

```
In [ ]: !pip install pandas
!pip install basic-image-eda
!pip install scikit-image
import pandas as pd
```

```
In [10]: from basic_image_eda import BasicImageEDA

data_dir = "./data/train22A/train2_new"
# BasicImageEDA.explore(data_dir)
extensions = ['png', 'jpg', 'jpeg']
threads = 0
dimension_plot = True
channel_hist = True
nonzero = False
hw_division_factor = 1.0
print('train images: ')
BasicImageEDA.explore(data_dir, extensions, threads, dimension_plot, channel_hist, nonzero, hw_d:
```

```
train images:

found 7000 images.
Using 8 threads. (max:8)

100%|██████████████████| 7000/7000 [00:09<00:00, 775.91it/s]
*-----*
number of images                | 7000

dtype                           | uint8
channels                        | [3]
extensions                      | ['jpg']

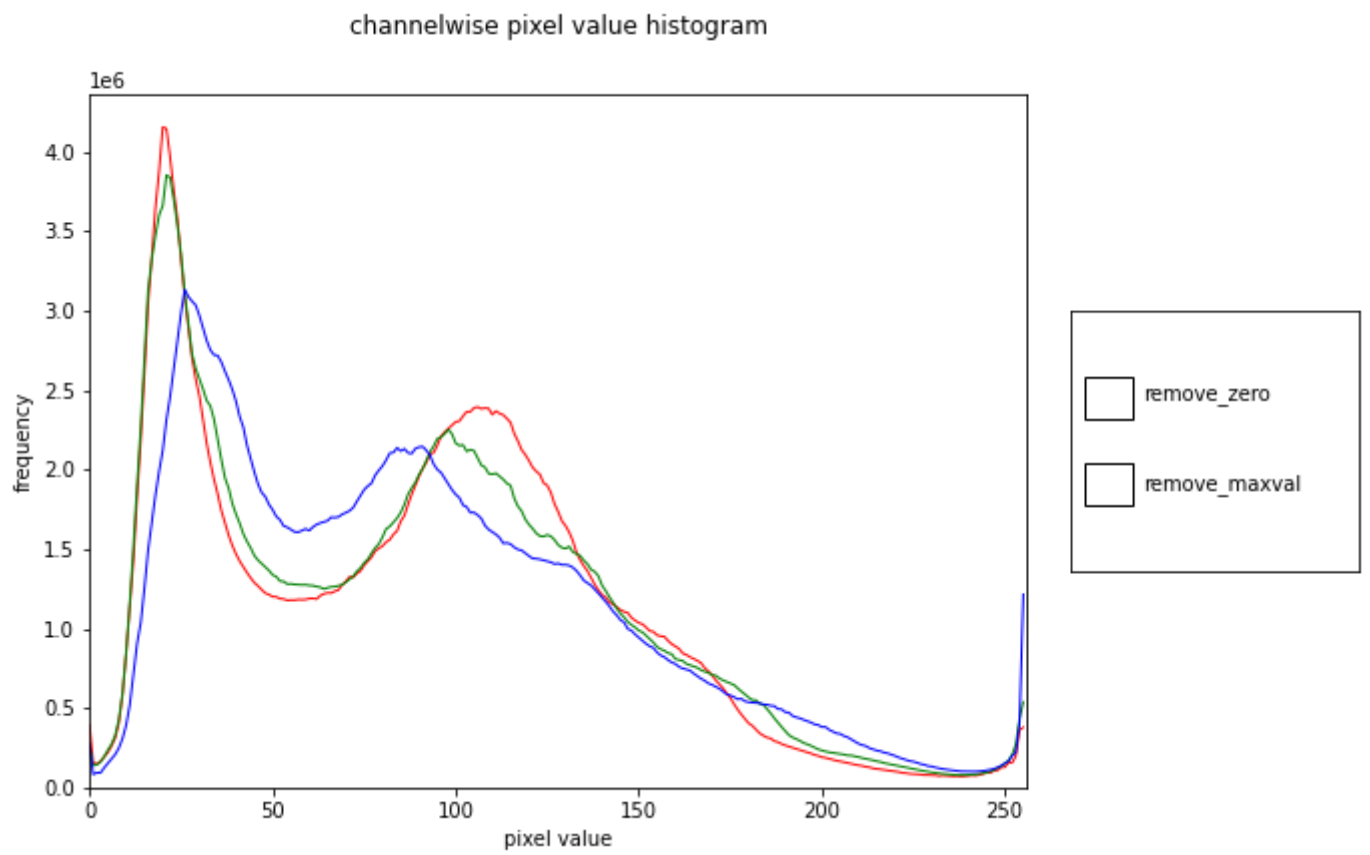
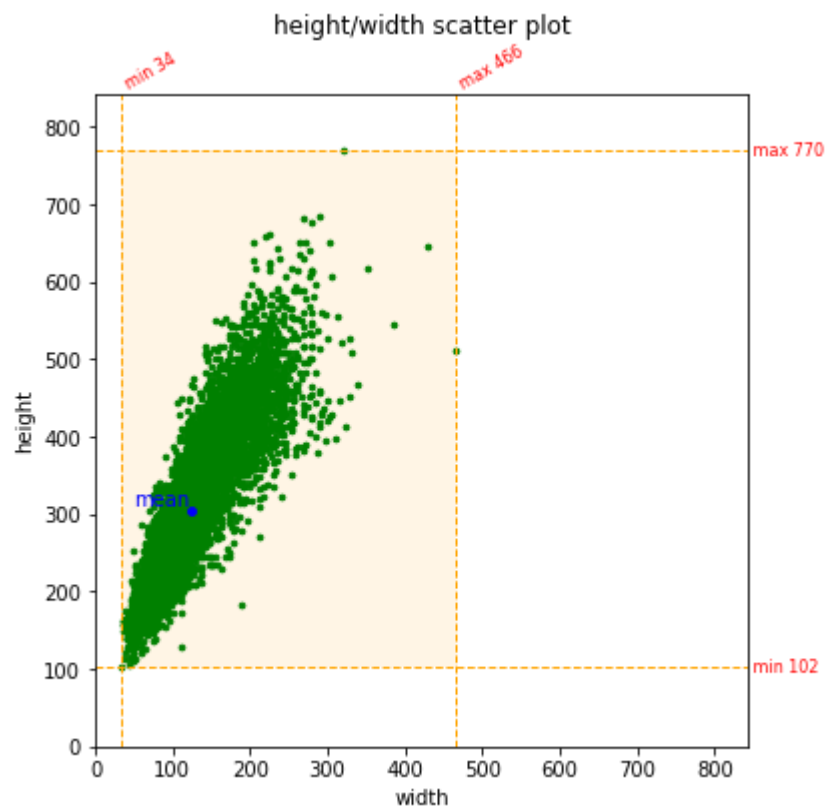
min height                      | 102
max height                      | 770
mean height                     | 305.29742857142855
median height                   | 288

min width                       | 34
max width                       | 466
