

Final_Project

December 9, 2024

0.1 # Final Project

DSC-360

Brian Chaffee & Leo Sanchez

Date: 12/09/2024

0.2 Intro:

0.2.1 Problem:

Large scale image processing for machine learning applications. - Art institutions contain a large amount of works - Machine Learning can be used to better classify works for their associated genre - The Machine Learning algorithms require a large amount of information to accurately train the model - Complex process of transforming the image into tangible information

Dataset: <https://www.kaggle.com/datasets/ipythonx/wikiart-gangogh-creating-art-gan>

0.3 Imports

```
[1]: # Basic Imports
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import os
import time

from pathlib import Path

# Image Libraries
import cv2
from PIL import Image
```

```
[2]: # Dask Imports
import dask
import dask_image

import dask_image.imread
```

```
import dask_image.ndfilters
import dask_image.ndmeasure
import dask.array as da
```

0.4 # Image Opening

0.5 Example Image Opening

```
[3]: %%time
output_filename = os.path.join('/users/PSS0009/brach21/final_project/data/
↳images/animal-painting', 'dd02a8c8f0ca6567ff33563bf9ea42e9c.jpg')
output_filename2 = os.path.join('/users/PSS0009/brach21/final_project/data/
↳images/animal-painting', '55a592b632d75ae6eab87f418dc7d962c.jpg')
output_filename3 = os.path.join('/users/PSS0009/brach21/final_project/data/
↳images/animal-painting', 'd1e6257ed13facbe8990f21afb1dac9ec.jpg')
output_filename4 = os.path.join('/users/PSS0009/brach21/final_project/data/
↳images/animal-painting', 'd20cacc077f752bb16cc27e3f7652296c.jpg')
output_filename5 = os.path.join('/users/PSS0009/brach21/final_project/data/
↳images/animal-painting', 'd918cc3486fb789c9c15b26ef42f373ec.jpg')
```

CPU times: user 7 s, sys: 2 s, total: 9 s
Wall time: 11 s

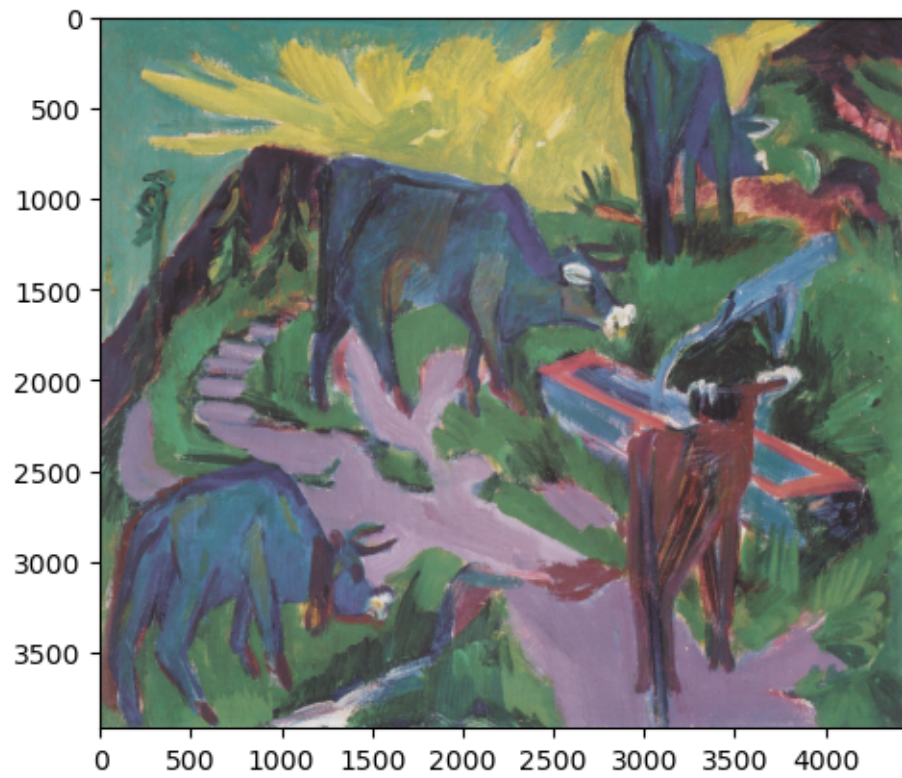
```
[4]: im = Image.open(output_filename)

print(im.format, im.size, im.mode)
```

JPEG (4464, 3918) RGB

```
[5]: plt.imshow(im)
```

```
[5]: <matplotlib.image.AxesImage at 0x14ea4a9f2870>
```



0.6 Directories

```
[6]: # All directories
base_dirs = [
    '/users/PSS0009/brach21/final_project/data/images/abstract',
    '/users/PSS0009/brach21/final_project/data/images/animal-painting',
    '/users/PSS0009/brach21/final_project/data/images/cityscape',
    '/users/PSS0009/brach21/final_project/data/images/figurative',
    '/users/PSS0009/brach21/final_project/data/images/flower-painting',
    '/users/PSS0009/brach21/final_project/data/images/genre-painting',
    '/users/PSS0009/brach21/final_project/data/images/landscape',
    '/users/PSS0009/brach21/final_project/data/images/marina',
    '/users/PSS0009/brach21/final_project/data/images/mythological-painting',
    '/users/PSS0009/brach21/final_project/data/images/nude-painting-nu',
    '/users/PSS0009/brach21/final_project/data/images/religious-painting',
    '/users/PSS0009/brach21/final_project/data/images/still-life',
    '/users/PSS0009/brach21/final_project/data/images/symbolic-painting'
]

# Testing Sample (for size)
base_dir_test = ['/users/PSS0009/brach21/final_project/data/images/marina',
                 '/users/PSS0009/brach21/final_project/data/images/still-life']
```

