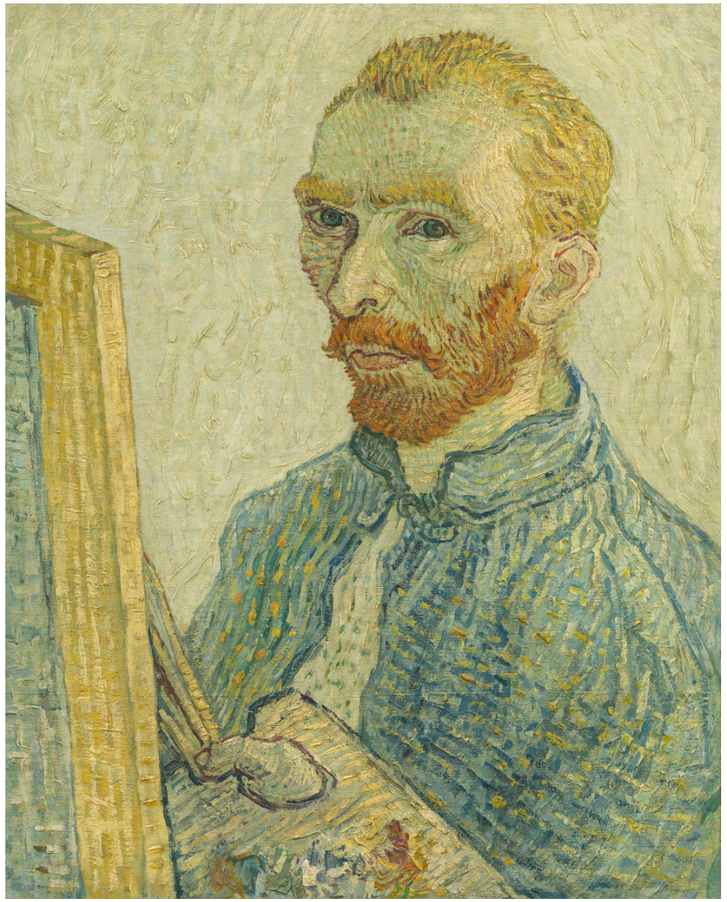
DID VAN GOGH PAINT THIS?



PRESENTED BY

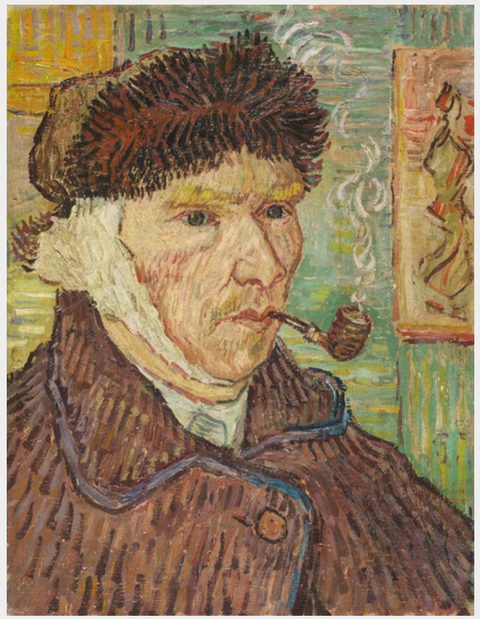
YIYANG XU

BOSTON UNIVERSITY

Tracking fine art ownership is difficult due to private sales and limited public records.

Art buyers must verify authenticity to make sound investments, while art dealers must ensure authenticity to protect their reputation.

