Chemscape chemical lab visual dataset

General data set structure

The image and the annotation data for the specific image are stored in folders that have the image name (the name usually reflects the website or specific youtube channel from which the image was extracted).

In each of these folders contain the following:
Files:
Image.png file that contains the image
Ignore.png file that contains maps of regions that were not annotated. The ignore/unannotated regions for the specific instance and segmentation also appear inside each specific annotation.

Folders:

The semantics folder contains the semantic annotation for the image each label has its own annotation map, in a separate file.
Materials/Parts/Vessels folders contain masks of specific instances. Each instance is stored in a separate file.
Materials instance are simply instances of liquid/solids/foam phases etc…
Parts Instance: are simply instances of vessel parts (labels corks…)
Vessel instances: are instances of different vessels.

Semantic:

Semantic maps contain maps for any the following classes:

1   Vessel
2   Vessel Label
3   Vessel Cork

4   **Vessel Parts** general (Include corK and label)

5   **Ignore** (Image area that was not semantically annotated)

6   **Liquid general** (Include Suspension/Emulsion and gel)

7   Liquid Suspension/emulsion

8   Foam

9   Gel

10   **Solid General** (Include Granular/Powder/Bulk/Gel)

11   Granular

12   Powder

13   Solid Bulk

14   Vapor

15   Other Material (Unundetified material or materials that do not match any of the above classes)

16   Filled vessel (Include all materials except vapor)

Note that these classes are none exclusive and any pixel can be of several classes. Some classes like solid and liquid contain other subclasses. For example, the solid class includes powder and granular classes. Hence, any pixel that belongs to the powder class also belongs to the solid class.
Note that if a given class does not appear in the annotated region of the image there will be no segmentation file for this class (which is equivalent to having a map of all zeros)

Semantic annotation file structure:
Semantic map for each class is written in a separate png image:
The file name contains the class number and the class name.
(10_Solid_GENERAL.png for example)
The png images are standard 3 channels images.
The first channel is a binary mask of the regions containing the class in front (meaning not occluded by other material/vessel etc). Ence pixels containing the class will have value of 1 and the rest 0.
The second image channel is the binary mask of the regions belong to the class but which are beyond/occluded by other material/vessel (background).

The third channel is the merging of the first and second channels.
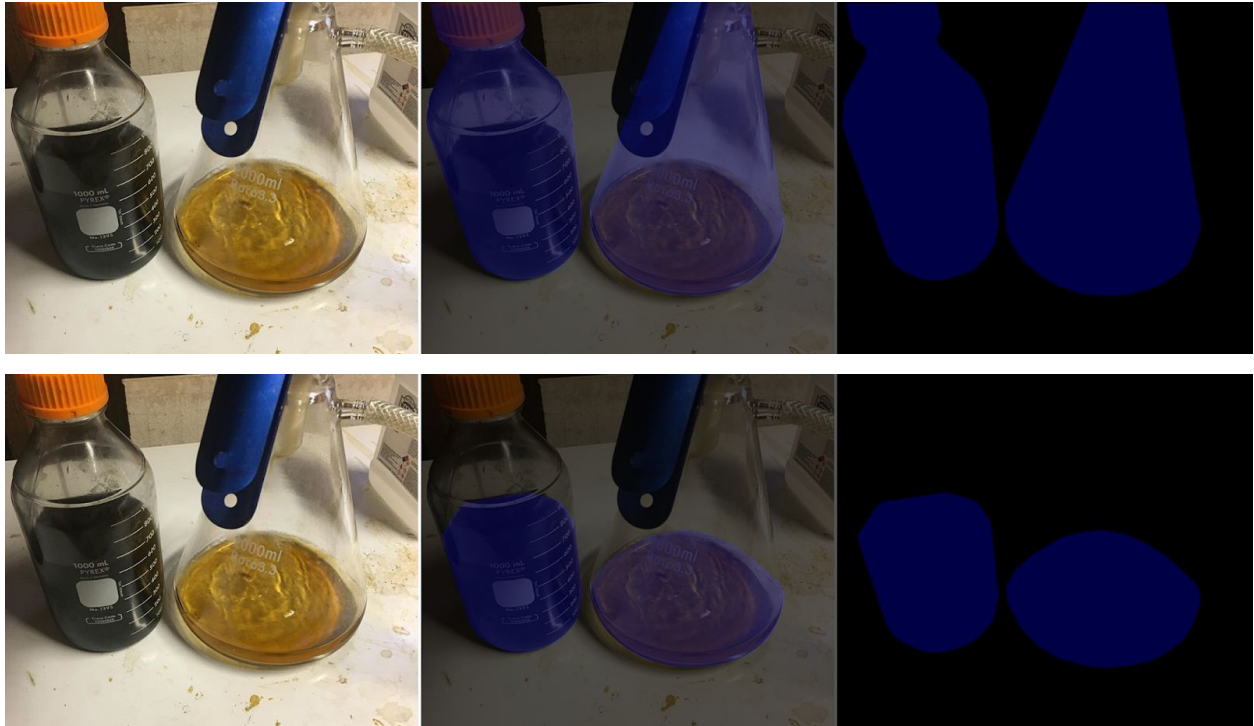
Liquid vessels semantic map

Semantic ignore regions:
If some regions were not semantically annotated or are ambiguous the semantic folder will contain a file called *5_Ignore.png* which contain binary mask for all the regions that were not properly annotated and should be ignored.

Instances

There are 3 types of instances:
1: Materials phases
2: Vessels parts
3: Vessels

Materials phases

(in the Material subfolder)

The material folder contains a separate png file to each instance of material saved as a 3 channels png image (uint8).

The file name of each instance contains the instance number the name and the ID of the categories for this material

Example liquid phase instance

png instance file structure

Channel 1: Material phase mask

The first image channel is the instance mask, this has the following values:

0: Not instance

1: Instance region (no intersection with other instances)

2: Instance region intersect with other instance regions but is foreground/priority.

3/4: Instance region intersects with other instance regions but is background/occluded/low priority.

5: Instance of material in a vessel that is beyond/occluded by another vessel.

Channel 2: Containing vessel mask

The second image channel is an instance mask of the vessel containing the material instance.

The same as the first channel only the mask is for vessel instance instead of material phase instance.

Channel 3: Ignore/unannotated regions, and additional containing vessel masks

The third channel might contain mask of the instance regions that were not properly annotated (Ignore) marked as 7.

In addition, on a rare occasion, the material is contained by two vessels the third channel will have the second vessel mask the same in channel 2.

Vessel parts instance

In Parts subfolder

Instances of vessel parts such as label cork valves…

The data file is exactly the same as the material instances but for vessel parts instances.

Vessel instances

Vessel subfolder

An instance of all marked vessels in the image as a 3 channels png image (uint8 format).

Cork instance mask

Channel 1: Vessel instance mask

The first image channel is the image segment, this has the following values:

0: Not instance

1: Instance region (no intersection with other instance)

2: Instance region intersect with other instance regions but is foreground/priority.

3/4: Instance region intersects with other instance regions but is foreground/occluded/priority.

5: Instance is in a vessel that is beyond/occluded by another vessel.