



LIGHTLY

Dataset Filtering and Analytics Report


22/01/2021 10:16:27

General Information

Job Information

| Metric | Value |
|--------------------------|------------------------------|
| Build Time | Fri Jan 22 09:49:11 UTC 2021 |
| Sampling Method | Coreset Algorithm |
| Number of Images | 2975 |
| Number of Corrupt Images | 0 |
| Number of Duplicates | 0 |
| Number of Removed Images | 52 |
| Number of Output Images | 2923 |
| Job Submitted | 22/01/2021 09:49:33 |
| Job Finished | 22/01/2021 10:14:34 |
| Total Processing Time | 25m 01s |

Estimated Savings

| Task | Annotation Savings* | CO2 Savings*  |
|-----------------------|---------------------|--|
| Image Classification | \$ 15.60 | 0.00 kg |
| Object Detection | \$ 62.40 | 0.01 kg |
| Semantic Segmentation | \$ 312.00 | 0.22 kg |

*<https://lightly.ai/report>

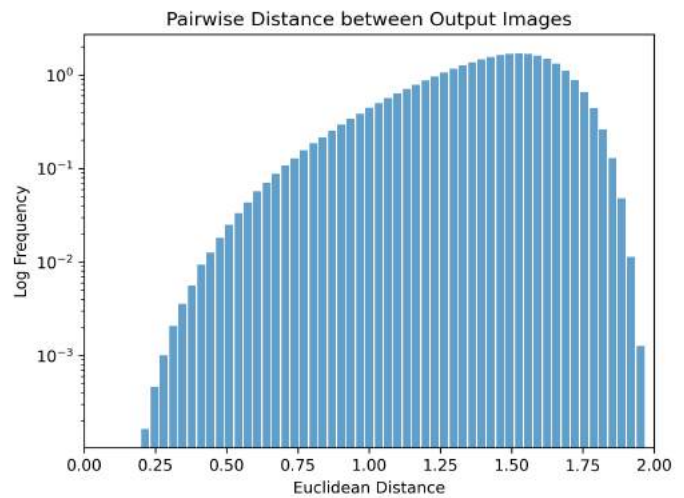
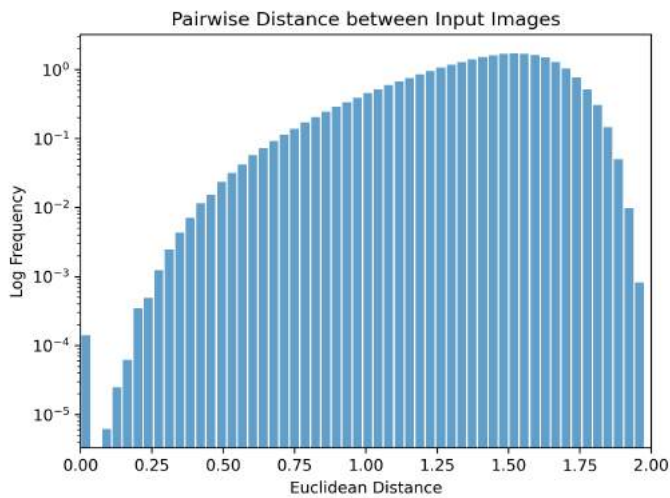
Statistics

| Metric | Before | After |
|--------------------------------------|--------|--------|
| Euclidean Distance (Mean) | 1.3901 | 1.3903 |
| Euclidean Distance (Min) | 0.0000 | 0.1974 |
| Euclidean Distance (Max) | 1.9766 | 1.9684 |
| Euclidean Distance (10th Percentile) | 1.0275 | 1.0285 |
| Euclidean Distance (90th Percentile) | 1.6869 | 1.6871 |

