DataGradients™

Training Data 3 Draft

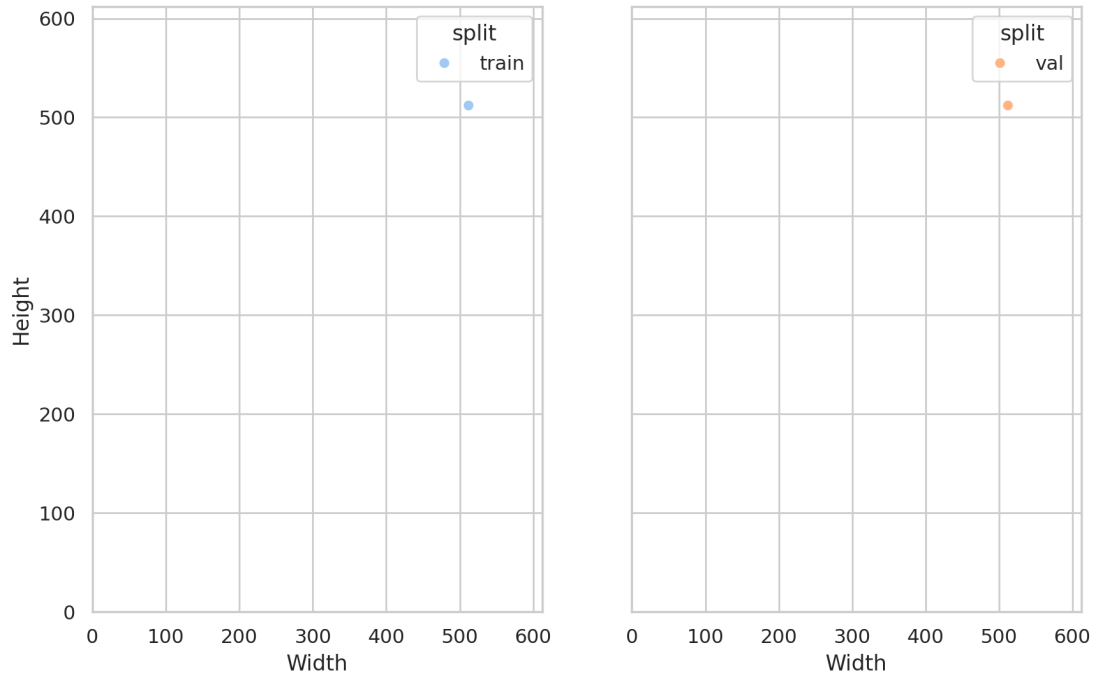
02:13 February 26, 2024

1. Image Features

1.1. General Statistics

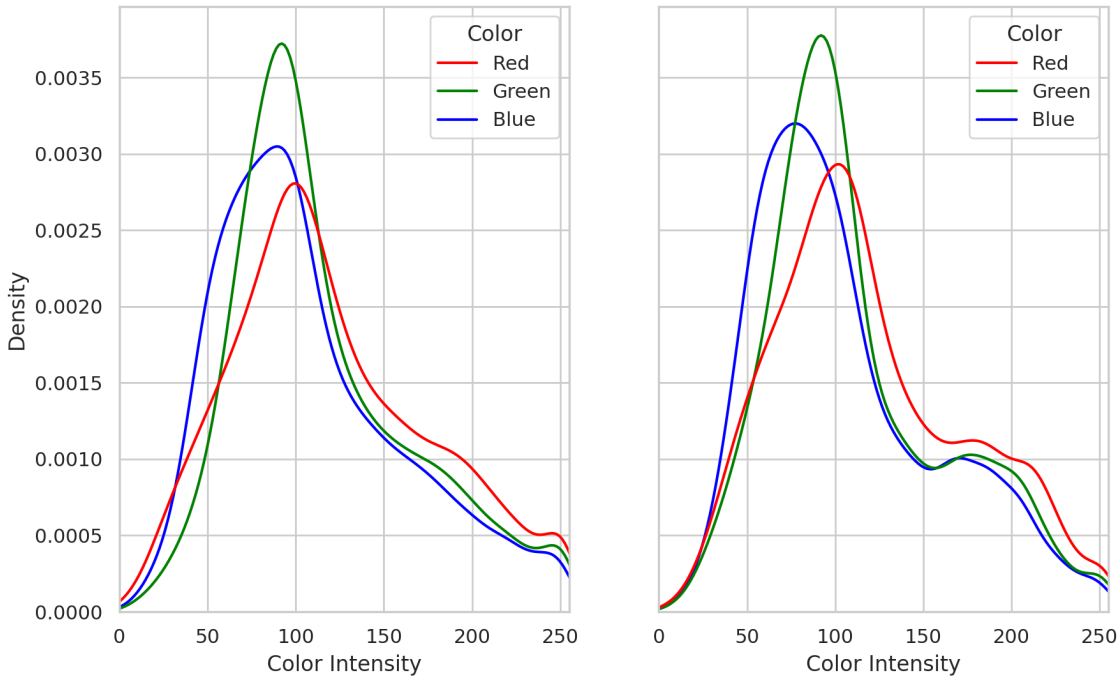
	Train	Validation
Images	39	8
Classes	4	4
Classes in use	4	4
Annotations	1008	103
Annotations per images	25.85	12.88
Images with no annotations	0	0
Median image resolution	512x512	512x512
Smallest annotation	360	750
Largest annotation	74100	77469
Most annotations in an image	69	47
Least annotations in an image	1	3