mean width                      | 123.07514285714285
median width                    | 112

mean height/width ratio         | 2.4805774869243646
median height/width ratio       | 2.5714285714285716
recommended input size(by mean) | [304 120] (h x w, multiples of 8)
recommended input size(by mean) | [304 128] (h x w, multiples of 16)
recommended input size(by mean) | [320 128] (h x w, multiples of 32)

channel mean(0~1)               | [0.36166626 0.3658852 0.3834917 ]
channel std(0~1)                | [0.23254155 0.24040939 0.24047028]
*-----*

eda ended in 00 hours 00 minutes 09 seconds
```



```
Out[10]: {'dtype': 'uint8',
'channels': [3],
'extensions': ['jpg'],
'min_h': 102,
'max_h': 770,
'mean_h': 305.29742857142855,
'median_h': 288,
'min_w': 34,
'max_w': 466,
'mean_w': 123.07514285714285,
'median_w': 112,
'mean_hw_ratio': 2.4805774869243646,
'median_hw_ratio': 2.5714285714285716,
'rec_hw_size_8': array([304, 120]),
'rec_hw_size_16': array([304, 128]),
'rec_hw_size_32': array([320, 128]),
```

```
'mean': array([0.36166626, 0.3658852 , 0.3834917 ], dtype=float32),  
'std': array([0.23254155, 0.24040939, 0.24047028], dtype=float32)}
```

