Hackathon Problem: Segmenting Medical Images toward Digital Twin in Healthcare

Image segmentation is an essential step for generating a digital twin (DT) of a patient body in healthcare. In the medical domain, typically, segmenting regions of interests for a DT requires medical images and volume models. Examples of such data are Digital Imaging and Communication in Medicine (DICOM) data, Picture Archiving and Communication System (PACS) data, and three-dimensional (3D) mesh data such as STereoLithography models. Although techniques related to above mentioned segmentation have been developed, medical image segmentation performed in the real-world has limitations due to data imbalance in practical scenarios.

The objective of this hackathon problem is to address a data-imbalance problem in medical image segmentation. Medical images will be provided, and participants are expected to complete the objective by gathering evidence as much as possible from the provided medical images to prove the participants' conclusion. Teams would be required to present their solutions and approaches for completing each benchmark to a panel of judges.

Challenges

- How can data-analytics techniques be developed for generalizability of segmenting medical images in a DT process?
- How can the imbalance in medical images be minimized in DTs?

Objectives

A 56-year-old patient has taken computerized tomography (CT) scans as a regular health check-up. At this time, the radiologist who examines the patient’s CT scan finds renal cell carcinoma (RCC) on the patient’s kidney. Therefore, a doctor decides to perform a surgery to remove the RCC. The doctor will preserve the patient's renal function as much as possible and figure out the best surgical approach by utilizing a 3D-printed model. The 3D reconstruction of the kidney and the RCC for 3D printing will be extracted from the patient's CT images. In order to accurately diagnose and measure the RCC, the doctor needs a contrast-enhanced abdominal CT scan with the standard radiation dose. However, there are a few limitations as follows.

1. Since the patient has taken several CT scans for a year, constant exposure of the patient to x-ray radiation has already been high. To minimize the radiation exposure, the doctor can consider a method such as decreasing the radiation dose, thickening the slice thickness of the CT scan, or increasing the pitch speed of the patient table. However, these methods could amplify noise in a CT image or a 3D model constructed using CT images, such as increased, stair-step effects in a 3D model caused by a thickened slice thickness.

2. Injecting a contrast agent in the patient body can enhance the contrast of the CT images, which can help the diagnosis and measurement of the RCC. However, this time, this method is not possible for the patient because of the patient’s kidney problem. A contrast

agent cannot be used for a patient with a kidney problem due to an allergic reaction or other safety issues the agent can potentially cause.

3. The CT scanner at the hospital has recently been replaced with a new scanner from another manufacturer. In general, the textures of the anatomic structures are different depending on the models of CT scanners, which means that changing the CT scanners may reduce the consistency of the diagnosis of a lesion. Therefore, the new CT scanner needs a reconstruction method and a reconstruction kernel different from the ones for the old scanner.

Although these limitations described above inevitably cause difficulties in segmenting the images for 3D printing, the only way to print the 3D kidney and RCC models specific for the patient is to segment the patient's CT images. To address the limitations, the doctor will additionally use the existing CT images of other patients generated in various conditions. Figure 1 shows an example of a CT image with a highlighted area of RCC.

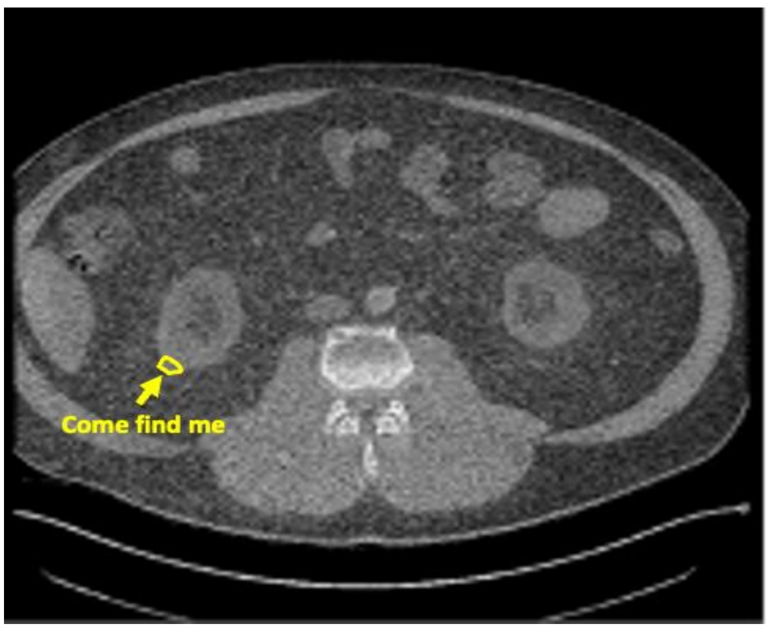

Figure 1. An example of a CT image with a highlighted area of RCC

Find the following from the 56-year-old patient's CT images.