0.7 # Trial without Parallelization

Not using a function

Using testing directory

```
[7]: %%time
images = {}
img_times_list = []

for base_dir in base_dir_test:

    print(f"\nCurrent directory: {base_dir}\n")

    # Look at files in directory with .jpg and not starting with ._
    filenames = [f for f in os.listdir(base_dir) if f.endswith('.jpg') and not
    ↪f.startswith('._')]

    # Joining path of base_dir and filename of image
    output_filenames = [os.path.join(base_dir, filename) for filename in
    ↪filenames]

    # For every image in the list of images
    for output_filename in output_filenames:
        # IMG START TIME
        img_start_time = time.time()
```

```

# Key is splitting path based on filename
key = os.path.splitext(os.path.basename(output_filename))[0]
try:
    # Open image using Pillow
    with Image.open(output_filename) as img:
        # Filter image based on width and height
        if img.width > 9500 or img.height > 9500:
            print(f"Skipping file {output_filename} due to size: {img.
↳width}x{img.height}")
            continue
        img.verify() # verify that file isn't corrupted
        # Read image information using Dask
        images[key] = dask_image.imread.imread(output_filename)
        # Raise Exception and skip image if error
except Exception as e:
    print(f"Skipping file {output_filename} due to error: {e}")
img_end_time = time.time()
# IMG END TIME
img_time = img_end_time - img_start_time
bullshit = (output_filename, img_time)

# Append list with filename and time to process
img_times_list.append(tuple(bullshit))

```

Current directory: /users/PSS0009/brach21/final_project/data/images/marina

Current directory: /users/PSS0009/brach21/final_project/data/images/still-life

/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:

UserWarning: Metadata Warning, tag 33723 had too many entries: 6, expected 1

warnings.warn(

/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:

UserWarning: Metadata Warning, tag 33723 had too many entries: 14, expected 1

warnings.warn(

Skipping file /users/PSS0009/brach21/final_project/data/images/still-life/7df98e66353960e802001c2ba698fc0fc.jpg due to error: All handlers returned exceptions:

<class 'pims.image_reader.ImageReader'> errored: image file is truncated (23 bytes not processed)

<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (23 bytes not processed)

<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (23 bytes not processed)

<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no attribute 'isJVMStarted'

```

Skipping file /users/PSS0009/brach21/final_project/data/images/still-
life/894cefe96a11b688816ffee4360275c5c.jpg due to error: All handlers returned
exceptions:
<class 'pims.image_reader.ImageReader'> errored: image file is truncated (82
bytes not processed)
<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (82
bytes not processed)
<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (82
bytes not processed)
<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no
attribute 'isJVMStarted'
Skipping file /users/PSS0009/brach21/final_project/data/images/still-
life/cc7a82b31b75d7f6084f6e3cd4be1771c.jpg due to error: All handlers returned
exceptions:
<class 'pims.image_reader.ImageReader'> errored: image file is truncated (61
bytes not processed)
<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (61
bytes not processed)
<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (61
bytes not processed)
<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no
attribute 'isJVMStarted'
Skipping file /users/PSS0009/brach21/final_project/data/images/still-
life/fcfe03cd47d2e2f2909723579956b8b2c.jpg due to error: All handlers returned
exceptions:
<class 'pims.image_reader.ImageReader'> errored: image file is truncated (48
bytes not processed)
<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (48
bytes not processed)
<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (48
bytes not processed)
<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no
attribute 'isJVMStarted'
CPU times: user 1min 6s, sys: 5.19 s, total: 1min 11s
Wall time: 1min 36s

```

Error Management:

Producing Error messages with image name for future analysis

```

[8]: # Create Dataframe
df_np_test = pd.DataFrame(img_times_list, columns=['IMG', 'TIME'])
print(df_np_test.head())

```

	IMG	TIME
0	/users/PSS0009/brach21/final_project/data/imag...	0.016376
1	/users/PSS0009/brach21/final_project/data/imag...	0.034035
2	/users/PSS0009/brach21/final_project/data/imag...	0.012344
3	/users/PSS0009/brach21/final_project/data/imag...	0.013765

4 /users/PSS0009/brach21/final_project/data/imag... 0.007730

```
[9]: print("Longest Image Time:", df_np_test['TIME'].max())
```

