

Base Dataset

[Print to PDF](#) ▶

Contents

- `AnomalibDataset`

Anomalib dataset base class.

```
class anomalib.data.base.dataset.AnomalibDataset(task, transform=None)
```

Bases: `Dataset`, `ABC`

Anomalib dataset.

The dataset is based on a dataframe that contains the information needed by the dataloader to load each of the dataset items into memory.

The samples dataframe must be set from the subclass using the setter of the *samples* property.

The DataFrame must, at least, include the following columns:

- *split* (str): The subset to which the dataset item is assigned (e.g., 'train', 'test').
- *image_path* (str): Path to the file system location where the image is stored.
- *label_index* (int): Index of the anomaly label, typically 0 for 'normal' and 1 for 'anomalous'.
- *mask_path* (str, optional): Path to the ground truth masks (for the anomalous images only).

Required if task is 'segmentation'.

Example DataFrame:

	image_path	label	label_index	mask_path	split
0	path/to/image.png	anomalous	1	path/to/mask.png	train

Note

The example above is illustrative and may need to be adjusted based on the specific dataset structure.

Parameters:

- **task** (*str*) – Task type, either 'classification' or 'segmentation'
- **transform** (*Transform, optional*) – Transforms that should be applied to the input images. Defaults to `None`.

property category: *str* / *None*

Get the category of the dataset.

property has_anomalous: *bool*

Check if the dataset contains any anomalous samples.

property has_normal: *bool*

Check if the dataset contains any normal samples.

property name: *str*

Name of the dataset.

property samples: *DataFrame*

Get the samples dataframe.

subsample(*indices*, *inplace=False*)

Subsamples the dataset at the provided indices.

Parameters:

- **indices** (*Sequence[int]*) – Indices at which the dataset is to be subsampled.
- **inplace** (*bool*) – When true, the subsampling will be performed on the instance itself. Defaults to `False`.

Return type:

`AnomalibDataset`

< Previous
[Base Data](#)

Next >
[Base Datamodules](#)