1) The volume of the right kidney in a 3D model
2) The volume of the RCC in a 3D model
3) The mean Hounsfield Unit (HU) of the right kidney
4) The mean HU of the RCC

Data Types and Formats

X-ray CT images from various image acquisition conditions changed depending on manufacturers, enhanced contrast, slice thickness, and so on. The data types include dicom, nifti, and mip files.

Results Submission Table

Region of Interest (ROI)	Volume (mm^3)	Mean HU	Brief description of the solution method
Right Kidney			
RCC			

Judgment Criteria

Category	Criteria	Scoring
Results (60%): Output solution	Accuracy of the volumes and the mean HUs of the kidney and the RCC. Clear and concise explanation of the obtained solution	100*(1/RMSE)
Creativity (20%): A new direction in the field to approach the problem	A solution through critical thinking. A major departure from other submissions Creativity in solving the problem Use of appropriate software to aide in problem solving	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)
Overall presentation (20%): Organization, structure, and message conveying	Title, headings, and labels with - appropriate sizes, locations, spelling, and contents The demonstration of teamwork Structure and clarity Clear and concise explanation of approaches takes	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)

Submission

- The presentation slides describing the overall approach to obtain the solution and outlining the difficulties faced.
- Each team will submit a zip file containing:
    a. A detailed word document which includes:
        i. The completed submission table from above
        ii. A description of the brainstorming process
        iii. A summary of any other approach attempted that may not have been successful to provide insight into your effort level and thought process.
    b. Any supplementary file to support your report (e.g., images, STL files, dicom/nifti files, and programming scripts)

References

- Bhosale, P. (2015). Comparing CNR, SNR, and image quality of CT images reconstructed with soft kernel, standard kernel, and standard kernel plus ASIR 30% techniques. *International Journal of Radiology*, *2*(2), 60-65.
- Kubo, T., Lin, P. J. P., Stiller, W., Takahashi, M., Kauczor, H. U., Ohno, Y., & Hatabu, H. (2008). Radiation dose reduction in chest CT: a review. *American journal of roentgenology*, *190*(2), 335-343.
- Kuehn, B. M. (2010). FDA warning: CT scans exceeded proper doses. *JAMA*, *303*(2), 124-124.

Subject Matter Experts and Mentors:

Jong-Min Kim, PhD, received his PhD degree in ICT Convergence Technology at Korea University. Also, He was a research professor at Korea University. He is currently working as a Leader in Research and Science Division of Medical IP Co., Ltd.

Chorong Hong, MS, received her master of science in Modern Human Anatomy, University of Colorado, Anschutz Medical Campus. She is currently assistant manager of Data Science Division in Medical IP, Co., Ltd.

Han-Jae Chung, MS, earned his master of science in ICT Convergence Technology at Korea University. He is currently working as an assistant manager in Research and Science Division of Medical IP Co., Ltd.

Khwaja Monib Sediqi, MS, obtained his Master of Science in Computer Science and Engineering from Jeonbuk National University through the most prestigious Korean Government Scholarship Program (KGSP). He is currently working as an assistant manager in Research and Science Division of Medical IP Co., Ltd.

Hyunwoong Ko, PhD, is an assistant professor in the School of Manufacturing Systems and Networks at Arizona State University.

Hackathon Problem: Digital manufacturing cybersecurity strategies for protecting valuable information in design files

Problem Statement

A digital manufacturing (DM) process chain requires the use of computers, network connectivity, and cloud systems. Industry 4.0 continues to evolve towards the digital transformation of manufacturing, leading to concerns of hacking for sabotage and intellectual property protection. The unique threats faced by DM are side-channel attacks, direct sabotage, reverse engineering, and counterfeit production.

The objective of this hackathon problem is to assess the robustness of security strategies to hide information in the design files for DM and stimulate the critical thinking process. An STL file of a model will be provided, and participants are to complete the objective by gathering as much evidence from the provided files to prove their conclusion. Teams would be required to present their solution approaches for completing each benchmark to a panel of judges.