Longest Image Time: 0.6511249542236328

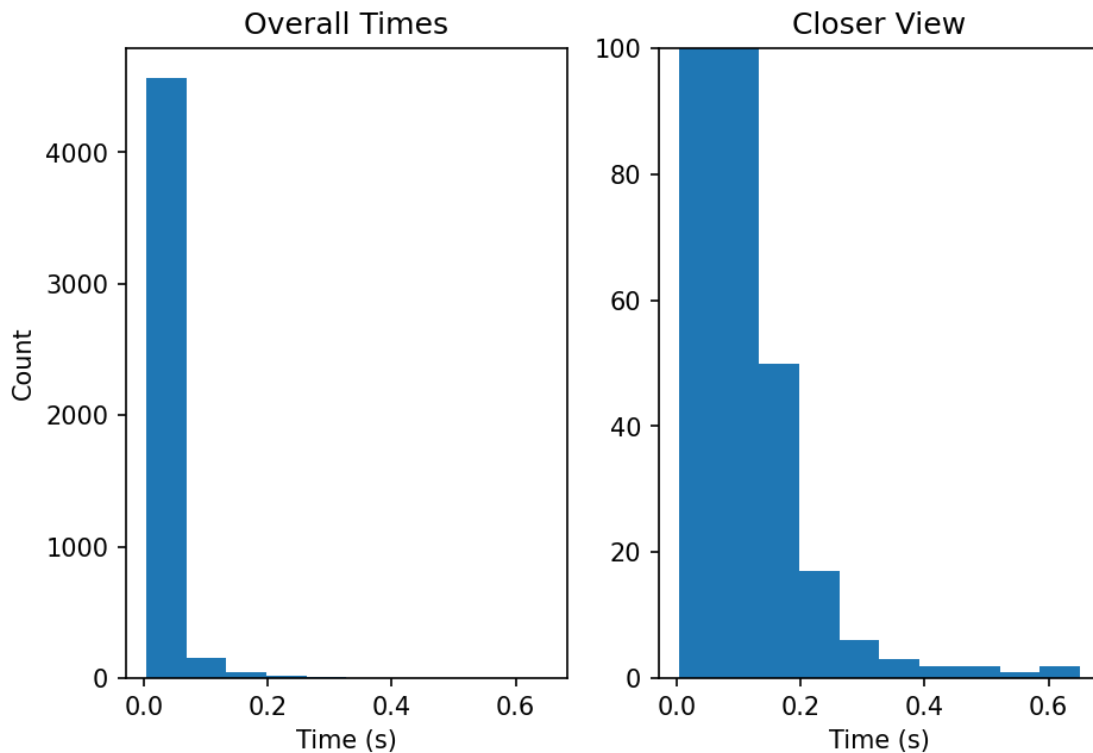
```
[10]: # Generating Plot
fig, axs = plt.subplots(1, 2, dpi=150)
# Overall Times
axs[0].hist(df_np_test['TIME'])
axs[0].set_title("Overall Times")
axs[0].set_ylabel("Count")
axs[0].set_xlabel("Time (s)")

# Enhanced View of Times
axs[1].hist(df_np_test['TIME'])
axs[1].set_title("Closer View")
axs[1].set_xlabel("Time (s)")

axs[1].set_ylim([0, 100])

# General
fig.suptitle("Non-Parallelized Times of Sample", weight='bold', size='x-large')
fig.tight_layout()
```

Non-Parallelized Times of Sample



0.8 # Functions

Create Functions for non-parallized and parallized

0.9 Non-Parallelized

```
[11]: # Non-Parallized
      ## Take list of directories
      def np_image_process(base_dirs):

          # Instantiate empty list and dictionary used in function
          images = {}
          img_times_list_func = []

          for base_dir in base_dirs:

              print(f"\nCurrent directory: {base_dir}\n")

              filenames = [f for f in os.listdir(base_dir) if f.endswith('.jpg') and
↪not f.startswith('.')]
```



```

        output_filenames = [os.path.join(base_dir, filename) for filename in
↪filenames]

    for output_filename in output_filenames:
        # IMG START TIME
        img_start_time = time.time()
        key = os.path.splitext(os.path.basename(output_filename))[0]
        try:
            with Image.open(output_filename) as img:
                if img.width > 9500 or img.height > 9500:
                    print(f"Skipping file {output_filename} due to size:
↪{img.width}x{img.height}")
                    continue
            img.verify()
            images[key] = dask_image.imread.imread(output_filename)
        except Exception as e:
            print(f"Skipping file {output_filename} due to error: {e}")
        img_end_time = time.time()
        # IMG END TIME
        img_time = img_end_time - img_start_time
        bullshit = (output_filename, img_time)

        img_times_list_func.append(tuple(bullshit))

    # Make dataframe from image and times (img_times_list_func)
    data = pd.DataFrame(img_times_list_func, columns=['IMG', 'TIME'])

    return data # return dataframe

```

0.10 Parallelized

Utilize Dask for parallel computing

Run File processing first and image processing second

Both in parallel

```

[12]: # Parallel File processing
def parallel_file_process(base_dirs):
    # Takes Files from given directories

    for base_dir in base_dirs:

        print(f"\nCurrent directory: {base_dir}\n")

        filenames = [f for f in os.listdir(base_dir) if f.endswith('.jpg') and
↪not f.startswith('.') and not f.startswith('_')]

```

```

        output_filenames = [os.path.join(base_dir, filename) for filename in
↪filenames]

    return output_filenames    # output list of filenames within base_dirs

# Parallel Image Processing
def parallel_image_process(output_filenames):
    # Takes list of filenames for image processing

    images = {}
    img_times_list_func = []

    for output_filename in output_filenames:
        # IMG START TIME
        img_start_time = time.time()
        key = os.path.splitext(os.path.basename(output_filename))[0]
        try:
            with Image.open(output_filename) as img:
                if img.width > 9500 or img.height > 9500:
                    print(f"Skipping file {output_filename} due to size: {img.
↪width}x{img.height}")
                    continue
            img.verify()
            images[key] = dask_image.imread.imread(output_filename)
        except Exception as e:
            print(f"Skipping file {output_filename} due to error: {e}")
        img_end_time = time.time()
        # IMG END TIME
        img_time = img_end_time - img_start_time
        bullshit = (output_filename, img_time)

        img_times_list_func.append(tuple(bullshit))
    # Creates dataframe of images and processing time per image
    data = pd.DataFrame(img_times_list_func, columns=['IMG', 'TIME'])

    return data