In [11]:

```
from basic_image_eda import BasicImageEDA  
  
data_dir = "../data/test22A_new/"  
# BasicImageEDA.explore(data_dir)  
extensions = ['png', 'jpg', 'jpeg']  
threads = 0  
dimension_plot = True  
channel_hist = True  
nonzero = False  
hw_division_factor = 1.0  
print('test images: ')  
BasicImageEDA.explore(data_dir, extensions, threads, dimension_plot, channel_hist, nonzero, hw_d
```

test images:

found 2000 images.

Using 8 threads. (max:8)

100%|██████████████████| 2000/2000 [00:02<00:00, 718.56it/s]

\*-----\*

number of images		2000
------------------	--	------

dtype		uint8
channels		[3]
extensions		['jpg']

min height		150
max height		714
mean height		338.116
median height		326

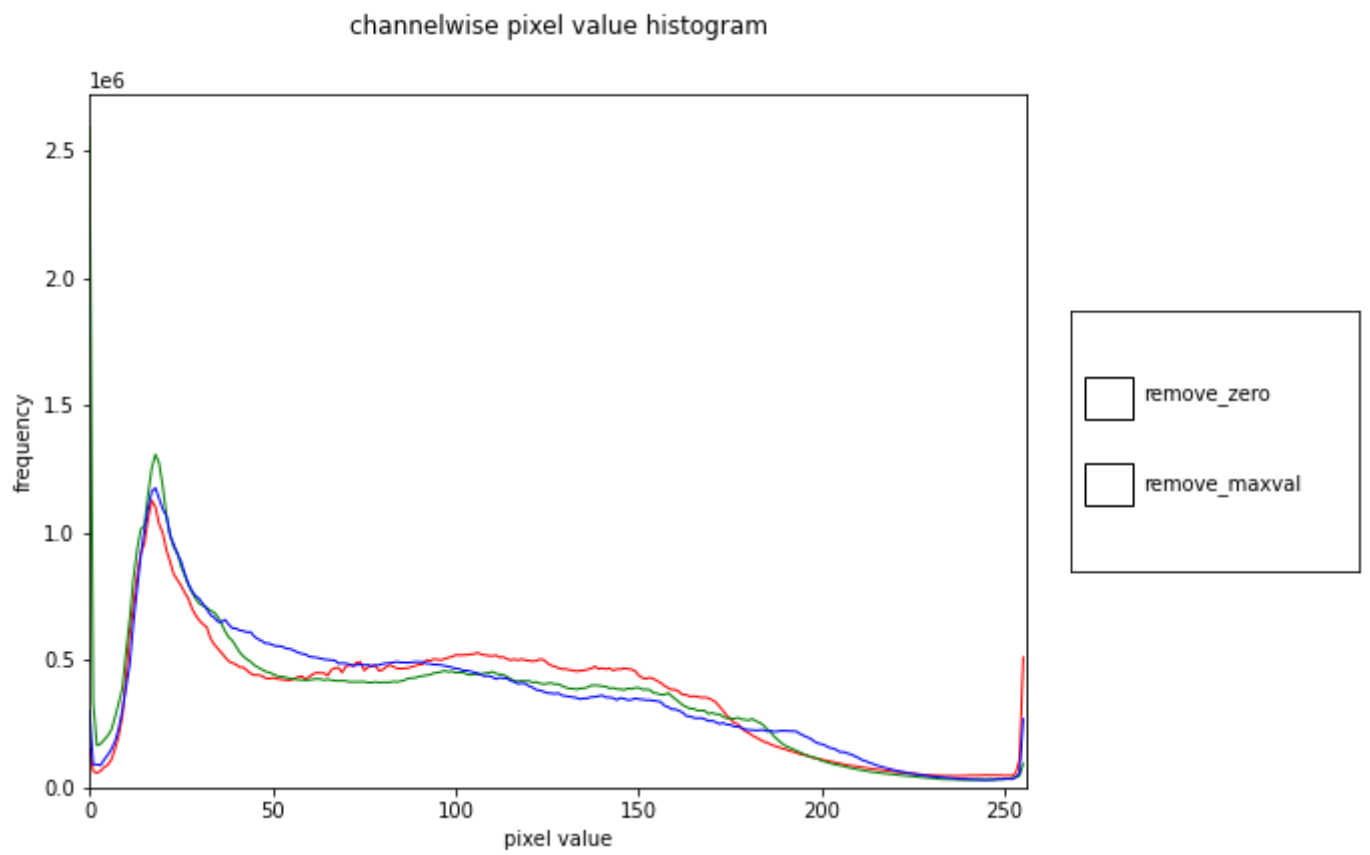
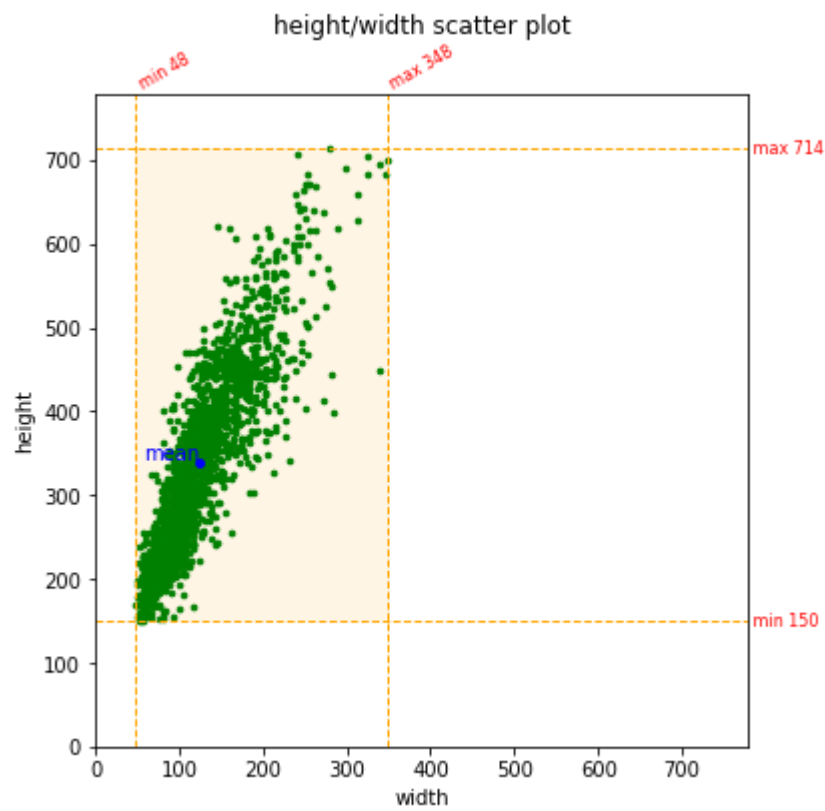
min width		48
max width		348
mean width		125.258
median width		116

mean height/width ratio		2.699356528125948
median height/width ratio		2.810344827586207
recommended input size(by mean)		[336 128] (h x w, multiples of 8)
recommended input size(by mean)		[336 128] (h x w, multiples of 16)
recommended input size(by mean)		[352 128] (h x w, multiples of 32)

channel mean(0~1)		[0.36428225 0.3349935 0.3507986 ]
channel std(0~1)		[0.23110583 0.23254822 0.23205678]

\*-----\*

eda ended in 00 hours 00 minutes 02 seconds



```
Out[11]: {'dtype': 'uint8',
'channels': [3],
'extensions': ['jpg'],
'min_h': 150,
'max_h': 714,
'mean_h': 338.116,
'median_h': 326,
'min_w': 48,
'max_w': 348,
'mean_w': 125.258,
'median_w': 116,
'mean_hw_ratio': 2.699356528125948,
'median_hw_ratio': 2.810344827586207,
'rec_hw_size_8': array([336, 128]),
'rec_hw_size_16': array([336, 128]),
'rec_hw_size_32': array([352, 128]),
```

```
'mean': array([0.36428225, 0.3349935 , 0.3507986 ], dtype=float32),  
'std': array([0.23110583, 0.23254822, 0.23205678], dtype=float32)}
```