Challenges

- How can security strategies be developed and incorporated into a DM cyber-physical system? [1]
- What is the optimal approach to test the effectiveness of developing security strategies and to account for every classification of attacks in the DM supply chain? [2]
- How can the cybersecurity threats be minimized in digital manufacturing?
- Is current 3D printing technology safe from threats?

Objective

You and your friends are traveling across the globe to multiple locations for summer vacation. One day, you discovered that your passport was missing, and you received a mysterious email containing a file. The anonymous sender requires you to solve the puzzle in the attachments to be able to find where your passport is located. You and your friends are only given 24 hours to locate your passport.

The STL file shows a 3D model of an object and there are five hints that are hidden throughout the files. Each hint that you can decode will get you closer to the location of the lost passport. Teams will receive points based on how many puzzles they can decode correctly and their method of solving the challenges.

Results submission table:

Hint #	The location provided by the hint	Brief description of the hint and solution method
1	?	
2	?	
3	?	
4	?	
5	?	

Judgment Criteria

Category	**Criteria**	**Scoring**
Results (60%): Output solution	● The objective is achieved by determining the exact GPS location of the passport ● Clear and concise explanation of obtaining solution	Correctly determining: 12 points for each clue
Creativity (20%): A new direction in the field to approach the problem	● Derived solution through critical thinking ● The approach is a major departure from other submissions ● Team demonstrates creativity in solving each puzzle ● Use of appropriate software to aide in problem solving	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)

Overall presentation (20%): Organization, structure, and message conveying	• Title, headings, labels: Appropriate size, location, spelling, and content • The demonstration of teamwork • Structure and Clarity	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)

Submission

1. The presentation slides describing the overall approach to obtain the solution for each benchmark and outlining the difficulties faced.
2. Each team will submit a zip file containing:
   a. A detailed word document which includes:
      i. The completed submission table from above
      ii. A description of the brainstorming process and each clue
      iii. A summary of any other approach attempted that may not have been successful to provide insight into your effort level and thought process.
   b. Any supplementary file to support your report (CAD/STL files, programming scripts, images)

Sample Data Set

Click for sample data set

References

1. Mahesh, P., et al., *A Survey of Cybersecurity of Digital Manufacturing.* Proceedings of the IEEE, 2021. **109**(4): p. 495-516.

2. Linares, M., et al. *HACK3D: Crowdsourcing the Assessment of Cybersecurity in Digital Manufacturing*. 2020. arXiv:2005.04368.

3. Practice problems and previous challenges are available at: https://www.csaw.io/hack3d

Subject Matter Experts and Mentors:

Nikhil Gupta, Professor, Department of Mechanical and Aerospace Engineering, New York University

Gary Mac, Ph.D. Candidate in Mechanical Engineering, Department of Mechanical and Aerospace Engineering, New York University

Hammond Pearce Postdoctoral Associate, Department of Mechanical and Aerospace Engineering, New York University

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.

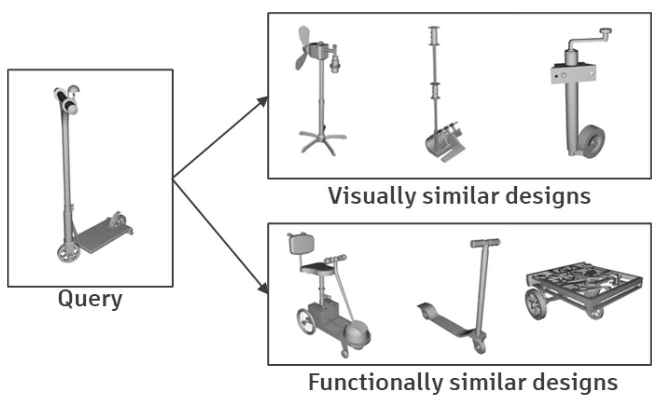

Figure 1. The model of a scooter could be used to find visually similar designs, or search for functionally similar designs.

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
	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.json	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.

Judgment Criteria

Category	Criteria	Score
Similarity criteria (30%)	• Teams will present an overview of their definitions of similarity between 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.	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts)

	• 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.	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)

Subject Matter Experts and Mentors

Ye Wang, Principal Research Scientist, Autodesk, Inc.

Daniele Grandi, Sr. Research Engineer, Autodesk, Inc.

Binyang Song, Postdoctoral Associate, Department of Mechanical Engineering, Massachusetts Institute of Technology

Zhenghui Sha, Assistant Professor, J. Mike Walker Department of Mechanical Engineering, The University of Texas at Austin