Black Patches

July 9, 2025

Jakob Balkovec Wed Jul 9th 2025

1 Intermediate Analysis of Results

This is to determine the ideal threshold for the black patches.

[ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

PICKLE_PATH = r'/Users/jbalkovec/Desktop/DR/data/patches/frame/patch_frame.pkl'
THRESHOLD = 10

# same function as in the pipeline
def count_black_pixels(patch, threshold=10):
    # docs... (removed for brevity)
    if isinstance(patch, np.ndarray):
        if patch.ndim == 3:
            return np.sum(np.all(patch < threshold, axis=-1))
        elif patch.ndim == 2:
            return np.sum(patch < threshold)
    return 0

def plot_patches(df, patch_col="patch", n_per_fig=8, start=0, stop=100,␣
 ↪cmap=None):
    # pre: df exists
    # post: plots patches in df
    # desc: plots patches from a df in groups of n_per_fig.

    for i in range(start, stop, n_per_fig):
        subset = df.iloc[i:i+n_per_fig]
        fig, axes = plt.subplots(1, len(subset), figsize=(15, 3))
        if len(subset) == 1:
            axes = [axes]
        for ax, (_, row) in zip(axes, subset.iterrows()):
            patch = row[patch_col]
            ax.imshow(patch, cmap=cmap)
            ax.set_title(f"black pixels: {row['black_pixel_count']}")
```

```
            ax.axis("off")
        plt.tight_layout()
        plt.show()
```

[ ]:
```
df = pd.read_pickle(PICKLE_PATH)
display(df.head(n=2))
```

```
                                              patch  patch_no    x    y  \
0  [[105, 94, 81, 89, 91, 79, 86, 102, 86, 69, 71…         1  938  156
1  [[82, 90, 101, 104, 107, 109, 117, 120, 109, 9…         2  530  350

  lesion_type                                      patch_polygon  \
0      lesion  POLYGON ((926 144, 950 144, 950 168, 926 168, …
1      lesion  POLYGON ((518 338, 542 338, 542 362, 518 362, …

            label                                      lesion_shape image_id  \
0  microaneurysms  [POLYGON ((938.9980392156863 161, 939 160.9980…   0068_1
1  microaneurysms  [POLYGON ((533.9980392156863 357, 533.99803921…   0068_1

                          file_name
0  0068_1_microaneurysms_938_156.png
1  0068_1_microaneurysms_530_350.png
```

[ ]:
```
df["black_pixel_count"] = df["patch"].apply(count_black_pixels)
df_sorted = df.sort_values(by="black_pixel_count", ascending=False)

df_sorted.head(n=2)
```

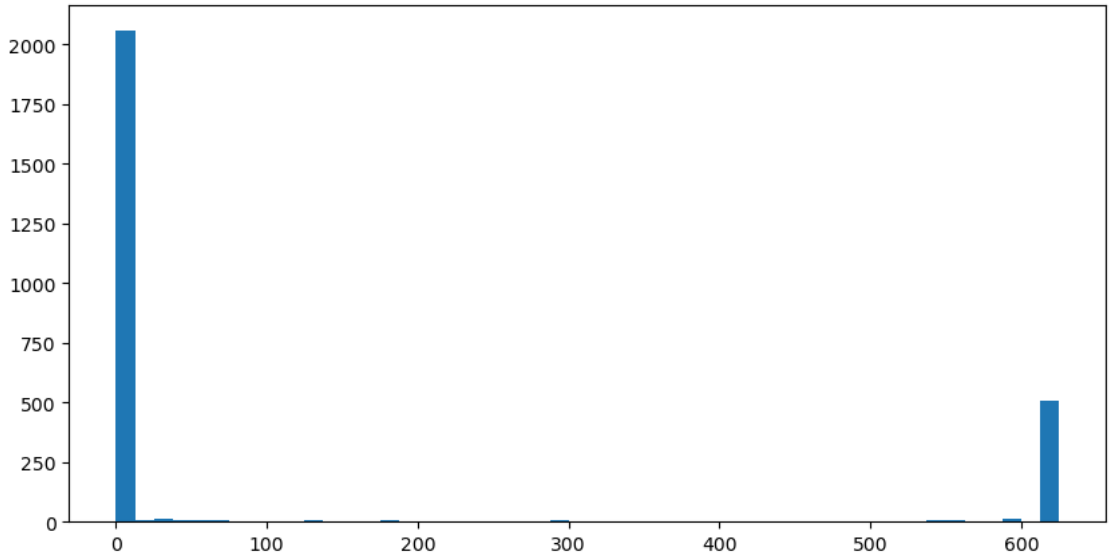
[ ]:
```
                                              patch  patch_no     x     y  \
2735  [[3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,…      2736  1260  1260
289   [[3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,…       290  1236   108

     lesion_type                                      patch_polygon  label  \
2735     healthy  POLYGON ((1248 1248, 1272 1248, 1272 1272, 124…  black
289      healthy  POLYGON ((1224 96, 1248 96, 1248 120, 1224 120…  black

