

Data_report

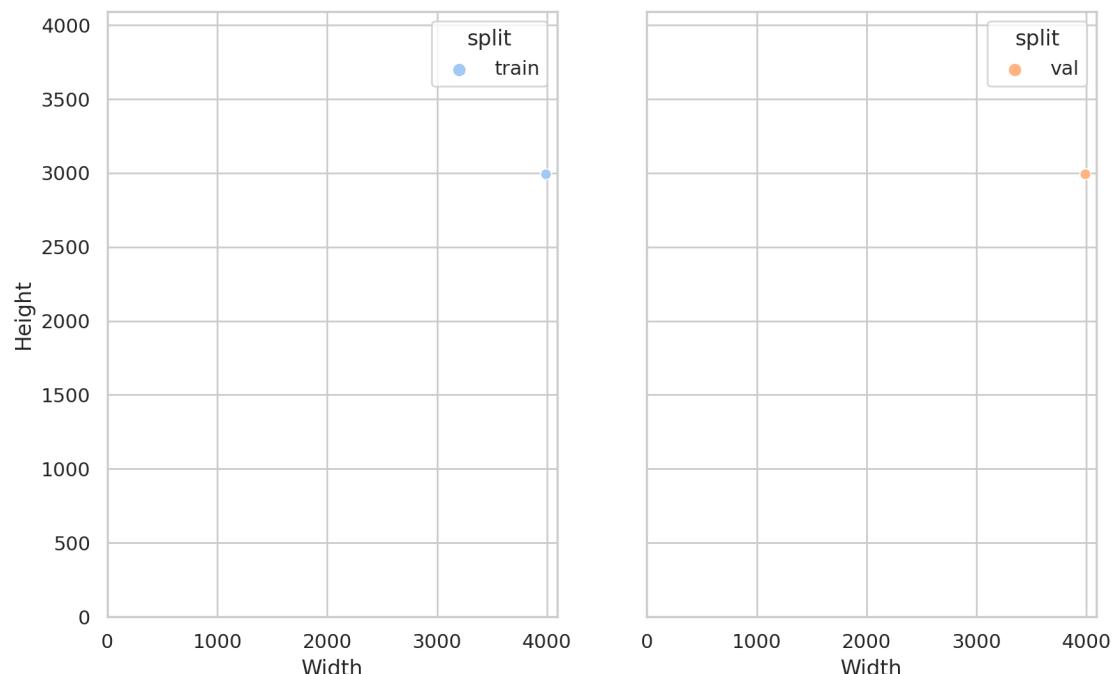
08:12 August 25, 2023

1. Image Features

1.1. General Statistics

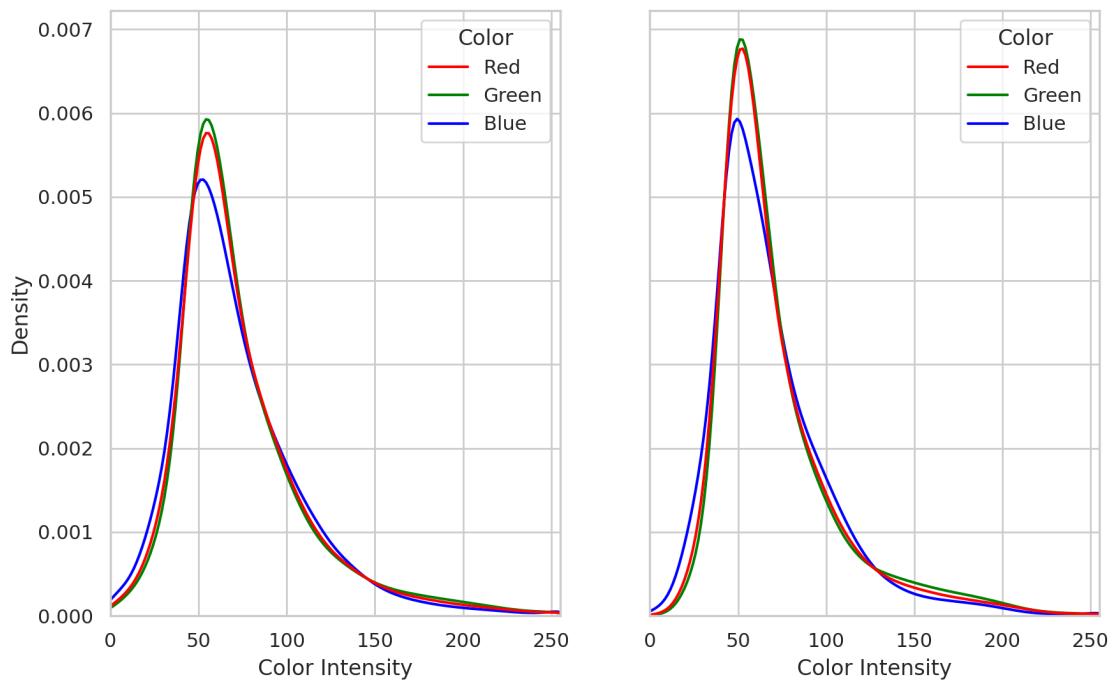
	Train	Validation
Images	155	30
Classes	7	7
Classes in use	1	1
Annotations	734	113
Annotations per images	4.74	3.77
Images with no annotations	0	0
Median image resolution	2992x3992	2992x3992
Smallest annotation	638	638
Largest annotation	115920	26398
Most annotations in an image	12	9
Least annotations in an image	1	1