```

0.10.1 Dask Parallelization

```

[13]: bs_output_filenames = []

# Verify that directory is present in list
for base_dir in base_dirs:
    if not os.path.isdir(base_dir):
        print(f"Skipping invalid directory: {base_dir}")
        continue
    # dask delay file processing from base_dirs

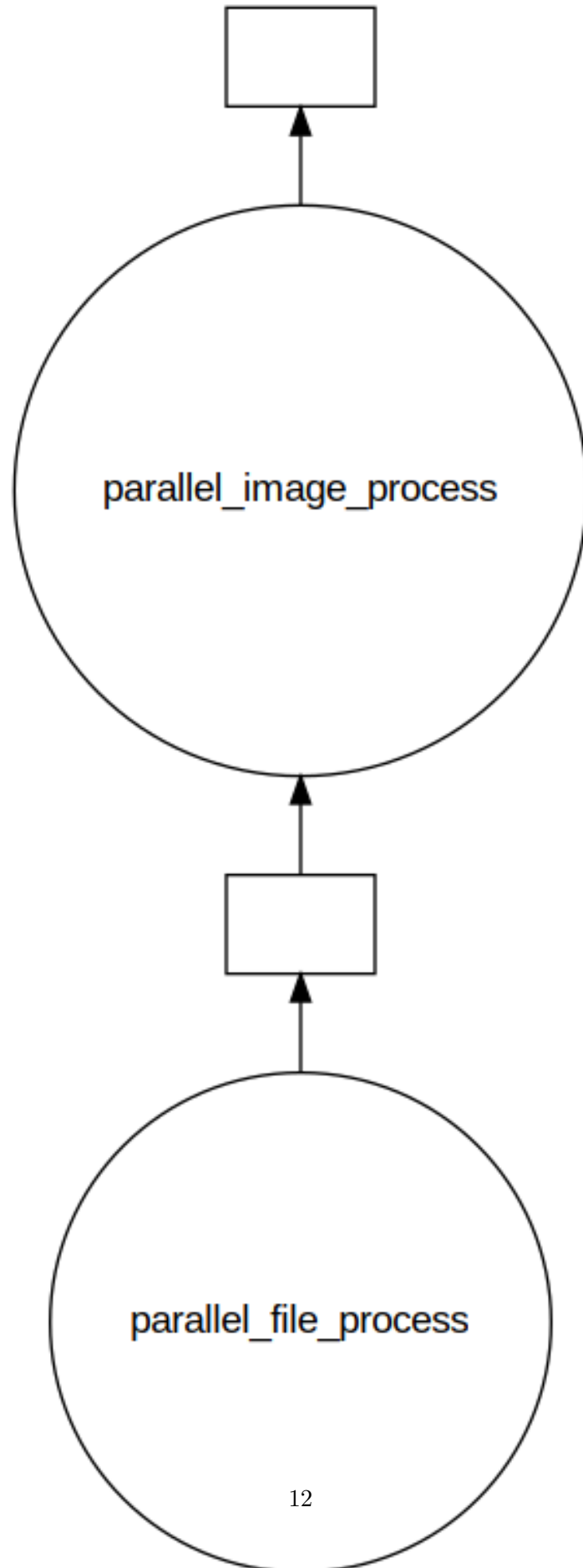
```

```
output_files = dask.delayed(parallel_file_process)([base_dir])
bs_output_filenames.append(output_files) # append filenames to empty list

# process images
image_processing_tasks = []
for output_filenames in bs_output_filenames:
    # dask delay task for image processing using list of filenames per base_dir
    dataframe_img_time = dask.delayed(parallel_image_process)(output_filenames)
    image_processing_tasks.append(dataframe_img_time)
```

```
[14]: dataframe_img_time.visualize()
```

```
[14]:
```



0.11 # Running Non-Parallel and Parallel

0.12 Non-parallel

```
[15]: %%time
      # Non-parallel
      np_time_df = np_image_process(base_dirs)
```

Current directory: /users/PSS0009/brach21/final_project/data/images/abstract

/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:

UserWarning: Metadata Warning, tag 33723 had too many entries: 26, expected 1
warnings.warn(

Skipping file /users/PSS0009/brach21/final_project/data/images/abstract/23317e3a7151dc264f62b62c221315e5.jpg due to size: 3609x9541

/apps/python/3.12/lib/python3.12/site-packages/PIL/Image.py:3218:

DecompressionBombWarning: Image size (99962094 pixels) exceeds limit of 89478485 pixels, could be decompression bomb DOS attack.
warnings.warn(

Skipping file /users/PSS0009/brach21/final_project/data/images/abstract/621belee87852917e9870107f2d65ba4.jpg due to size: 14559x6866

/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:

UserWarning: Metadata Warning, tag 33723 had too many entries: 2, expected 1
warnings.warn(

/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:

UserWarning: Metadata Warning, tag 33723 had too many entries: 17, expected 1
warnings.warn(

/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:

UserWarning: Metadata Warning, tag 33723 had too many entries: 12, expected 1
warnings.warn(

Skipping file /users/PSS0009/brach21/final_project/data/images/abstract/c90b7a03b0c904e0af01dc9c9e347b59.jpg due to error: All handlers returned exceptions:

<class 'pims.image_reader.ImageReader'> errored: image file is truncated (4054 bytes not processed)

<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (4054 bytes not processed)

<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (4054 bytes not processed)

<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no attribute 'isJVMStarted'

```
/apps/python/3.12/lib/python3.12/site-packages/PIL/Image.py:3218:
DecompressionBombWarning: Image size (107327830 pixels) exceeds limit of
89478485 pixels, could be decompression bomb DOS attack.
  warnings.warn(

Skipping file /users/PSS0009/brach21/final_project/data/images/abstract/dca9d9c3
d515db31c88eb255e18996c4.jpg due to size: 15530x6911

/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:
UserWarning: Metadata Warning, tag 33723 had too many entries: 4, expected 1
  warnings.warn(
```

