Programming Task: Parsing 2D CAD Sketch JSONs

Motivation: The SketchGraphs dataset [1] contains 15 million human-made sketches for 3D CAD models. These sketches are saved in a JSON format that is not suitable for downstream LLM training to generate 2D sketches and 3D CAD models. We would like to convert these JSONs into a custom programming language for CAD that is Python-like and more suitable for LLMs.

Description: The goal of this programming task is to write a script to parse these JSONs and convert them into a text-based format. You are free to structure your code as you like and make any simplifying assumptions. You are allowed to use any tool or existing codebases. Here is a possible example of the target code:

```
sketch0 = Sketch() # initializes a sketch object
loop0_0 = Loop()
line0_0 = Line(start_point=[-27.79, -19.17], end_point=[-27.79, 13.62])
arc0_1 = Arc(start_point=[-27.79, 13.62], middle_point=[-26.68, 16.39], end_point=[-24.45, 19.17])
line0_2 = Line(start_point=[-24.45, 19.17], end_point=[2.50, 19.17])
line0_3 = Line(start_point=[2.50, 19.17], end_point=[2.50, 0.00])
line0_4 = Line(start_point=[2.50, 0.00], end_point=[27.79, 0.00])
line0_5 = Line(start_point=[27.79, 0.00], end_point=[27.79, -19.17])
line0_6 = Line(start_point=[27.79, -19.17], end_point=[-27.79, -19.17])
loop0_0.add([line0_0, arc0_1, line0_2, line0_3, line0_4, line0_5, line0_6])

loop0_1 = Loop()
circle0_1 = Circle(center = [100.0, 100.0], radius = 25.0)
loop0_1.add([circle0_1]) # circles are loops on their own
sketch0.add([loop0_0, loop0_1])
```

For this task, you can ignore construction lines and constraints. Focus only on getting the sketch elements formatted as a programming language as you see above.

Deliverables:

- Your own fork of this repository with a python script to convert JSONs from the SketchGraphs format to a text-based representation (files are under `data/` directory)
- Validation of results by visualizing the sketches from the JSONs and your outputs side-by-side
- Short documentation of the code (as a README.md or a Jupyter notebook)

Evaluation Criteria: It is okay if you do not get all the functionality right or there are missing parts. *The goal of this task is to understand how you approach complex problems and to see your coding style.* Here are important aspects to consider:

- Code structure and cleanliness
- Python coding style
- In-depth validation of the functionality and testing the code
- Quality of the documentation and explaining your approach to the problem

Timeframe: 1 week

Feel free to reach out to me if you have any questions or concerns. Good luck!

References

[1] Seff, Ari, et al. "Sketchgraphs: A large-scale dataset for modeling relational geometry in computer-aided design." *arXiv preprint arXiv:2007.08506* (2020).