1.2. Image Width and Height Distribution



These histograms depict the distributions of image height and width. It's important to note that if certain images have been rescaled or padded, the histograms will represent the size after these operations.

1.3. Color Distribution

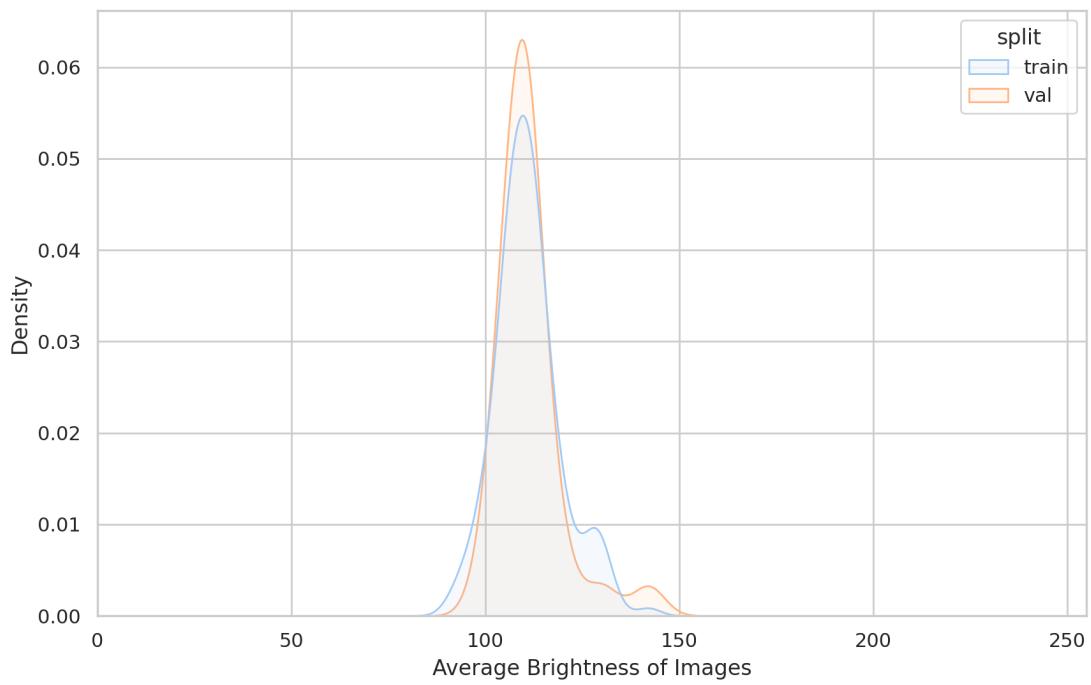


Here's a comparison of RGB or grayscale intensity distributions across the entire dataset, assuming RGB channel ordering.

It can reveal discrepancies in the image characteristics between the two datasets, as well as potential flaws in the augmentation process.

E.g., a notable difference in the mean value of a specific color between the two datasets may indicate an issue with the augmentation process.