Current directory: /users/PSS0009/brach21/final_project/data/images/animal-
painting

```
Skipping file /users/PSS0009/brach21/final_project/data/images/animal-
painting/49ef80973448ea527ae9b762873b6784c.jpg due to error: All handlers
returned exceptions:
<class 'pims.image_reader.ImageReader'> errored: image file is truncated (57
bytes not processed)
<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (57
bytes not processed)
<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (57
bytes not processed)
<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no
attribute 'isJVMStarted'
```

Current directory: /users/PSS0009/brach21/final_project/data/images/cityscape

```
Skipping file /users/PSS0009/brach21/final_project/data/images/cityscape/26e4f3f
ba28d0e44337a3f3f904b1685c.jpg due to error: All handlers returned exceptions:
<class 'pims.image_reader.ImageReader'> errored: image file is truncated (53
bytes not processed)
<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (53
bytes not processed)
<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (53
bytes not processed)
<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no
attribute 'isJVMStarted'
Skipping file /users/PSS0009/brach21/final_project/data/images/cityscape/564e8ca
ed5f754cc37744e64aa5db15cc.jpg due to error: All handlers returned exceptions:
<class 'pims.image_reader.ImageReader'> errored: image file is truncated (79
bytes not processed)
<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (79
bytes not processed)
<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (79
bytes not processed)
<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no
```

attribute 'isJVMStarted'

Current directory: /users/PSS0009/brach21/final_project/data/images/figurative

Current directory: /users/PSS0009/brach21/final_project/data/images/flower-painting

Current directory: /users/PSS0009/brach21/final_project/data/images/genre-painting

/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:
UserWarning: Metadata Warning, tag 33723 had too many entries: 22, expected 1
warnings.warn(

Skipping file /users/PSS0009/brach21/final_project/data/images/genre-painting/ece40a957a51a7cfb9e24bbccaa0673fc.jpg due to error: All handlers returned exceptions:
<class 'pims.image_reader.ImageReader'> errored: image file is truncated (0 bytes not processed)
<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (0 bytes not processed)
<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (0 bytes not processed)
<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no attribute 'isJVMStarted'

Current directory: /users/PSS0009/brach21/final_project/data/images/landscape

/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:890:
UserWarning: Corrupt EXIF data. Expecting to read 2 bytes but only got 0.
warnings.warn(str(msg))

Current directory: /users/PSS0009/brach21/final_project/data/images/marina

Current directory:
/users/PSS0009/brach21/final_project/data/images/mythological-painting

Skipping file /users/PSS0009/brach21/final_project/data/images/mythological-painting/219821bca439d3774690c4d0b7b22b1bc.jpg due to size: 25528x3000
Skipping file /users/PSS0009/brach21/final_project/data/images/mythological-painting/fbcd0b972339c781482a466fd1448ebc.jpg due to error: All handlers returned exceptions:
<class 'pims.image_reader.ImageReader'> errored: image file is truncated (18 bytes not processed)

```
<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (18
bytes not processed)
<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (18
bytes not processed)
<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no
attribute 'isJVMStarted'
```

Current directory: /users/PSS0009/brach21/final_project/data/images/nude-
painting-nu

```
/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:
UserWarning: Metadata Warning, tag 33723 had too many entries: 31, expected 1
warnings.warn(
```

Current directory: /users/PSS0009/brach21/final_project/data/images/religious-
painting

Current directory: /users/PSS0009/brach21/final_project/data/images/still-life

```
/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:
UserWarning: Metadata Warning, tag 33723 had too many entries: 6, expected 1
warnings.warn(
/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:
UserWarning: Metadata Warning, tag 33723 had too many entries: 14, expected 1
warnings.warn(
```

Skipping file /users/PSS0009/brach21/final_project/data/images/still-
life/7df98e66353960e802001c2ba698fc0fc.jpg due to error: All handlers returned
exceptions:

```
<class 'pims.image_reader.ImageReader'> errored: image file is truncated (23
bytes not processed)
<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (23
bytes not processed)
<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (23
bytes not processed)
<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no
attribute 'isJVMStarted'
```

Skipping file /users/PSS0009/brach21/final_project/data/images/still-
life/894cefe96a11b688816ffee4360275c5c.jpg due to error: All handlers returned
exceptions:

```
<class 'pims.image_reader.ImageReader'> errored: image file is truncated (82
bytes not processed)
<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (82
bytes not processed)
<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (82
bytes not processed)
```



```

<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no
attribute 'isJVMStarted'
Skipping file /users/PSS0009/brach21/final_project/data/images/still-
life/cc7a82b31b75d7f6084f6e3cd4be1771c.jpg due to error: All handlers returned
exceptions:
<class 'pims.image_reader.ImageReader'> errored: image file is truncated (61
bytes not processed)
<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (61
bytes not processed)
<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (61
bytes not processed)
<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no
attribute 'isJVMStarted'
Skipping file /users/PSS0009/brach21/final_project/data/images/still-
life/fcfe03cd47d2e2f2909723579956b8b2c.jpg due to error: All handlers returned
exceptions:
<class 'pims.image_reader.ImageReader'> errored: image file is truncated (48
bytes not processed)
<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (48
bytes not processed)
<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (48
bytes not processed)
<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no
attribute 'isJVMStarted'

```

Current directory: /users/PSS0009/brach21/final_project/data/images/symbolic-
painting

```

Skipping file /users/PSS0009/brach21/final_project/data/images/symbolic-
painting/e20070688563b891915c82c5d9d0b103c.jpg due to error: All handlers
returned exceptions:
<class 'pims.image_reader.ImageReader'> errored: image file is truncated (86
bytes not processed)
<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (86
bytes not processed)
<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (86
bytes not processed)
<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no
attribute 'isJVMStarted'
CPU times: user 21min 14s, sys: 1min 58s, total: 23min 12s
Wall time: 32min 37s