Visualizations

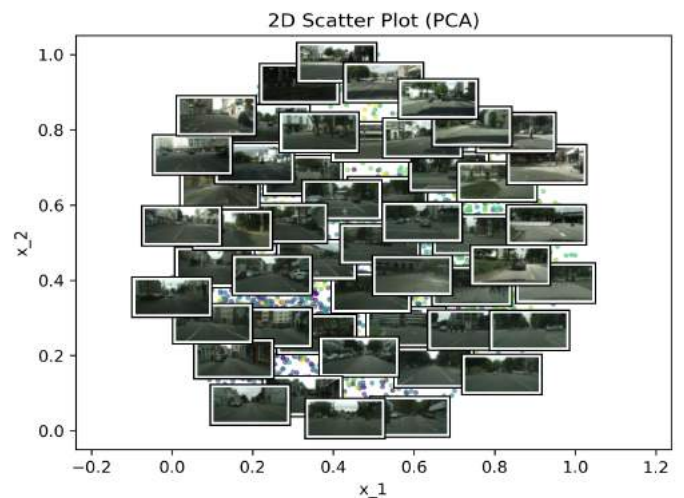
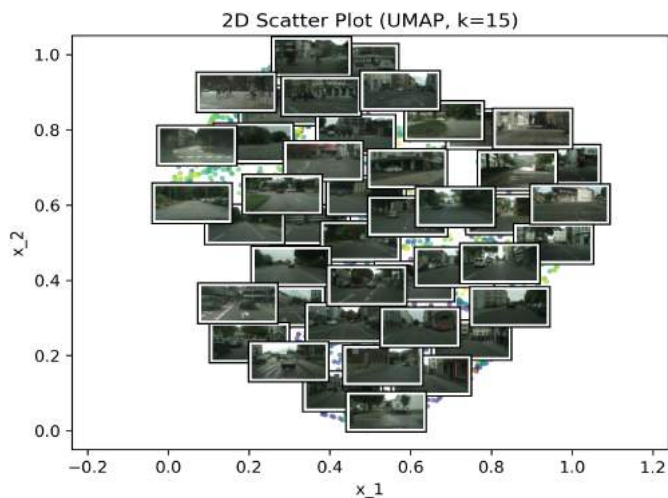
Image Similarity in Input and Output Data

The plots below show the distribution of the pairwise distance between images in the input and output data. The histograms allow you to get information about the diversity of the dataset and whether the filter strength is well-chosen.



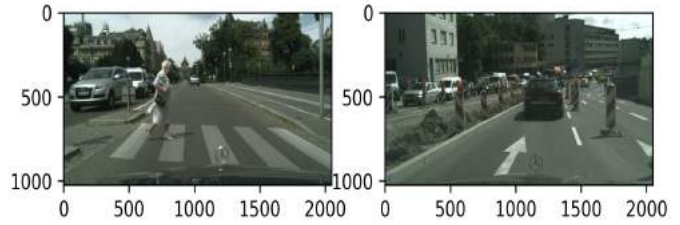
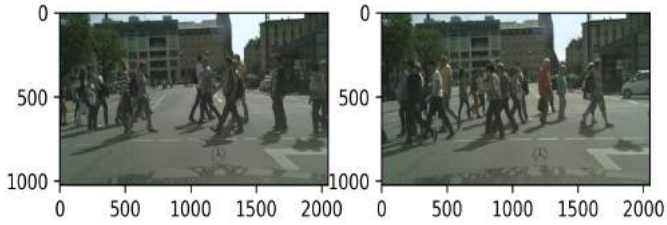
2D Scatter Plots of Output Data

Two-dimensional scatter plots help to understand the distribution of the data and may enable quick insights about outlier cases, dataset bias, or class imbalances.



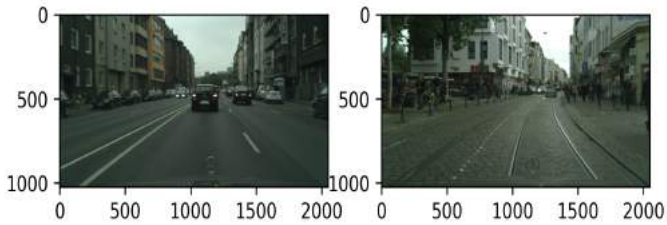
Retained (Left) and Removed (Right) Image with $d = 0.09$

Retained (Left) and Removed (Right) Image with $d = 0.45$



Retained (Left) and Removed (Right) Image with $d = 0.58$

Retained (Left) and Removed (Right) Image with $d = 0.68$



Retained (Left) and Removed (Right) Image with $d = 0.77$

Retained (Left) and Removed (Right) Image with $d = 0.87$