1.4. Image Brightness Distribution



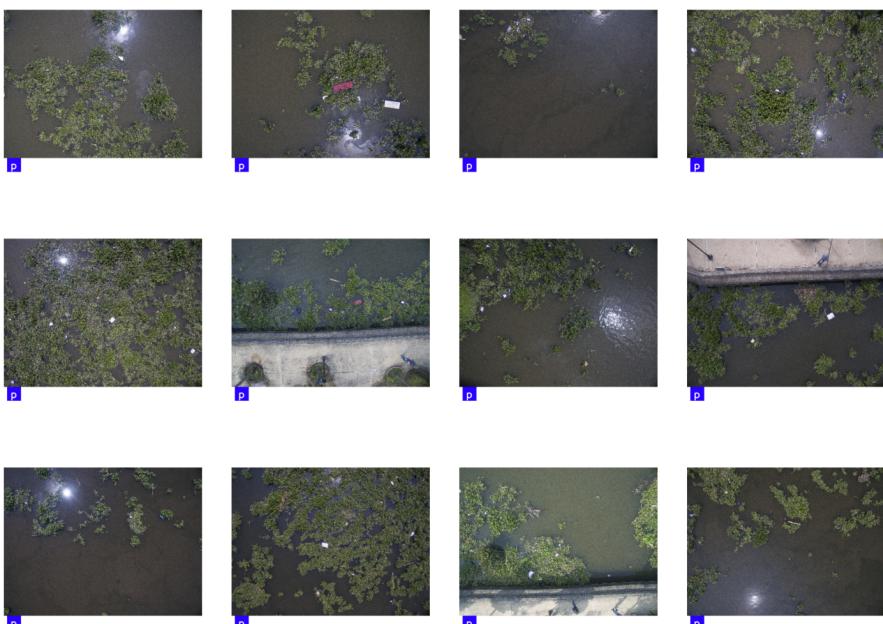
This graph shows the distribution of the brightness levels across all images.

This may for instance uncover differences between the training and validation sets, such as the presence of exclusively daytime images in the training set and nighttime images in the validation set.

