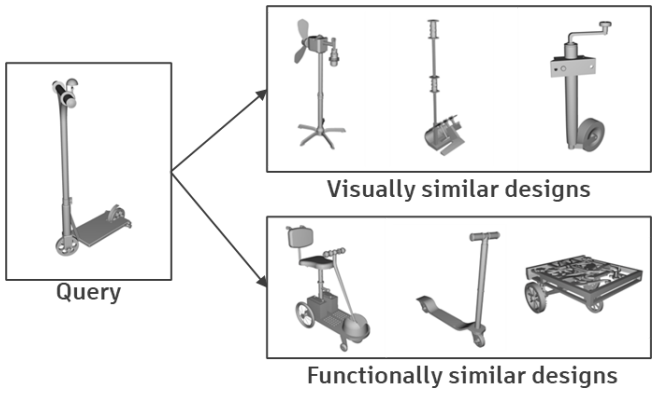
Characterizing Similarity from Computer-Aided Design (CAD) Assemblies

Introduction

In mechanical CAD software, assemblies are collections of parts represented as 3D shapes, that together represent an overall design artifact, or object. In recent years, large collections of CAD datasets have been collected and made public, which have brought new opportunities for searching for design knowledge. By searching for similar shapes, designers could get inspiration from past examples, and in doing so they might more quickly arrive at novel solutions to design problems.

There exist many methods for retrieving similar parts based on geometric and visual similarity. Although these methods work on single parts, they do not take into account all the diverse aspects characterizing an assembly, such as the relations between parts. Moreover, there might be other aspects of the design that designers might be interested in searching for, such as finding assemblies with similar materials, names, functions, or sourced from the same industry. There are many situations that might arise: designers could search large collections of CAD models for inspiration, or they might be looking for more specific instances of a part in different assemblies to learn how it was integrated by others in their designs.



The model of a scooter could be used to find visually similar designs, or search for functionally similar designs.

In this challenge, you will come up with an open-ended solution for characterizing similarities between designs in the Autodesk Fusion 360 Gallery Assembly Dataset. The dataset contains design data from CAD assemblies containing multiple parts. Due to the 24-hour time constraint for the hackathon, we will only focus on the assembly graph (the structure of parts forming the assembly) data, parts' attributes data (e.g. names and materials), and the image data of the

assembly and its parts, and not on the 3D shape data. The dataset is large and rich, so the identification of an appropriate size of the data for a satisfactory algorithm performance is part of the challenge.

Dataset

The dataset used in this hackathon is based on the Fusion 360 Gallery Assembly Dataset, which contains 8,251 assemblies and a total of 154,468 separate parts (i.e., bodies). To simplify the search space, we have provided a smaller subset of this dataset to be used as the official dataset of this hackathon, which you can download following the link below towards the end of this section.

Specifically, each of the assemblies contains the following information: **assembly-level information** (e.g. semantic name, physical properties, assembly tree hierarchy, etc.), as well as the individual bodies along with their connection information that make up the assemblies. Each body that belongs to the assembly also has its **body-level information** (e.g. semantic name, material category, etc.).

The table below summarizes the feature properties and their corresponding short descriptions for each assembly and body file in the IDETC-hackathon-2022 dataset. The "**File**" column shows the file name, the "**Feature Properties**" column shows the assembly and body level features present in the corresponding file, and the "**Feature Description**" column shows the brief description of the corresponding feature property.

File	Feature Properties	Feature Description
assembly.json	Body - Semantic Name	The semantic name of individual bodies, as assigned by the designers
	Body - Material Category	The hierarchical material category of individual bodies
	Body - Physical Properties	The physical properties of individual bodies (e.g., center of mass, area, volume, density, and mass)
	Assembly - Physical Properties	The physical properties of the entire assembly (e.g., center of mass, area, volume, density, and mass). There are also some additional physical properties (e.g., vertex count, edge count, etc.) for the assembly level.
	Assembly - Design	The design information of the entire assembly (e.g., design category, design industries, design type)
	Assembly - Community	The community statistics of the entire assembly as collected on the Fusion 360 Gallery (e.g., the number of views, comments, and likes)
[body_id].jpg	Body - 2D Geometry	A thumbnail image of the body geometry.
[assembly_id].jpg	Assembly - 2D Geometry	A thumbnail image of the assembly geometry.

Here are some useful links that you may reference or use:

- Code repository and documentation to the IDETC-Hackathon-2022: **link**
- Documentation of the Fusion 360 Gallery Assembly Dataset: **link**

Submission

The dataset will contain a validation set composed of 3 assemblies, and a test set composed of 7 assemblies. The test assemblies will be used by the judges to qualitatively evaluate the performance of the search. The test set will be released in last hours of the hackathon. Teams should evaluate their similarity search methods against each of the 7 test assemblies, and return the top-5 most similar results. The results should be included in the final presentation

deck, with the top-5 most similar assemblies for each of the 7 test assemblies on individual slides.

Score Sheets and Judgment Criteria

Category	Criteria	Score
Similarity criteria (30%)	● Teams will present an overview of their definitions of similarity brtween two assemblies. ● Creativity of the similarity metrics ● Data exploration and preparation ● Feature selection and combination	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)
Model development (30%)	● Teams will present an overview of their approaches for calculating the similarity metrics. ● Scientific soundness of the approach ● Readiness of the idea and the approach ● Model comparison and evaluation ● Judges will consider more favorably multi-modal search methods that take into consideration visual, functional, semantic, relational, local, and global similarity aspects.	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)
Qualitative evaluation (30%)	● For each assembly ID in the test set, each team must identify the 5 most similar assemblies in the training set.	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)
Overall Presentation (10%)	● Title, headings, labels: appropriate size, location, spelling, and content ● The demonstration of teamwork ● Structure and clarity ● Boarder impact of the idea to ME subfields	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)