↘	Train	Test
↘ Class 0: Forgeries and other	100	24
↘ Class 1: Real	100	24



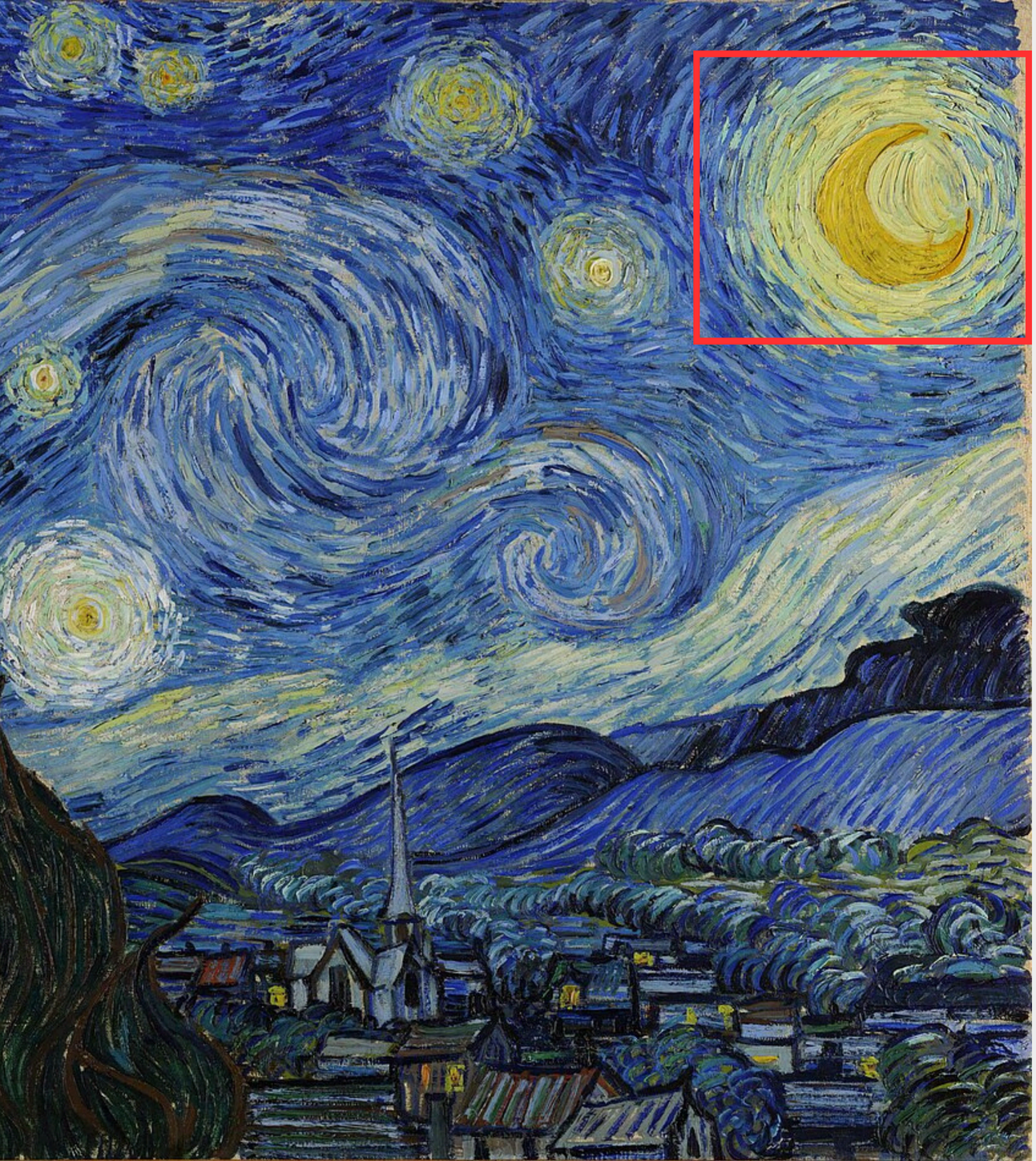
Human forgery

AI forgery

Real

HEURISTIC FEATURE EXTRACTION

What do you notice first?