```

```
[16]: np_time_df.iloc[81000,1]
```

```
[16]: 0.028087139129638672
```

```
[17]: np_time_df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 81009 entries, 0 to 81008
Data columns (total 2 columns):
#   Column  Non-Null Count  Dtype
---  -
0    IMG      81009 non-null    object
1    TIME      81009 non-null    float64
dtypes: float64(1), object(1)
memory usage: 1.2+ MB

```

```

[18]: # Non-parallelized Times Plotting
fig, axs = plt.subplots(1, 2, dpi=150)
# Overall Times
axs[0].hist(np_time_df['TIME'])
axs[0].set_title("Overall Times")
axs[0].set_ylabel("Count")
axs[0].set_xlabel("Time (s)")

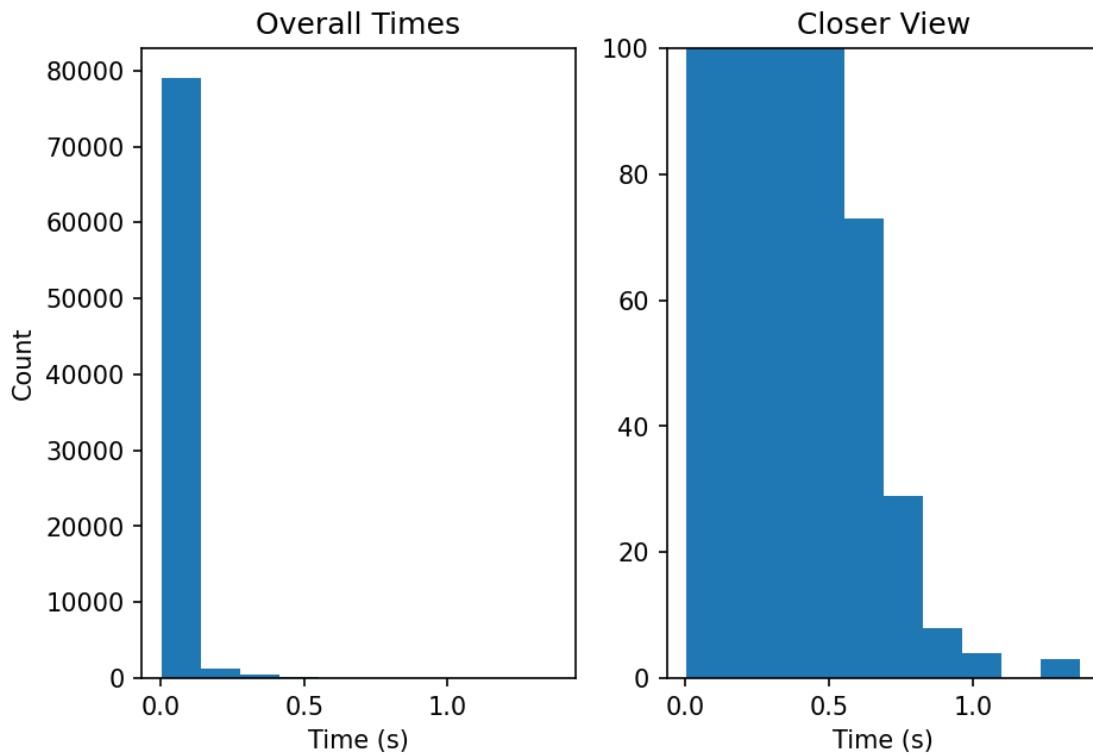
# Enhanced View of Times
axs[1].hist(np_time_df['TIME'])
axs[1].set_title("Closer View")
axs[1].set_xlabel("Time (s)")

axs[1].set_ylim([0, 100])

# General
fig.suptitle("Non-Parallelized Times", weight='bold', size='x-large')
fig.tight_layout()

```

Non-Parallelized Times



0.13 Parallel (using Dask)

```
[19]: %%time
      # Parallel
      # Execute all tasks
      results = dask.compute(*image_processing_tasks)

      # create dataframe with filename and time
      p_time_df = pd.concat(results, ignore_index=True)
```

Current directory: /users/PSS0009/brach21/final_project/data/images/religious-painting

Current directory: /users/PSS0009/brach21/final_project/data/images/symbolic-painting

Current directory: /users/PSS0009/brach21/final_project/data/images/cityscape

Current directory: /users/PSS0009/brach21/final_project/data/images/genre-painting

Current directory: /users/PSS0009/brach21/final_project/data/images/abstract

Current directory: /users/PSS0009/brach21/final_project/data/images/animal-painting

Current directory: /users/PSS0009/brach21/final_project/data/images/figurative

Current directory: /users/PSS0009/brach21/final_project/data/images/landscape

Current directory: /users/PSS0009/brach21/final_project/data/images/still-life

Current directory: /users/PSS0009/brach21/final_project/data/images/marina

Current directory:
/users/PSS0009/brach21/final_project/data/images/mythological-painting

Current directory: /users/PSS0009/brach21/final_project/data/images/flower-painting

Skipping file /users/PSS0009/brach21/final_project/data/images/mythological-painting/219821bca439d3774690c4d0b7b22b1bc.jpg due to size: 25528x3000

Skipping file /users/PSS0009/brach21/final_project/data/images/animal-painting/49ef80973448ea527ae9b762873b6784c.jpg due to error: All handlers

returned exceptions:

<class 'pims.image_reader.ImageReader'> errored: image file is truncated (57 bytes not processed)

<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (57 bytes not processed)

<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (57 bytes not processed)

<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no attribute 'isJVMStarted'

/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:

