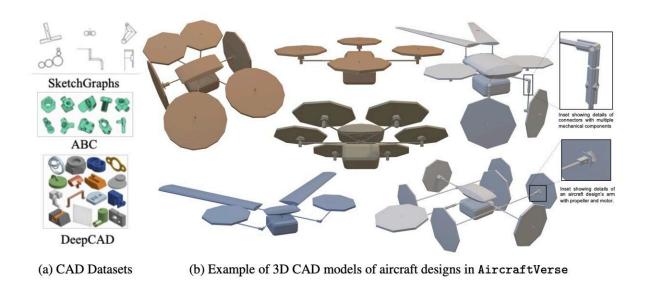# A large dataset to train machine learning models for aerial vehicle design

July 10 2023, by Ingrid Fadelli



(a) CAD Datasets    (b) Example of 3D CAD models of aircraft designs in AircraftVerse

Credit: Cobb et al.

Designing reliable aircraft can be both challenging and time-consuming, as it often entails several steps and analyses. Deep learning models could potentially help to speed up aircraft design and deployment, helping developers to identify the most promising solutions or potential flaws with a specific aircraft.

To train these models, researchers would need comprehensive datasets containing a wide range of aerial vehicle designs. However, these

datasets can be difficult to compile, as many designs are protected by proprietary contracts or difficult to source.

Researchers at SRI International, the Southwest Research Institute and Vanderbilt University recently created AircraftVerse, a large-scale dataset containing thousands of aircraft designs of varying complexities. Their dataset, presented in a paper pre-published on *arXiv* could be used to train machine learning to assist aerial vehicle designers.

"Aircraft design encompasses different physics domains and, hence, multiple modalities of representation," Adam D. Cobb, Anirban Roy and their colleagues wrote in their paper. "The evaluation of these cyber-physical system (CPS) designs requires the use of scientific analytical and simulation models ranging from computer aided design tools for structural and manufacturing analysis, computational fluid dynamics tools for drag and lift computation, battery models for energy estimation, and simulation models for flight control and dynamics."

Most existing datasets to train machine learning for computer assisted design (CAD), such as the SketchGraphs, DeepCAD and ABC datasets, primarily contain data related to individual mechanical parts. The dataset introduced by Cobb, Roy and their colleagues, on the other hand, contains fully-fledged aircraft designs that combine multiple components, such as propellers, wings, motors, batteries, and so on.

"AircraftVerse contains 27,714 diverse air vehicle designs—the largest corpus of engineering designs with this level of complexity," Cobb, Roy and their colleagues explained in their paper.

"Each design comprises the following artifacts: a symbolic design tree describing topology, propulsion subsystem, battery subsystem, and other design details; a Standard for the Exchange of Product (STEP) model data; a 3D CAD design using a stereolithography (STL) file format; a 3D

point cloud for the shape of the design; and evaluation results from high fidelity state-of-the-art physics models that characterize performance metrics such as maximum flight distance and hover-time."

The designs included in the AircraftVerse dataset were created using a deep learning-based approach, based on general rules provided by expert aircraft designers. The researchers ran the final versions of these designs through engineering models that produced metadata summarizing each of their unique characteristics and performance.

"We also present baseline surrogate models that use different modalities of design representation to predict design performance metrics, which we provide as part of our dataset release," Cobb, Roy and their colleagues wrote. "Finally, we discuss the potential impact of this dataset on the use of learning in aircraft design and, more generally, in CPS."

The new dataset created by this team of researchers is now publicly available online, along with its baseline models and underlying code. This means that it could soon be used by designers and developers worldwide, assisting them with the design and performance evaluation of new aerial vehicles.