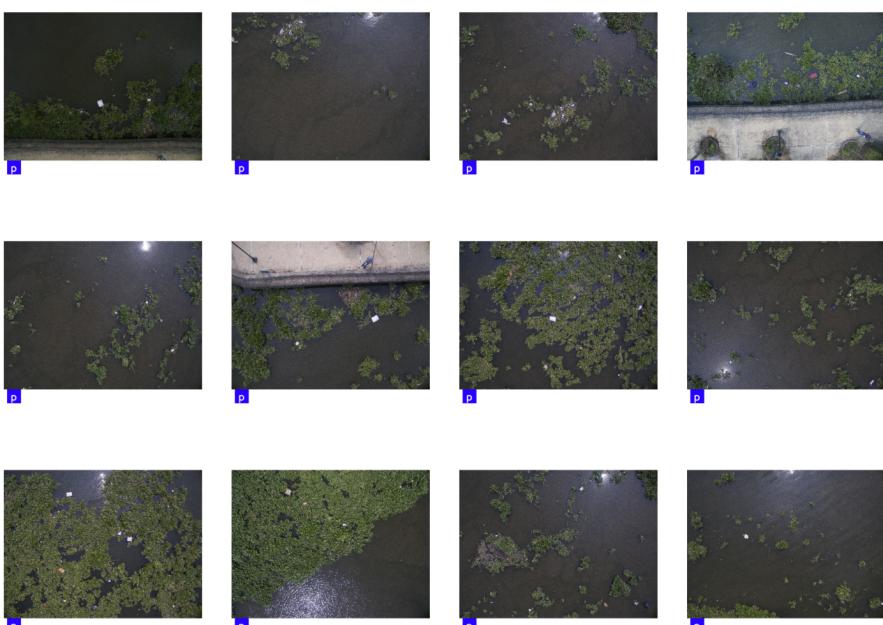
2. Object Detection Features

2.1. Visualization of Samples

train



val



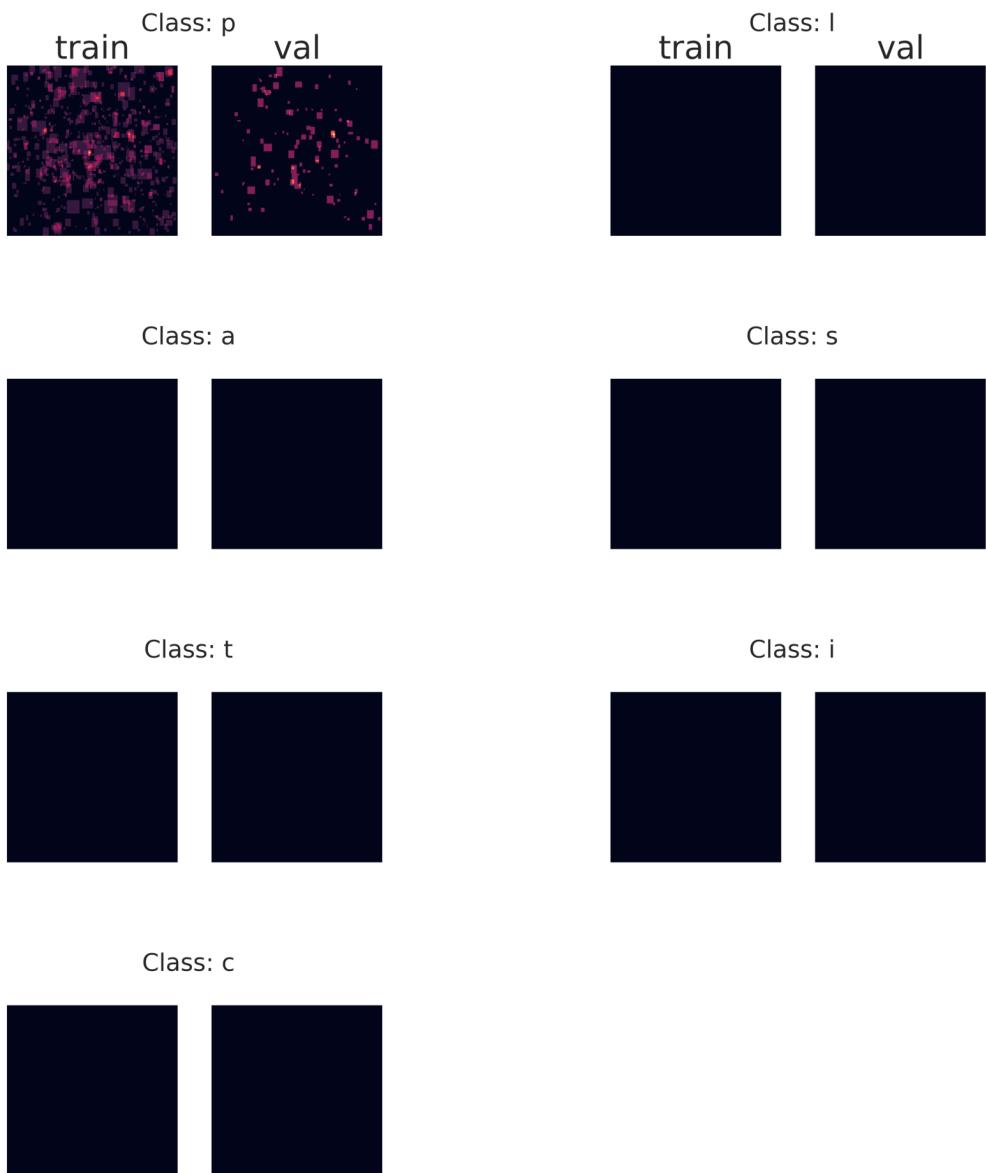
The sample visualization feature provides a visual representation of images and labels. This visualization aids in understanding of the composition of the dataset.



Notice: Only 12 random samples are shown.

You can increase the number of images by changing `n_cols` and `n_rows` in the configuration file.

