

Kather datasets

We evaluate S0oD on two H&E stained publicly available CRC datasets, Kather-19 (K19) [1] and Kather-16 (K16) [2], as the source domain and target domain

[1] Kather_2019 (Kather, Jakob Nikolas, Halama, Niels, & Marx, Alexander. (2018). 100,000 histological images of human colorectal cancer and healthy tissue (Version v0.1) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.1214456>).

[2] Kather_2016 (Kather, J. N., Zöllner, F. G., Bianconi, F., Melchers, S. M., Schad, L. R., Gaiser, T., ... Weis, C.-A. (2016). Collection of textures in colorectal cancer histology [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.53169>).

How to download the data:

1) Navigate to the corresponding link for each dataset and download the appropriate zip file into `datasets/Kather_2016` and `datasets/Kather_2019` folder. 2) Unzip the files and rename them `Images`.

Data splits:

For each dataset, we split the data into the train, val and a test json files.

Each json file has the following format:

```
images :
  - {path: path_to_image1, label: image1 label, brut_label: image1 original
    label}
  - {path: path_to_image2, label: image2 label, brut_label: image2 original
    label}
  ...
labels: [List of labels]
brut_labels: [List of original labels]
```

How to generate splits (please do not run it if you want to replicate our results).

Run:

```
python Kather.py
```