1.2. Image Width and Height Distribution



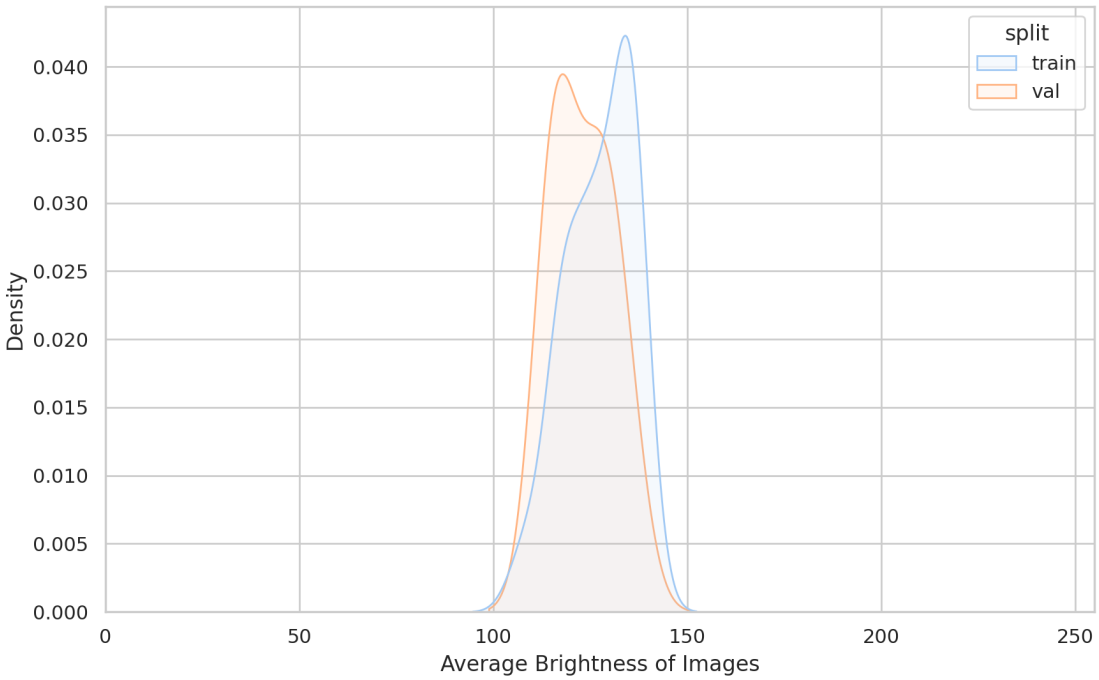
These histograms depict the distributions of image height and width. It's important to note that if certain images have been rescaled or padded, the histograms will represent the size after these operations.

1.3. Color Distribution



Visualize the spread of color intensities with a frequency distribution for each channel, delineated from darkest (0) to brightest (255). By comparing these distributions between training and validation sets, you can identify any significant variations that might affect model performance. For instance, if one dataset shows a higher concentration of darker values, it could suggest a need for lighting correction in preprocessing.

