

ChemScape visual chemistry lab dataset

Simple Dataset

This is a simple version of the chemscape dataset with instance and semantic maps. But no occlusion.

General:

The dataset contains two folders:

- 1) semantic which contain semantic segmentation maps (class per pixel.)
- 2) Instance map that contain for each pixel the instance to which it belongs.

Semantic segmentation:

This contain folder corresponding to each category of the following list.

- 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 (undefined material or materials that do not match any of the above classes)
- 16 Filled vessel (Include all materials except vapor)

Each folder contain annotations mask for all of the figures in the dataset. The name of the annotation is same as the name of the corresponding image in the “image” folder only with .png instead of .jpg. If the image contain no pixel with this class no annotation file for this image will appear in this class folder.

Annotation structure: The annotation is one channel image file in size of the corresponding jpg image. The pixel values are as follows:

0: Pixel does not belong to the class

- 1: Pixel belong to the class
- 2: Pixel belong but the material/part class but is occluded by other object belonging to this class.
- 254: Pixel was not annotated and the area should be ignored

Note that these classes are non 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).

Note that the Semantic folder also contain folder called **5_Ignore** which is simply map of the areas that were not annotated and should be ignored (same as the 254 pixels in the other annotations)

Supporting script: *EvaluateSemanticMaps.py*

Instance Segmentation:

The instance folder contain instance maps for corresponding to each image.

The instance maps are 3 channels:

Channel 1(b): Material instances

Channel 2(g): Parts instances

Channel 3(r): Vessels instances

Each pixel and each channel contain a number corresponding to the instance for which its belongs. Pixel which does not correspond to any instance are marked 0. Pixels which were not annotated and should be ignored are marked 254.

Instance classes

The categories of the instances are given in the *InstCategory.json* file in the main folder.

This file contains a dictionary of all the image Instance annotation. For each material and part instance it contains a list of all the classes for which it belongs.

In addition for each annotation it contains the list of all instance of materials and vessels that are part of multiphase system (vessel that contain more than one phase)

Supporting script: *EvaluateInstanceDataSet.py*