     lesion_shape image_id                  file_name  black_pixel_count
2735         None   0068_1  0068_1_black_1260_1260.png                625
289          None   0068_1   0068_1_black_1236_108.png                625
```

[ ]:
```
df_sorted["black_pixel_count"].hist(bins=50, grid=False, figsize=(10, 5))
```

[ ]:
```
<Axes: >
```

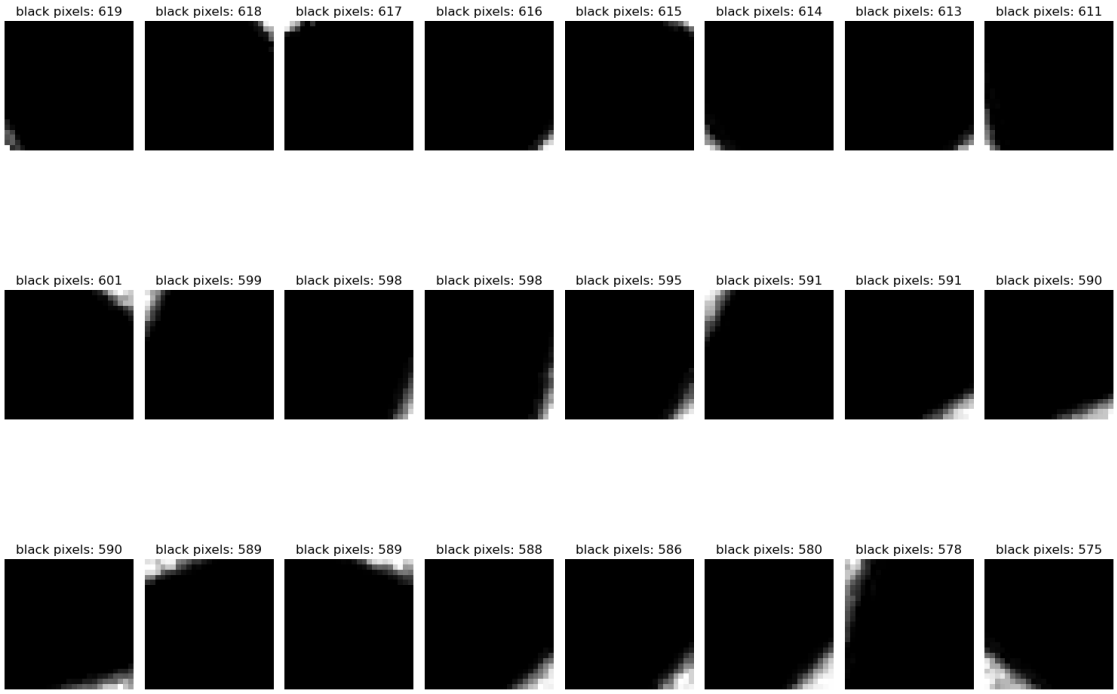


1.0.1 All-Black Patches

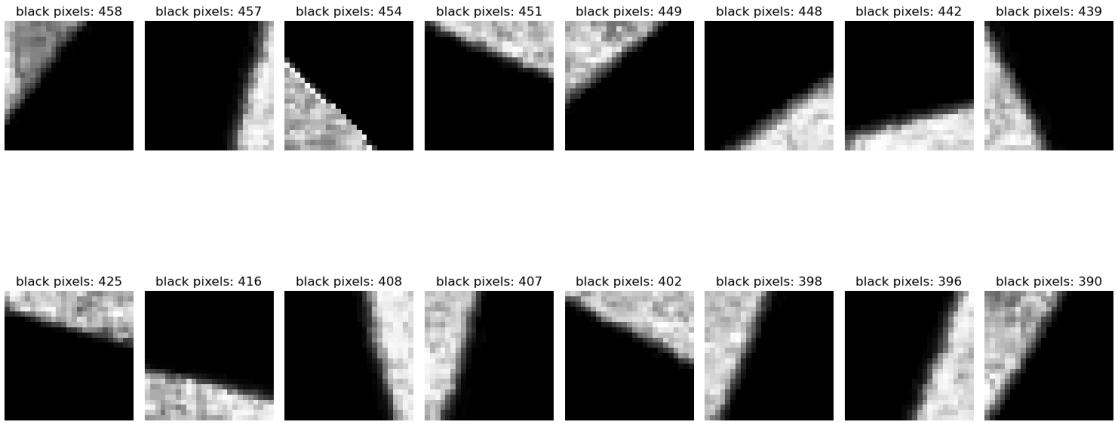
[ ]:
```
plot_patches(df_sorted, patch_col="patch", n_per_fig=8, start=1, stop=24,␣
↪cmap='gray')
```

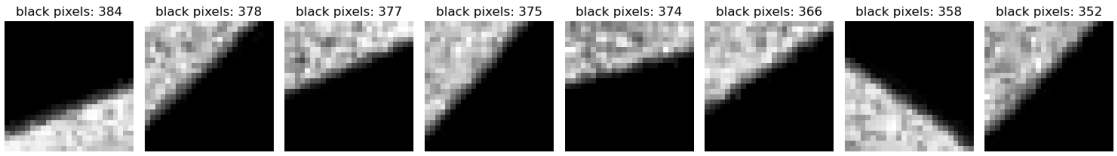
1.0.2 By # of Black Pixels

[ ]:
```
plot_patches(df_sorted, patch_col="patch", n_per_fig=8, start=500, stop=524,␣
↪cmap='gray')
```



[ ]:
```
plot_patches(df_sorted, patch_col="patch", n_per_fig=8, start=550, stop=574,␣
↪cmap='gray')
```





550 is the sweet-spot I my opinion