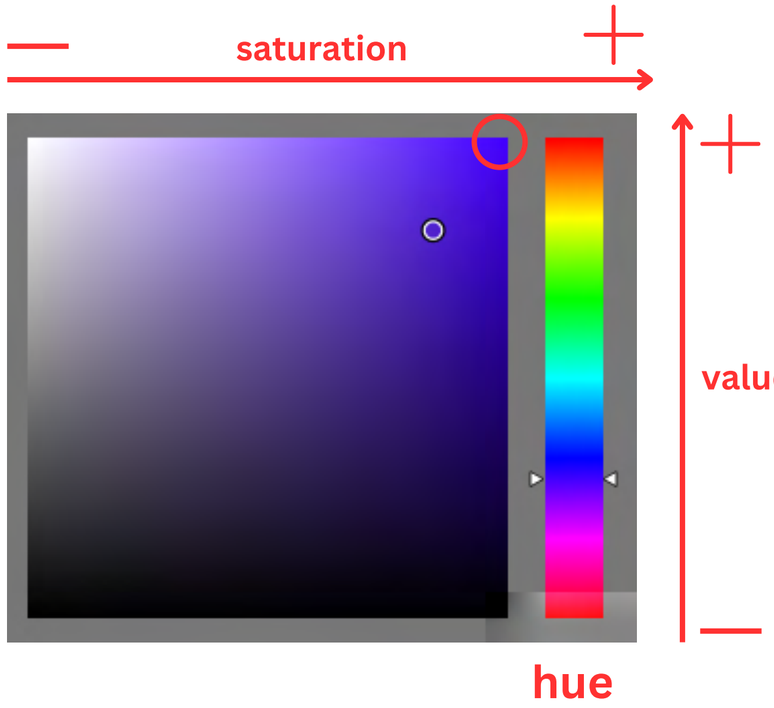


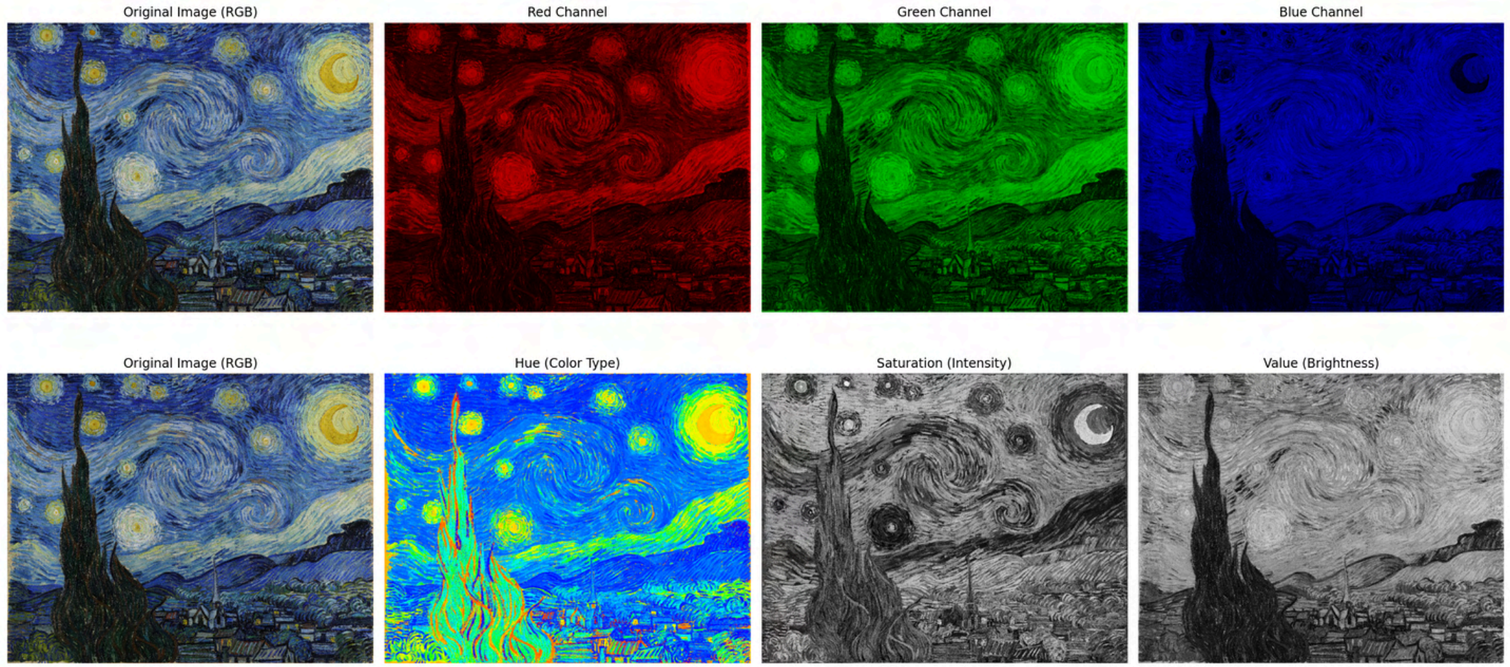
How yellow is this yellow?

What is Van Gogh’s unique palette?