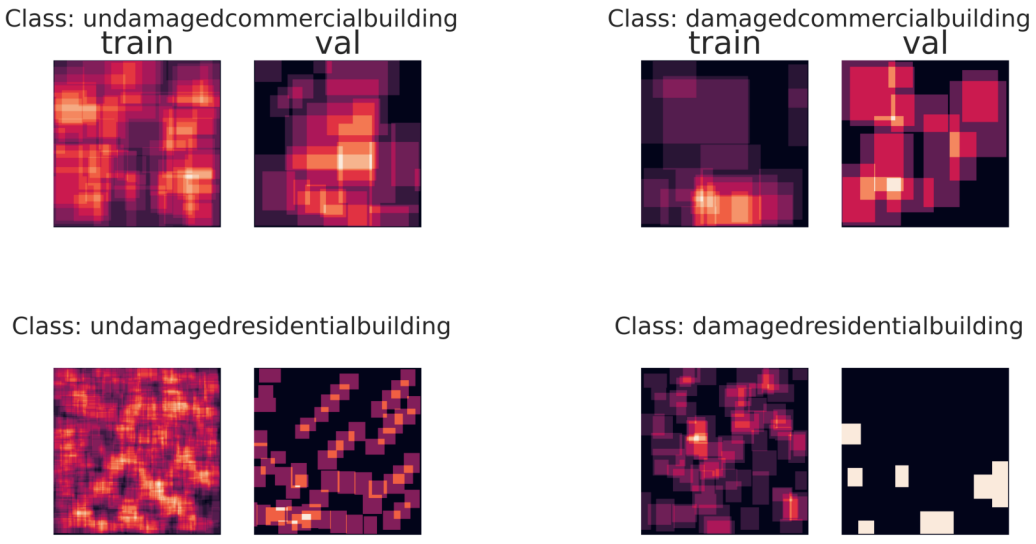
1.4. Image Brightness Distribution



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.

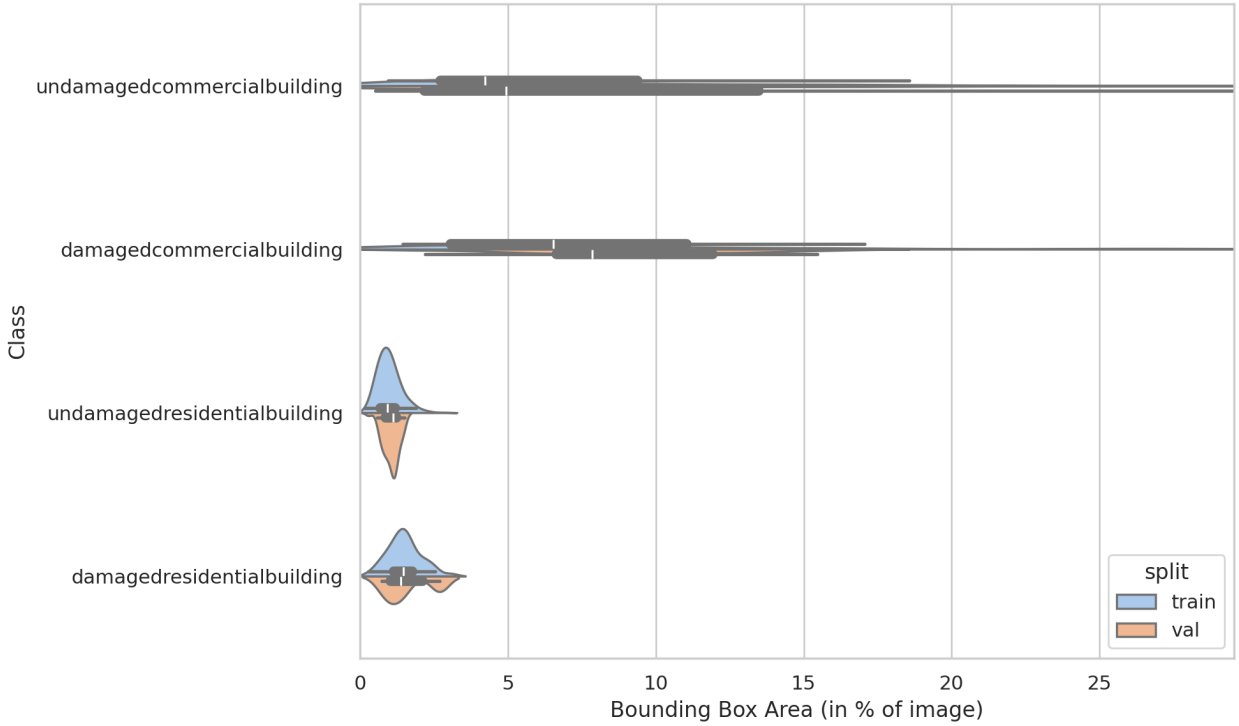
2. Object Detection Features

2.1. Bounding Box Density



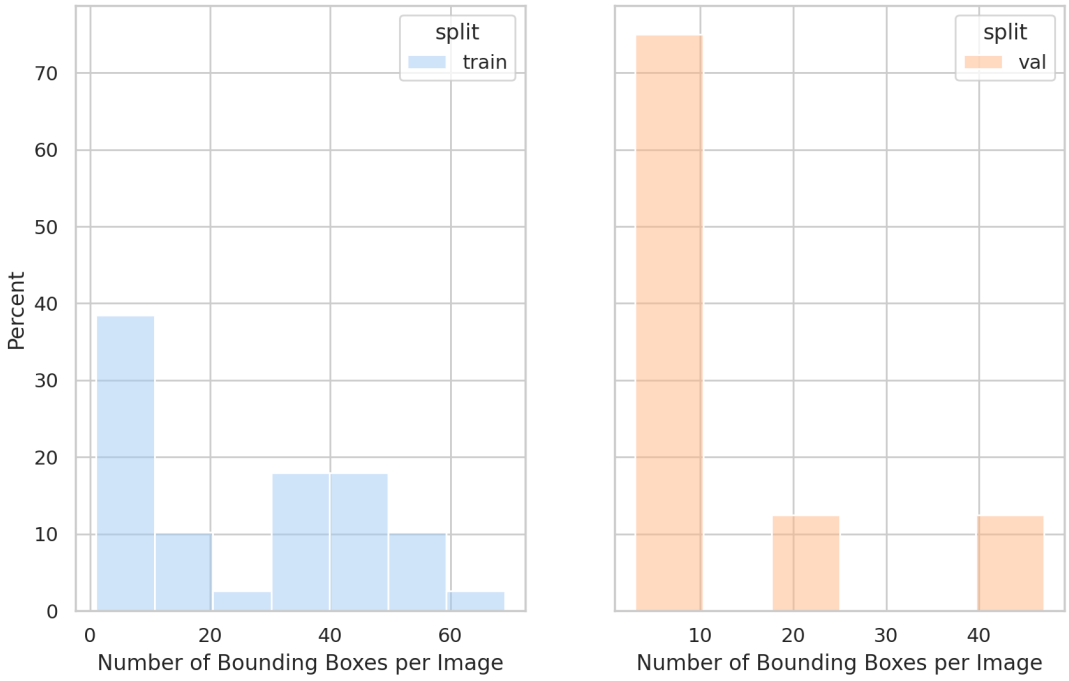
The heatmap represents areas of high object density within the images, providing insights into the spatial distribution of objects. By examining the heatmap, you can quickly detect whether objects are predominantly concentrated in specific regions or if they are evenly distributed throughout the scene. 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.