UserWarning: Metadata Warning, tag 33723 had too many entries: 2, expected 1
warnings.warn(

```

/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:
UserWarning: Metadata Warning, tag 33723 had too many entries: 26, expected 1
    warnings.warn(
/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:
UserWarning: Metadata Warning, tag 33723 had too many entries: 22, expected 1
    warnings.warn(
/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:
UserWarning: Metadata Warning, tag 33723 had too many entries: 6, expected 1
    warnings.warn(
/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:
UserWarning: Metadata Warning, tag 33723 had too many entries: 14, expected 1
    warnings.warn(

Skipping file /users/PSS0009/brach21/final_project/data/images/cityscape/26e4f3f
ba28d0e44337a3f3f904b1685c.jpg due to error: All handlers returned exceptions:
<class 'pims.image_reader.ImageReader'> errored: image file is truncated (53
bytes not processed)
<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (53
bytes not processed)
<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (53
bytes not processed)
<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no
attribute 'isJVMStarted'
Skipping file /users/PSS0009/brach21/final_project/data/images/still-
life/7df98e66353960e802001c2ba698fc0fc.jpg due to error: All handlers returned
exceptions:
<class 'pims.image_reader.ImageReader'> errored: image file is truncated (23
bytes not processed)
<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (23
bytes not processed)
<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (23
bytes not processed)
<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no
attribute 'isJVMStarted'
Skipping file /users/PSS0009/brach21/final_project/data/images/still-
life/894cefe96a11b688816ffee4360275c5c.jpg due to error: All handlers returned
exceptions:
<class 'pims.image_reader.ImageReader'> errored: image file is truncated (82
bytes not processed)
<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (82
bytes not processed)
<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (82
bytes not processed)
<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no
attribute 'isJVMStarted'
Skipping file /users/PSS0009/brach21/final_project/data/images/abstract/23317e3a
7151dc264f62b62c221315e5.jpg due to size: 3609x9541

```

Current directory: /users/PSS0009/brach21/final_project/data/images/nude-painting-nu

Skipping file /users/PSS0009/brach21/final_project/data/images/still-life/cc7a82b31b75d7f6084f6e3cd4be1771c.jpg due to error: All handlers returned exceptions:

<class 'pims.image_reader.ImageReader'> errored: image file is truncated (61 bytes not processed)

<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (61 bytes not processed)

<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (61 bytes not processed)

<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no attribute 'isJVMStarted'

Skipping file /users/PSS0009/brach21/final_project/data/images/symbolic-painting/e20070688563b891915c82c5d9d0b103c.jpg due to error: All handlers returned exceptions:

<class 'pims.image_reader.ImageReader'> errored: image file is truncated (86 bytes not processed)

<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (86 bytes not processed)

<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (86 bytes not processed)

<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no attribute 'isJVMStarted'

Skipping file /users/PSS0009/brach21/final_project/data/images/mythological-painting/fbcd0b972339c781482a466fd1448ebc.jpg due to error: All handlers returned exceptions:

<class 'pims.image_reader.ImageReader'> errored: image file is truncated (18 bytes not processed)

<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (18 bytes not processed)

<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (18 bytes not processed)

<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no attribute 'isJVMStarted'

Skipping file /users/PSS0009/brach21/final_project/data/images/cityscape/564e8caed5f754cc37744e64aa5db15cc.jpg due to error: All handlers returned exceptions:

<class 'pims.image_reader.ImageReader'> errored: image file is truncated (79 bytes not processed)

<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (79 bytes not processed)

<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (79 bytes not processed)

<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no attribute 'isJVMStarted'

Skipping file /users/PSS0009/brach21/final_project/data/images/still-life/fcfe03cd47d2e2f2909723579956b8b2c.jpg due to error: All handlers returned

```

exceptions:
<class 'pims.image_reader.ImageReader'> errored: image file is truncated (48
bytes not processed)
<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (48
bytes not processed)
<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (48
bytes not processed)
<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no
attribute 'isJVMStarted'

/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:
UserWarning: Metadata Warning, tag 33723 had too many entries: 31, expected 1
    warnings.warn(
/apps/python/3.12/lib/python3.12/site-packages/PIL/Image.py:3218:
DecompressionBombWarning: Image size (99962094 pixels) exceeds limit of 89478485
pixels, could be decompression bomb DOS attack.
    warnings.warn(

Skipping file /users/PSS0009/brach21/final_project/data/images/abstract/621belee
87852917e9870107f2d65ba4.jpg due to size: 14559x6866

/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:
UserWarning: Metadata Warning, tag 33723 had too many entries: 17, expected 1
    warnings.warn(
/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:
UserWarning: Metadata Warning, tag 33723 had too many entries: 12, expected 1
    warnings.warn(
/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:890:
UserWarning: Corrupt EXIF data. Expecting to read 2 bytes but only got 0.
    warnings.warn(str(msg))

Skipping file /users/PSS0009/brach21/final_project/data/images/abstract/c90b7a03
b0c904e0af01dc9c9e347b59.jpg due to error: All handlers returned exceptions:
<class 'pims.image_reader.ImageReader'> errored: image file is truncated (4054
bytes not processed)
<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (4054
bytes not processed)
<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated
(4054 bytes not processed)
<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no
attribute 'isJVMStarted'

/apps/python/3.12/lib/python3.12/site-packages/PIL/Image.py:3218:
DecompressionBombWarning: Image size (107327830 pixels) exceeds limit of
89478485 pixels, could be decompression bomb DOS attack.
    warnings.warn(

Skipping file /users/PSS0009/brach21/final_project/data/images/abstract/dca9d9c3
d515db31c88eb255e18996c4.jpg due to size: 15530x6911

/apps/python/3.12/lib/python3.12/site-packages/PIL/TiffImagePlugin.py:704:

```

```
UserWarning: Metadata Warning, tag 33723 had too many entries: 4, expected 1
warnings.warn(

Skipping file /users/PSS0009/brach21/final_project/data/images/genre-
painting/ece40a957a51a7cfb9e24bbccaa0673fc.jpg due to error: All handlers
returned exceptions:
<class 'pims.image_reader.ImageReader'> errored: image file is truncated (0
bytes not processed)
<class 'pims.image_reader.ImageReaderND'> errored: image file is truncated (0
bytes not processed)
<class 'pims.imageio_reader.ImageIOReader'> errored: image file is truncated (0
bytes not processed)
<class 'pims.bioformats.BioformatsReader'> errored: 'NoneType' object has no
attribute 'isJVMStarted'
CPU times: user 22min 27s, sys: 1min 10s, total: 23min 37s
Wall time: 6min 52s
```

```
[20]: p_time_df.head()
```

```
[20]:
```

		IMG	TIME
0	/users/PSS0009/brach21/final_project/data/imag...	0.134480	
1	/users/PSS0009/brach21/final_project/data/imag...	0.031847	
2	/users/PSS0009/brach21/final_project/data/imag...	0.017490	
3	/users/PSS0009/brach21/final_project/data/imag...	0.007940	
4	/users/PSS0009/brach21/final_project/data/imag...	0.014027	

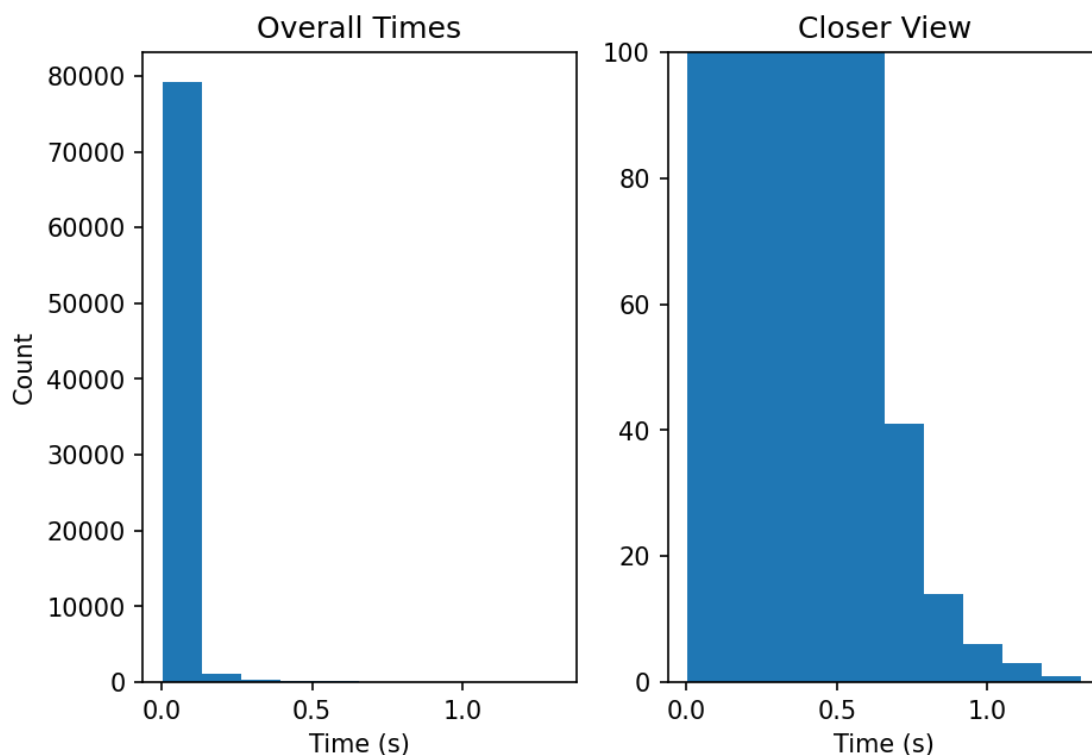
```
[21]: # Plotting Parallelized Times
fig, axs = plt.subplots(1, 2, dpi=150)
# Overall Times
axs[0].hist(p_time_df['TIME'])
axs[0].set_title("Overall Times")
axs[0].set_ylabel("Count")
axs[0].set_xlabel("Time (s)")

# Enhanced View of Times
axs[1].hist(p_time_df['TIME'])
axs[1].set_title("Closer View")
axs[1].set_xlabel("Time (s)")

axs[1].set_ylim([0, 100])

# General
fig.suptitle("Parallelized Times", weight='bold', size='x-large')
fig.tight_layout()
```


Parallelized Times



0.14 # Conclusion

0.14.1 Non-Parallelized Code

CPU times: user 21min 7s, sys: 1min 51s, total: 22min 59s

Wall time: 24min 28s

CPU times: user 21min 14s, sys: 1min 58s, total: 23min 12s

Wall time: 32min 37s

0.14.2 Parallelized Code

CPU times: user 22min 31s, sys: 1min 5s, total: 23min 36s

Wall time: 6min 50s

CPU times: user 22min 27s, sys: 1min 10s, total: 23min 37s

Wall time: 6min 52s

```
[22]: print("Non-Parallelized Average Time:", np_time_df['TIME'].mean())  
      print("Parallelized Average Time:", p_time_df['TIME'].mean())
```

```
print("Is parallelized faster?", np_time_df['TIME'].mean() > p_time_df['TIME'].  
    ↪mean())
```

Non-Parallelized Average Time: 0.024145636610743527

Parallelized Average Time: 0.029193748044120622

Is parallelized faster? False

[]: