Data analysis

Analysis of rooftiles
Image dataset visualization and preprocessing

Template created by:
Christian Wanschers
Sara Eftekhar Azam
Bastiaan Verheul
Nino van Alphen

March 10, 2024

Table of Contents

1 Visualizations of Dataset

First, it is important to analyze and visualize the properties of the original images. We did this with the following steps...

1.1 BGR channels of original dataset

1

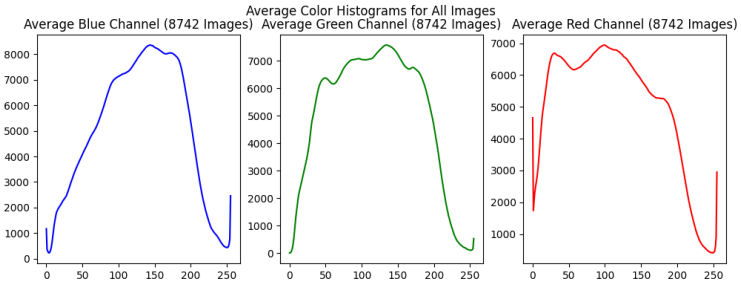

Figure 1: Average colors as colorchannels.

[1]We needed to analyze the color distribution of the colors to understand the average colors used within the dataset. The figure below shows the result of this analysis for the whole dataset and below that examples of 3 random images and their own color channels.

1.2 Visualizing aspect ratios

2

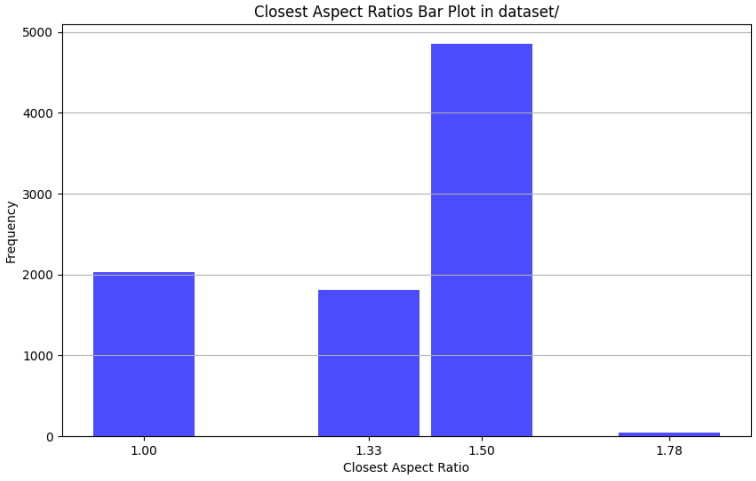

Figure 2: Aspect ratios.

[2]We also need to check the aspect ratios to check whether non-square images were present in our dataset. And as you can see, not all have a square aspect ratio. However, this is required for most and simpler YOLO algorithms

1.3 Visualizing resolutions

Before changing the properties of the images, we also need to check the resolutions to determine if we can safely decrease it to make it easier to work with the model without long loading times which comes with using big files. All images in the dataset we got from the client had roughly had a 4K resolution, which meant each image was around 4- to 6MB in size. This meant the total dataset was around 75GB for 7800 images. The size means that the dataset is difficult to process in code and the model generally does not require images with a resolution that large. We decided to scale the images down when the resolution was of a higher value than 3000 in either width or length. This meant a file size decrease of about 94%. This resulted in the total size of the dataset decreasing to less than 4GB and thus being much easier to process. The figure below shows the result resolution rescaling.

3

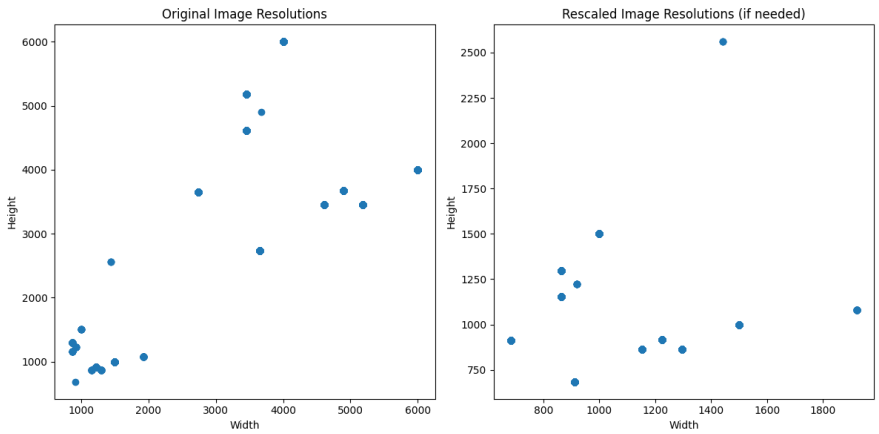

Figure 3: Resolutions of images before & after

[3] Rescaling was done immediatly since the 8742 images within the dataset resulted in a size of over 75GB which isn't easy to work with.

2 Changes to Dataset

We need to change some properties to prepare the dataset before feeding it to a model.

2.1 Updates to resolution

As we mentioned earlier, we decreased the resolution of all images in the whole dataset to make it more manageable. See 4.

2.2 File extension formatting

Most images were of .JPG format and some of .JPEG and .PNG format. To make all these images the same file format we chose to convert them all to the PNG format.

2.3 Applying grayscale

2.4 Applying normalization

2.5 Applying CLAHE

2.6 Updates to aspect ratio

4 4

Figure 4: Resolutions of images before & after

2.7 Data augmentation

2.7.1 Flipping

2.7.2 Rotating

2.7.3 Mirroring

2.7.4 Zoom

2.8 Splitting data

2.8.1 Training

2.8.2 Testing

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam auctor mi risus, quis tempor libero hendrerit at. Duis hendrerit placerat quam et semper. Nam ultricies metus vehicula arcu viverra, vel ullamcorper justo elementum. Pellentesque vel mi ac lectus cursus posuere et nec ex. Fusce quis mauris egestas lacus commodo venenatis. Ut at arcu lectus. Donec et urna nunc. Morbi eu nisl cursus sapien eleifend tincidunt quis quis est. Donec ut orci ex. Praesent ligula enim, ullamcorper non lorem a, ultrices volutpat dolor. Nullam at imperdiet urna. Pellentesque nec velit eget est pretium.[5]

[5] This is a sidenote. This template features a large margin specifically so you can put notes, figures, tables and other things into it as additional material to the main content in the text block.

Donec in elit ac ante vestibulum rhoncus. Pellentesque ligula tortor, aliquet malesuada nulla tristique vitae. Aliquam mi sem, varius eu pellentesque et, tristique nec quam. Vestibulum pellentesque in dui et venenatis. Sed malesuada elit pellentesque sapien aliquet porta. In at facilisis diam. Duis id ante tellus.[6]

[6] This sidenote has been pushed down the page manually with an optional parameter, otherwise it would be right under the one above.

2.9 Subsection Title

In diam libero, vulputate quis accumsan non, auctor in ipsum. Praesent cursus velit eget lacus sodales porta. Proin quis risus ut velit euismod scelerisque ut sed neque. Cras sagittis, dolor ac ullamcorper auctor, tortor dui facilisis diam, at sagittis nisi ipsum a neque. Nullam vel mattis nisi. Ut interdum ut diam at ornare. Nulla ultrices elit justo, vitae tristique massa vulputate sit amet.

This sidenote isn't numbered in the text or margin. This is useful for notes that apply anywhere on the page instead of one particular place.

Vestibulum erat felis, cursus vitae convallis ac, commodo eu nisi. Nulla facilisi. Mauris dignissim nisi felis, a mollis ex accumsan vel. Suspendisse bibendum vitae nibh in suscipit. Vestibulum et finibus eros. Nulla facilisi. Cras luctus aliquam

finibus. In nec justo nec orci malesuada faucibus.

2.9.1 Subsubsection Title

This is an example of a full width paragraph... Curabitur id placerat orci. Vivamus pulvinar augue ac feugiat blandit. Donec in ultricies mi. Nam eu lacus ac augue aliquet consectetur. Praesent dui risus, sollicitudin nec felis ut, posuere ultricies dolor. Sed massa nulla, dignissim eget sem sit amet, eleifend fermentum dui. Phasellus consequat sem vel turpis finibus, a aliquam risus malesuada.

Maecenas consectetur metus at tellus finibus condimentum. Proin arcu lectus, ultrices non tincidunt et, tincidunt ut quam. Integer luctus posuere est, non maximus ante dignissim quis. Nunc a cursus erat. Curabitur suscipit nibh in tincidunt sagittis. Nam malesuada vestibulum quam id gravida. Proin ut dapibus velit. Vestibulum eget quam quis ipsum semper convallis. Duis consectetur nibh ac diam dignissim, id condimentum enim dictum. Nam aliquet ligula eu magna pellentesque, nec sagittis leo lobortis. Aenean tincidunt dignissim egestas. Morbi efficitur risus ante, id tincidunt odio pulvinar vitae.

Paragraph Title Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam auctor mi risus, quis tempor libero hendrerit at. Duis hendrerit placerat quam et semper. Nam ultricies metus vehicula arcu viverra, vel ullamcorper justo elementum. Pellentesque vel mi ac lectus cursus posuere et nec ex.

The section titles below show how multi-line section titles look at the 3 top levels.

3 Fusce eleifend porttitor arcu, id accumsan elit pharetra eget

Section, subsection and subsubsection titles can span multiple lines, as shown here. Make sure to put a shorter version of these long titles in the optional parameter to the section commands so the title output to the table of contents is the short version.

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

3.1 Phasellus sit amet enim efficitur, aliquam nulla id, lacinia mauris viverra libero ac magna

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

3.1.1 In mi mauris, finibus non faucibus non, imperdiet nec leo. In erat arcu, tincidunt nec aliquam et, volutpat eget

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

4 Figure Examples

This statement automatically references the figure below using its label: Figure **??**.