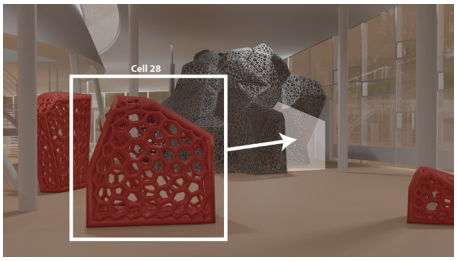
# One-Pager Computer Vision

*Group 3 – Oliver J. Post, Jiacheng Xu, and Jiahui Shi*

## Description

Our project is built up from so called Voronoi cells. These 3D shapes only fit together one certain way, a way which is not always immediately apparent to a human. Therefore, we want to use computer vision to solve this 3D puzzle for the people trying to reattach detachable pieces of our Voronoi patterned structure. This task is split up into two parts:

- Identify detached Voronoi cells
- Identify the places where each of these cells are supposed to go on the main structure

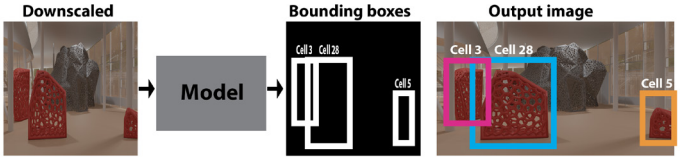



Render showing identification of detached cell (left) and instruction where it's supposed to go (right)

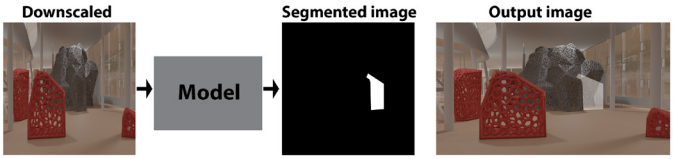
## Pipeline

For task 1 – identifying detached pieces – we intend to use object recognition. For task 2 – showing the place it's supposed to go – either image segmentation or image generation can be used.

The input for both task is a stream of images from the user's smartphone camera. The output is also a stream of images shown to the user on their screen. In the case of task 1 the actual output from the model are the bounding boxes, not images.




*Schematic representation of task 1*



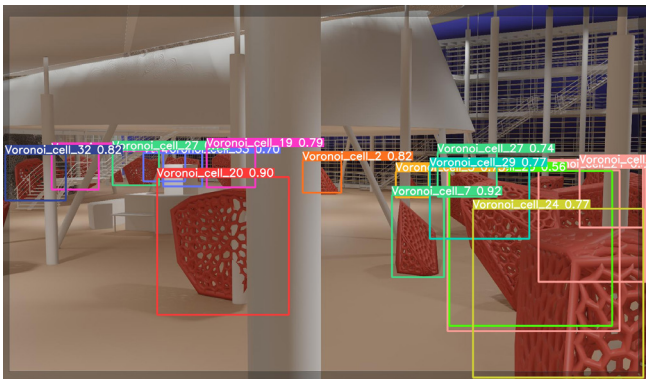

*Schematic representation of task 2*

## Data

To achieve these tasks, the following data is needed:

**Task 1:**

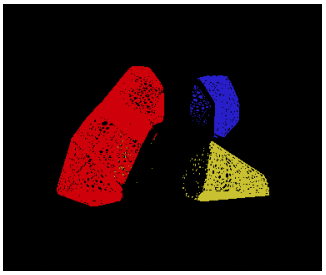
- Input: Image dataset showing one or several pieces in different configurations, scenarios and camera angles
- Ground truth: Bounding box data of the pieces that are visible in the frame, including their name/identifier




*Example for task 1 dataset.*

**Task 2:**

- Input: Random images of the main structures with the loose pieces missing
- Ground truth: RGB mask for the zones where the pieces are missing. Each color of the masks corresponds to a detachable piece.



*Example for task 2 dataset ground truth.*

There are no relatable datasets already available that solve this specific problem, so new ones are created.

## Software

Both datasets are created synthetically using renders and geometry data from the 3D software Blender. This approach both enables to showcase the models working in real time and to quickly and effectively create datasets for these complex problems.

For task 1 we use the YoloV5 object detection model, implemented in PyTorch. This model allows us to train without prior machine learning experience and to run the final implementation in real time.

For task 2 we will either use an image segmentation model or the pix2pix model.