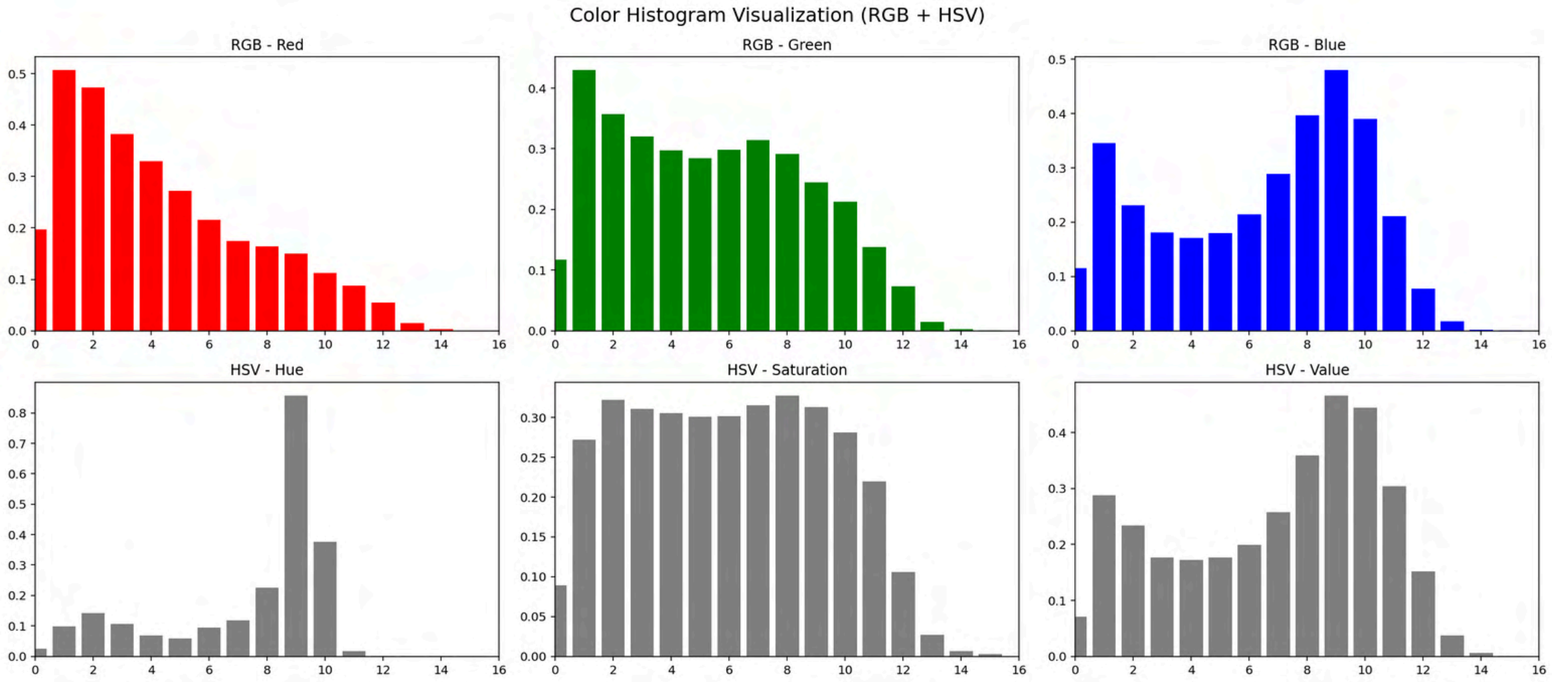
Feature Extraction #1:
Color Histograms (RGB + HSV)

1. **Calculate RGB and HSV for each pixel**





2. For each color channel, divide the range of values (0–255) into 16 bins, and count how many pixels fall into each bin.



For channel C and bin i, the histogram is:

$$H_C(i) = \frac{1}{N} \sum_{x=1}^{W} \sum_{y=1}^{H} \delta(I_C(x,y), i).$$

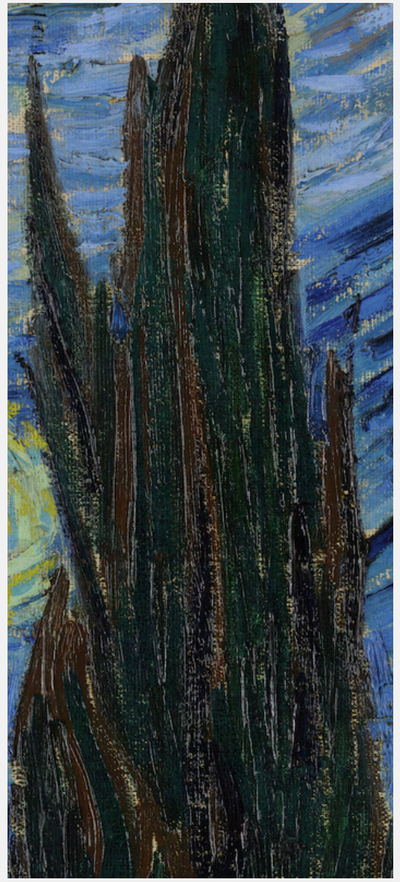
3. Normalize to make the sum of all bins in a histogram = 1.