2.2. Distribution of Bounding Box Area



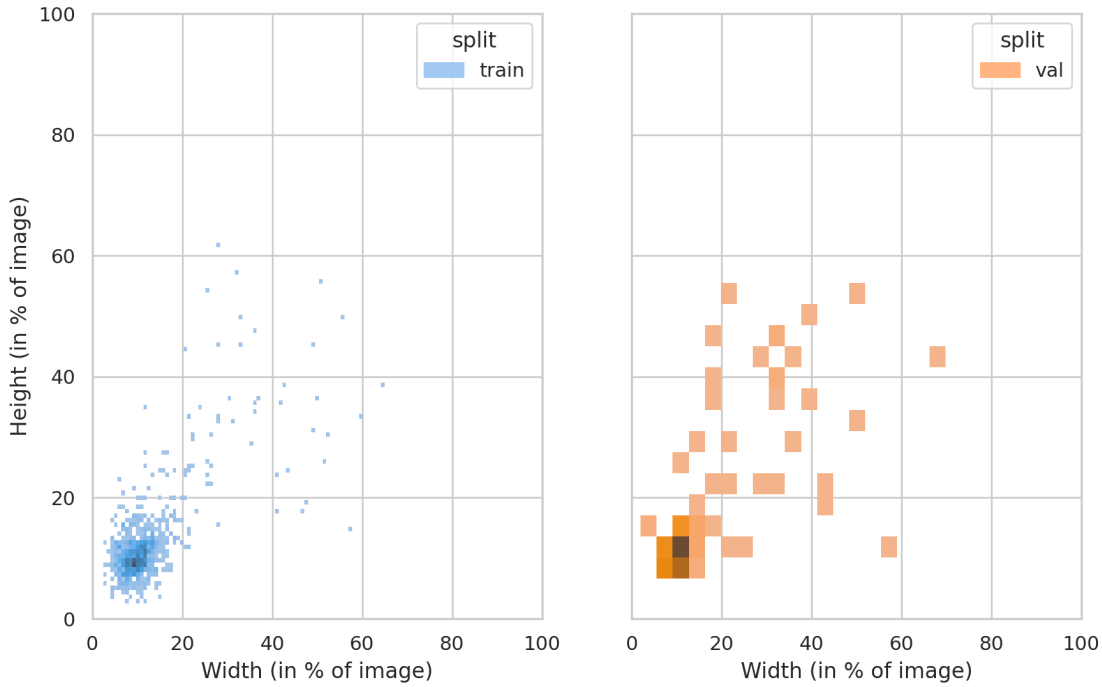
This chart displays the size of bounding boxes relative to their images, offering a clear view of object size variability within the training and validation datasets. It's particularly useful for spotting size imbalances or unusually small or large objects that could affect detection accuracy.

2.3. Distribution of Bounding Box per image



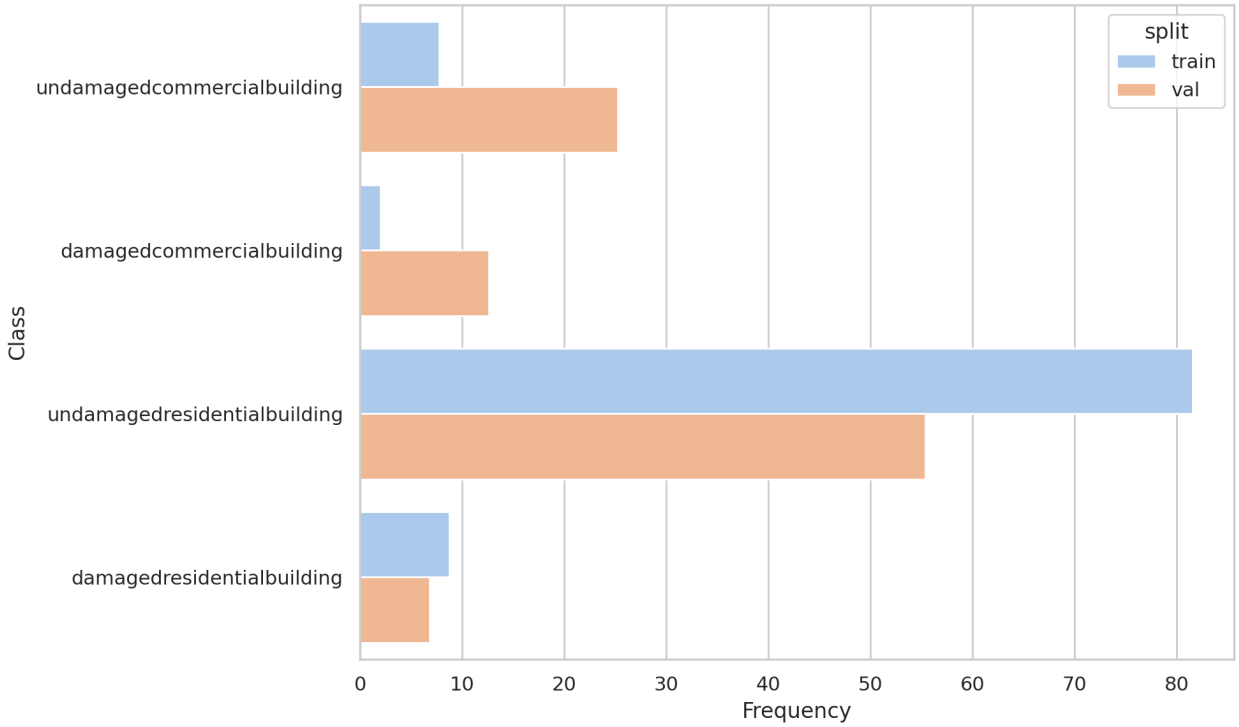
The histograms display the distribution of bounding box counts per image across dataset splits. They help to identify the commonality of bounding box frequencies, which can be instrumental in tuning detection models that process varying numbers of objects per image.