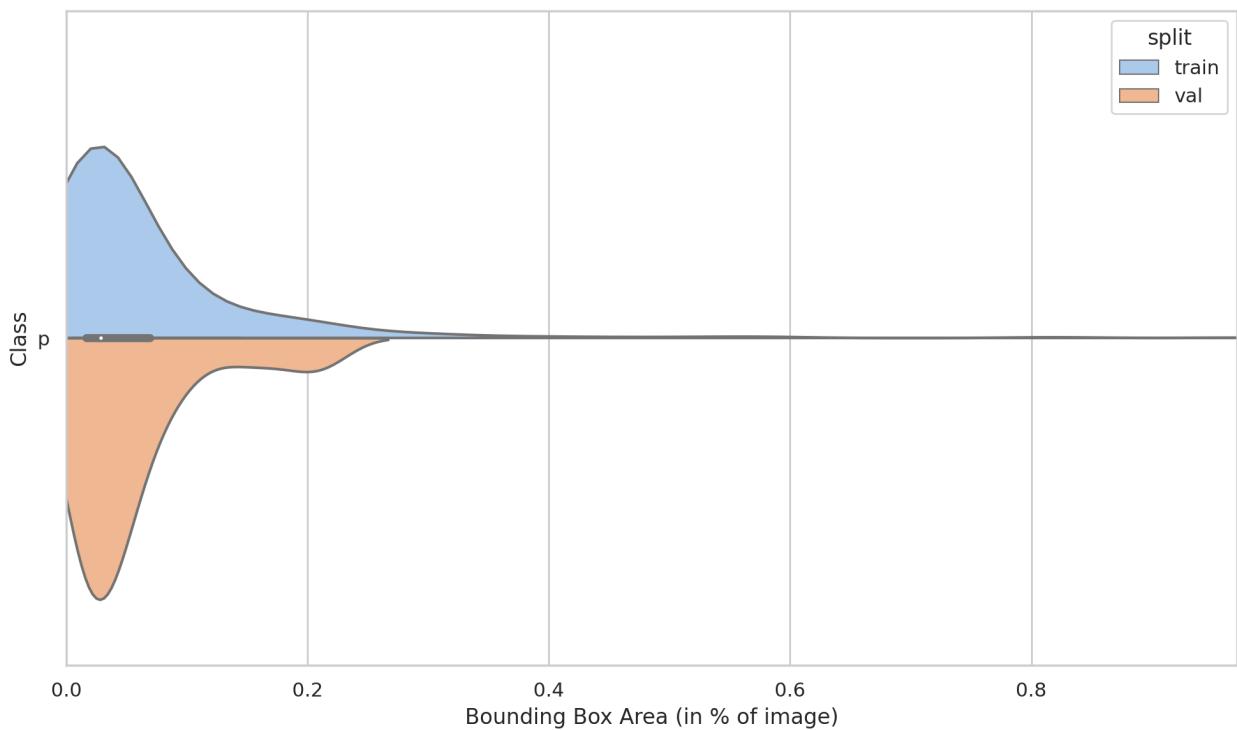
2.2. Bounding Box Density



The heatmap represents areas of high object density within the images, providing insights into the spatial distribution of objects. By examining the heatmap, you can quickly detect whether objects are predominantly concentrated in specific regions or if they are evenly distributed throughout the scene. This information can serve as a heuristic to assess if the objects are positioned appropriately within the expected areas of interest.

Note that images are resized to a square of the same dimension, which can affect the aspect ratio of objects. This is done to focus on localization of objects in the scene (e.g. top-right, center, ...) independently of the original image sizes.

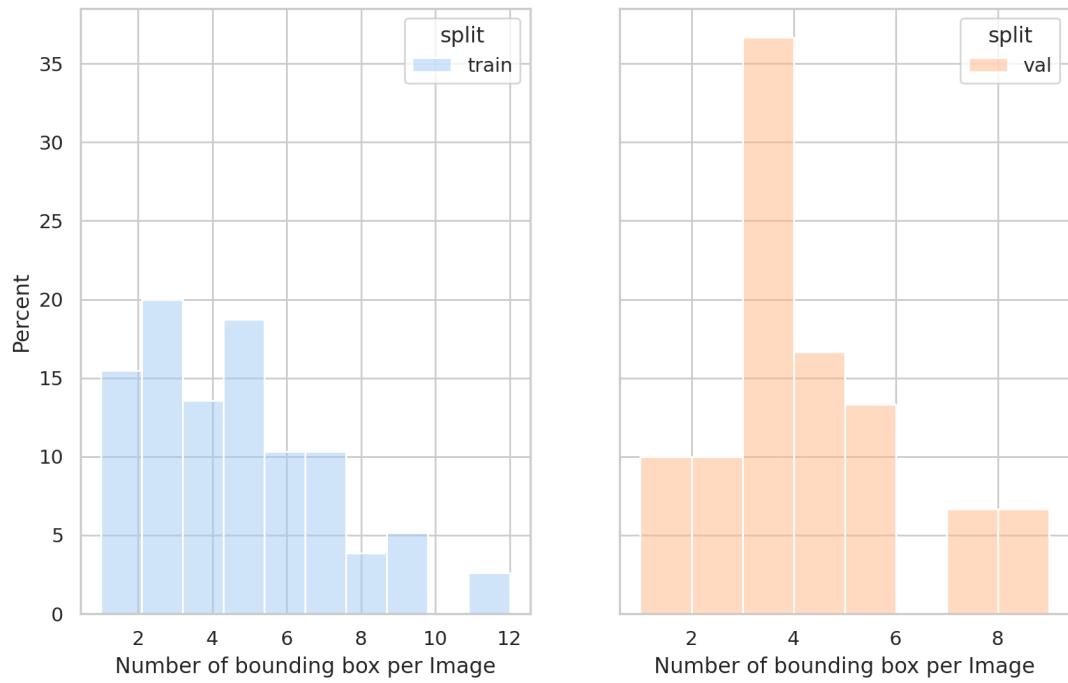
2.3. Distribution of Bounding Box Area



This graph shows the frequency of each class's appearance in the dataset. This can highlight distribution gap in object size between the training and validation splits, which can harm the model's performance.

Another thing to keep in mind is that having too many very small objects may indicate that you are downsizing your original image to a low resolution that is not appropriate for your objects.

