

Creating letter datasets with generateDataset.py

Inside of uavf\_2023 repo in the letter-model

Navigate to ./imaging/letter-detection/generateDataset.py

Update the size of the generated dataset and the distributions of each key in the main() method

```
def main():
    d = DataSet(2000) # Change the number here to change the size of the dataset

    custom_distributions = {}
    #add (key, 0 <= dist <= 1) to the custom_distributions
    #TODO Have them read from a file
    for key, dist in custom_distributions:
        d.setDistribution(key, dist)

    d.setRemainingDistributions()
    d.generate()
```

Run generateDataset.py to create a dataset. It will put the data and labels in the ./dataset directory.

The labels.txt has the name of each photo and the label of that photo.

Each label corresponds to a different character. Starting at 'a' = 0 and ending at '9' = 34

0

```
letter-detection > dataset > ≡ labels.txt
1  file, label
2  0.jpg, 0
3  1.jpg, 1
4  2.jpg, 2
5  3.jpg, 3
6  4.jpg, 4
7  5.jpg, 5
8  6.jpg, 6
9  7.jpg, 7
10 8.jpg, 8
11 9.jpg, 9
12 10.jpg, 10
```