2.4. Distribution of Bounding Box Width and Height



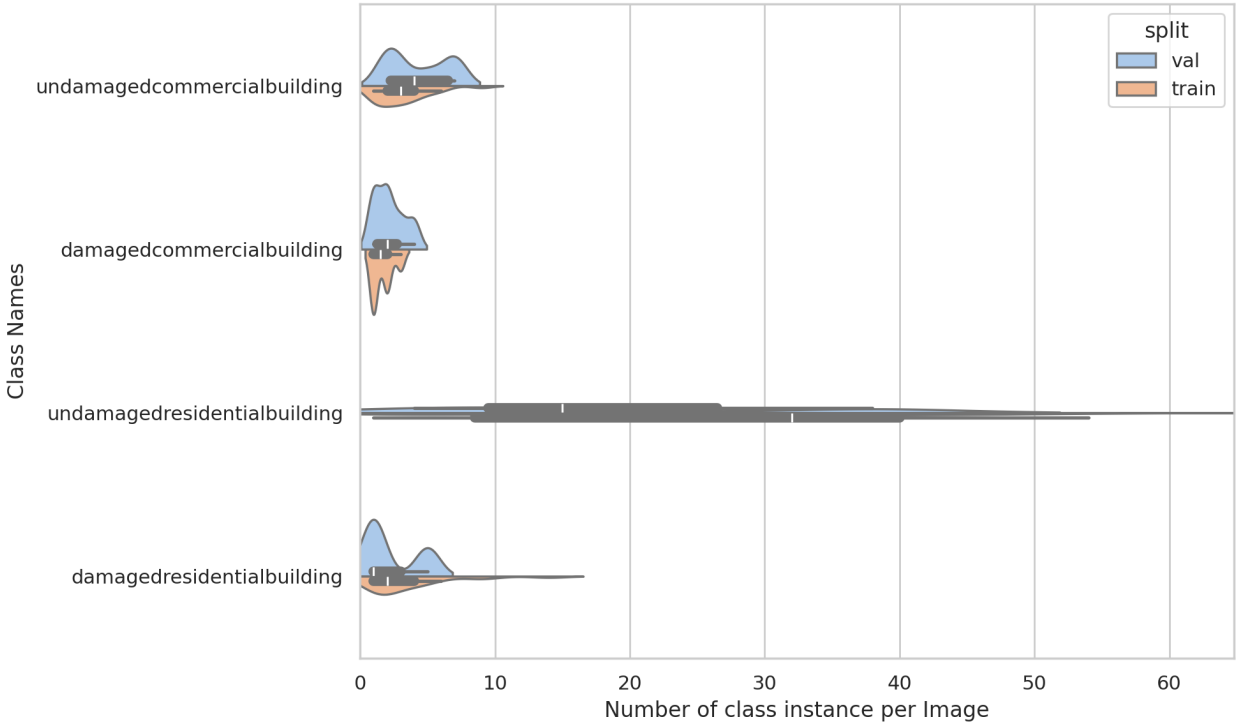
These heat maps illustrate the distribution of bounding box width and height per class.
Large variations in object size can affect the model's ability to accurately recognize objects.