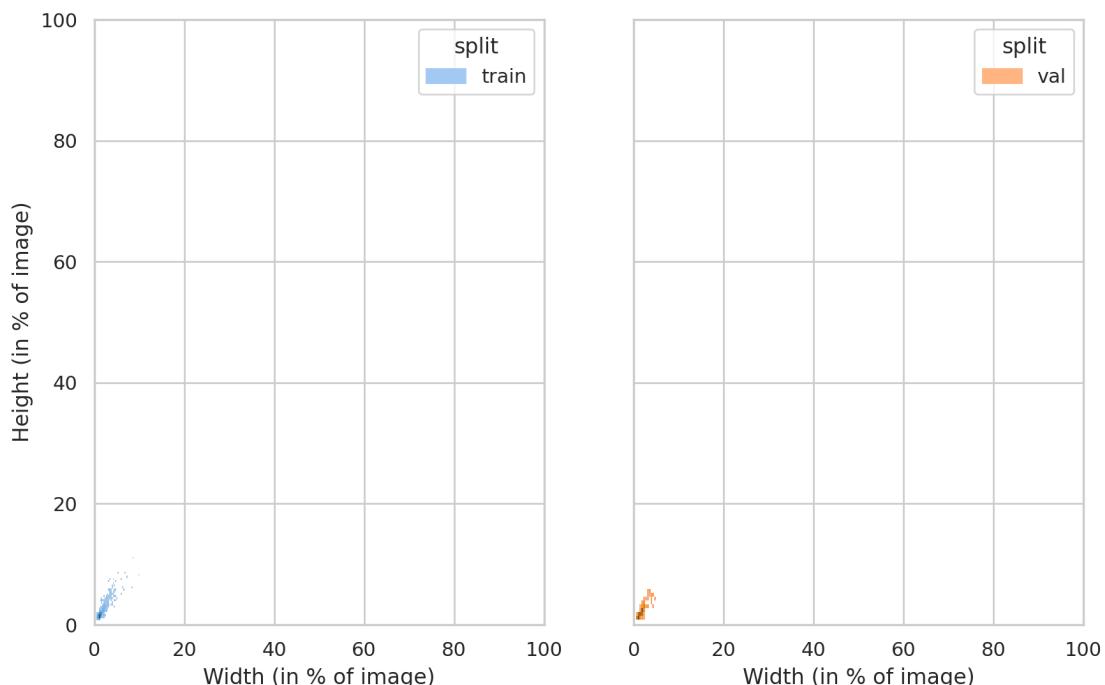
2.4. Distribution of Bounding Box per image



These graphs shows how many bounding boxes appear in images.

This can typically be valuable to know when you observe a very high number of bounding boxes per image, as some models include a parameter to filter the top k results.

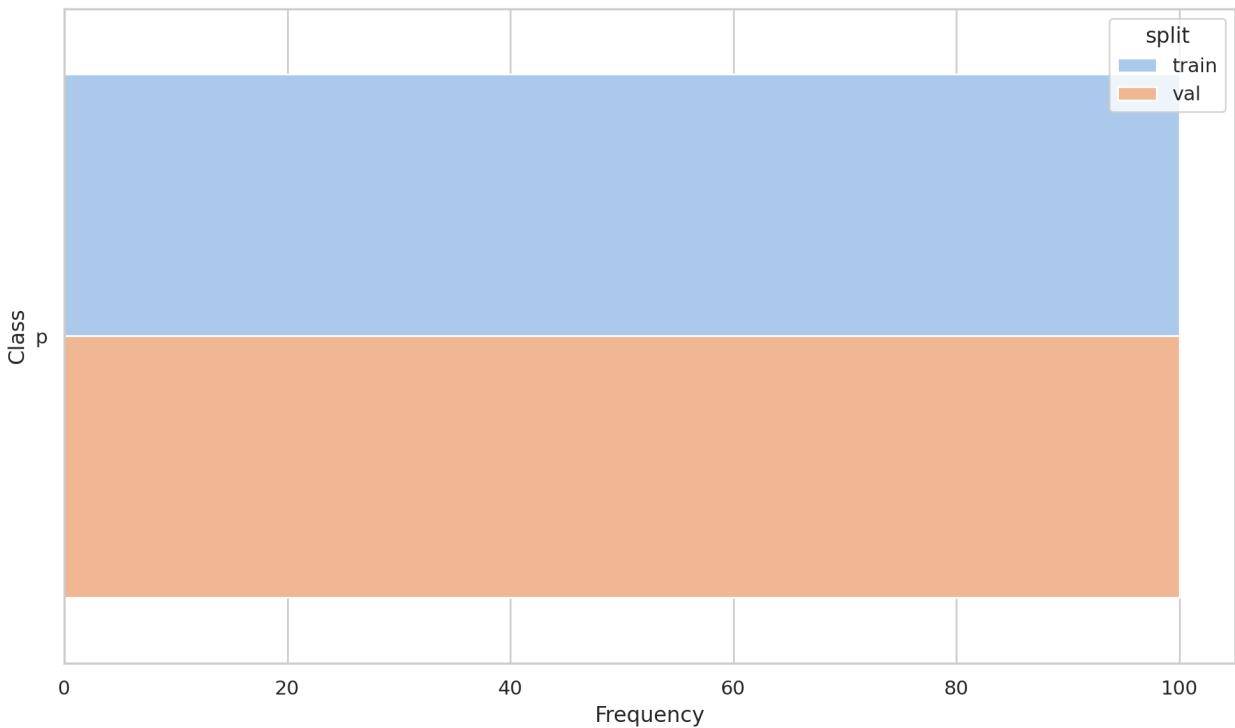
2.5. Distribution of Bounding Box Width and Height



These heat maps illustrate the distribution of bounding box width and height per class.

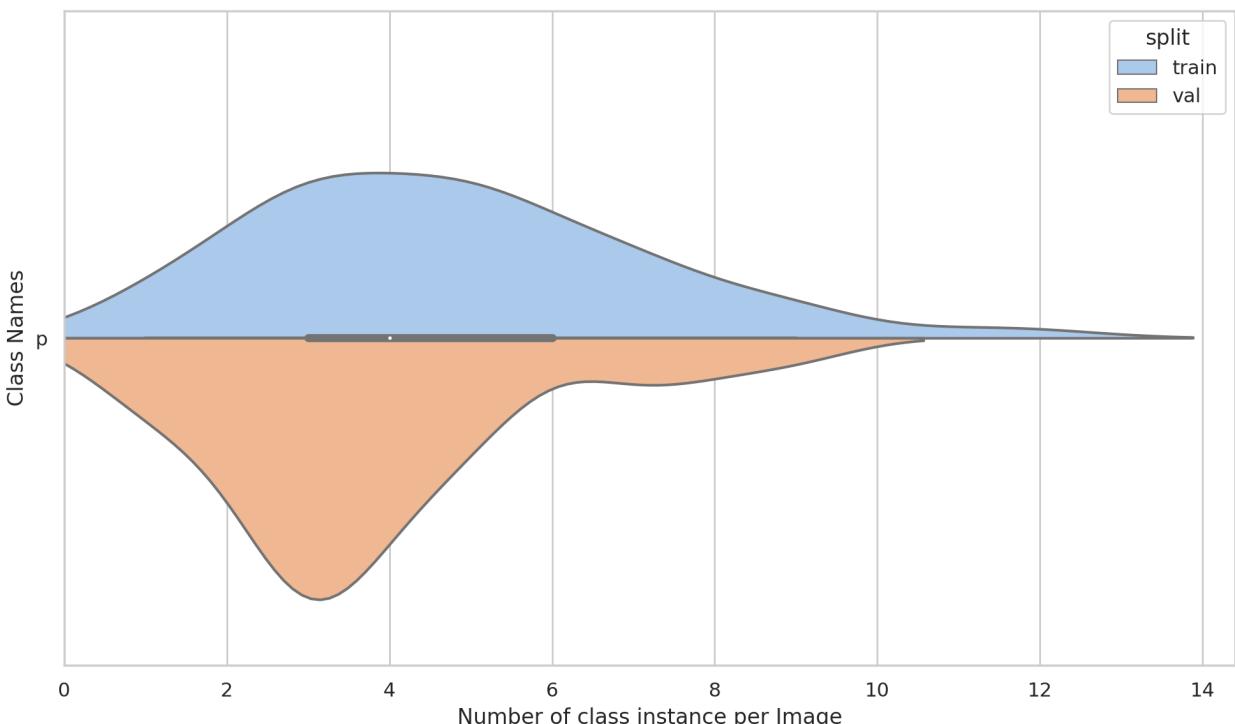
Large variations in object size can affect the model's ability to accurately recognize objects.

2.6. Class Frequency



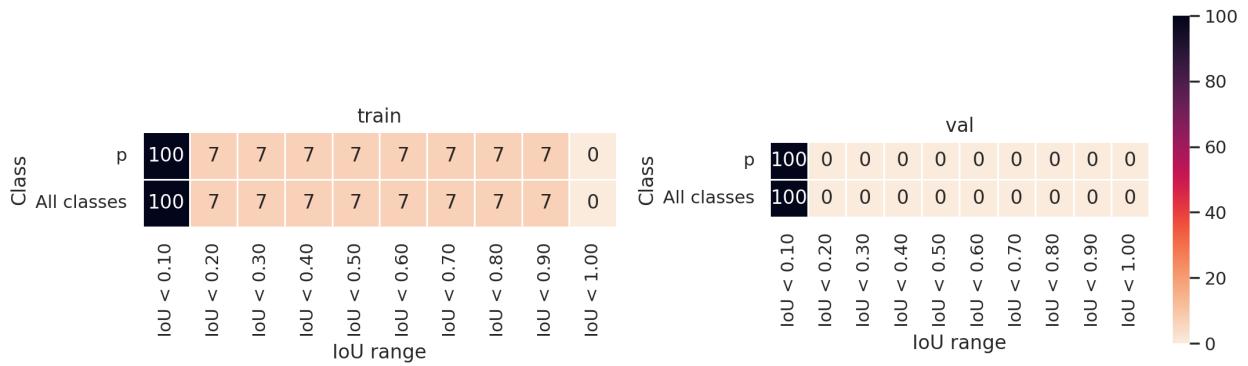
Frequency of appearance of each class. This may highlight class distribution gap between training and validation splits. For instance, if one of the class only appears in the validation set, you know in advance that your model won't be able to learn to predict that class.

2.7. Distribution of Class Frequency per Image



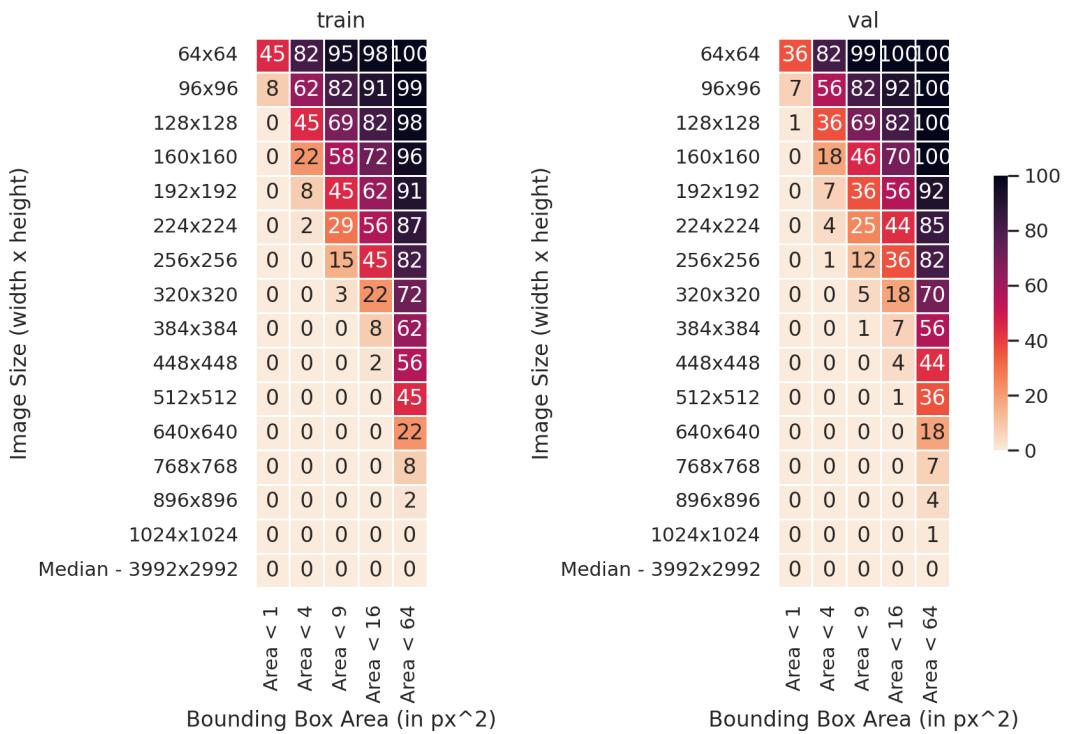
This graph shows how many times each class appears in an image. It highlights whether each class has a constant number of appearances per image, or whether there is variability in the number of appearances from image to image.

2.8. Intersection of Bounding Boxes



The distribution of the box Intersection over Union (IoU) with respect to other boxes in the sample. The heatmap shows the percentage of boxes that overlap with IoU in range [0..T] for each class. Intersection of all boxes is considered (Regardless of classes of corresponding bboxes).

2.9. Distribution of Bounding Boxes smaller than a given Threshold



This visualization demonstrates the consequences of rescaling images on the visibility of their bounding boxes.

By showcasing how bounding box sizes are affected upon varying the image resizing dimensions, we address a critical question: *"How far can we resize an image without causing its bounding boxes to shrink beyond a certain size, especially to less than 1px?"*.

Since an object, when scaled down to less than 1px, essentially disappears from the image, this analysis serves as a guide in identifying the optimal resizing limits that prevent crucial object data loss.

Understanding this is crucial, as inappropriate resizing can result in significant object detail loss, thereby adversely affecting the performance of your model.



Notice: To better understand how to tackle the data issues highlighted in this report, explore our comprehensive [course](#) on analyzing computer vision datasets.