2.5. Class Frequency



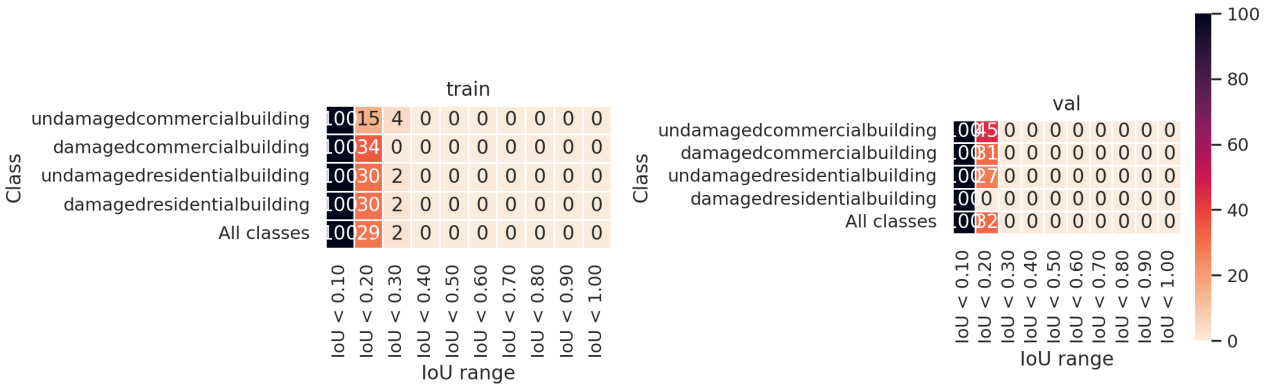
Frequency of appearance of each class. This may highlight class distribution gap between training and validation splits.
For instance, if one of the class only appears in the validation set, you know in advance that your model won't be able to learn to predict that class.

2.6. Distribution of Class Frequency per Image



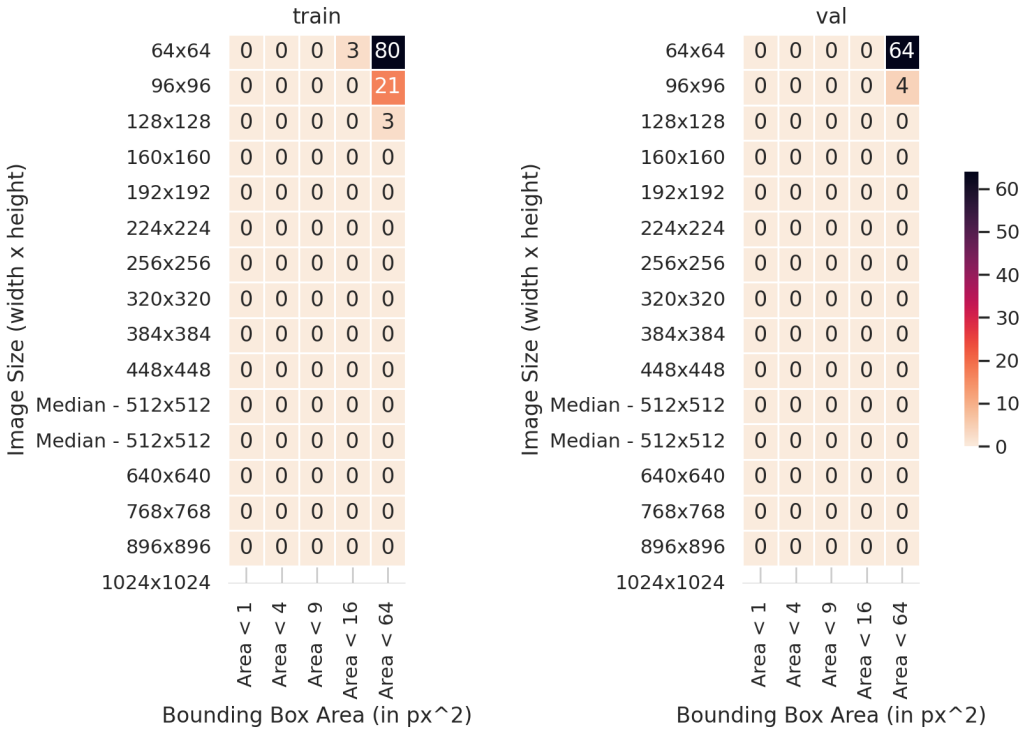
This graph shows how many times each class appears in an image. It highlights whether each class has a constant number of appearances per image, or whether there is variability in the number of appearances from image to image.

2.7. 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).

2.8. 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.