4. Concatenate the vector so it has 96 entries and do that for every painting.

Artists tend to have consistent, unqiue habits in how they paint subjects. This can be represented by flow of motion, direction and frequency of their brushstrokes.

Van Gogh often paint short spirals for sky and clouds.

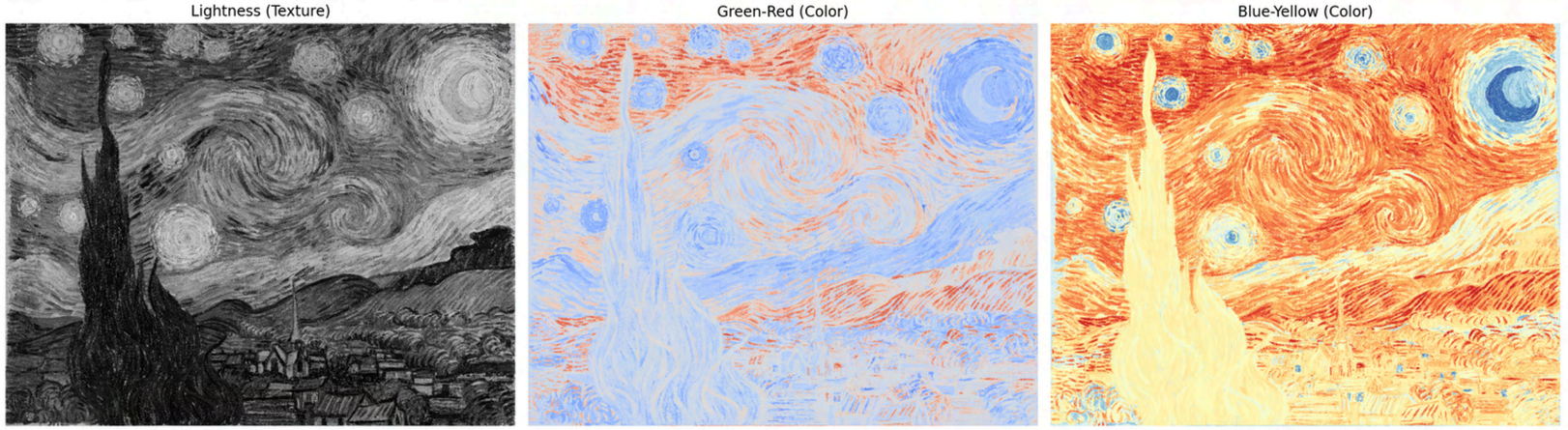
Clumped, thick, diagonal lines for shadows



Clumped, thick, vertical lines for trees

Feature Extraction #2:
HOG (Histogram of gradients) from LAB Color Space

1. **Convert each image into LAB color space**



2. Calculate the horizontal and vertical changes for each pixel in these three images

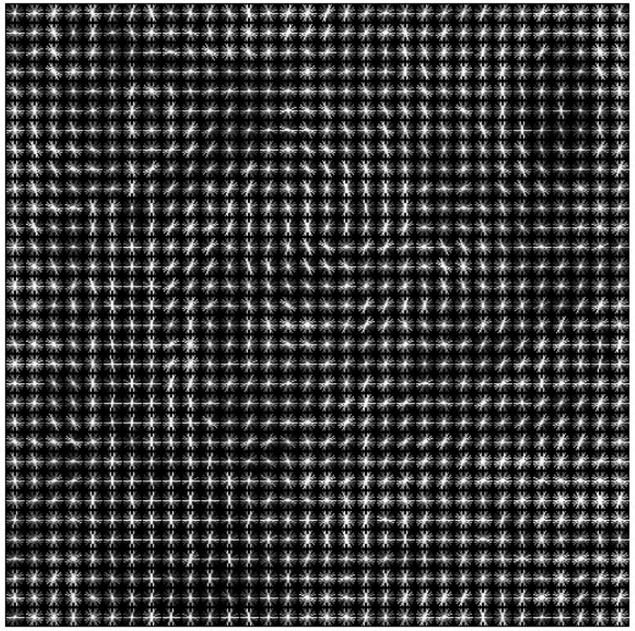
$$G_x(x,y) = I(x+1,y) - I(x-1,y) \qquad G_y(x,y) = I(x,y+1) - I(x,y-1)$$

3. Calculate the gradient(direction and magnitude) for each pixel

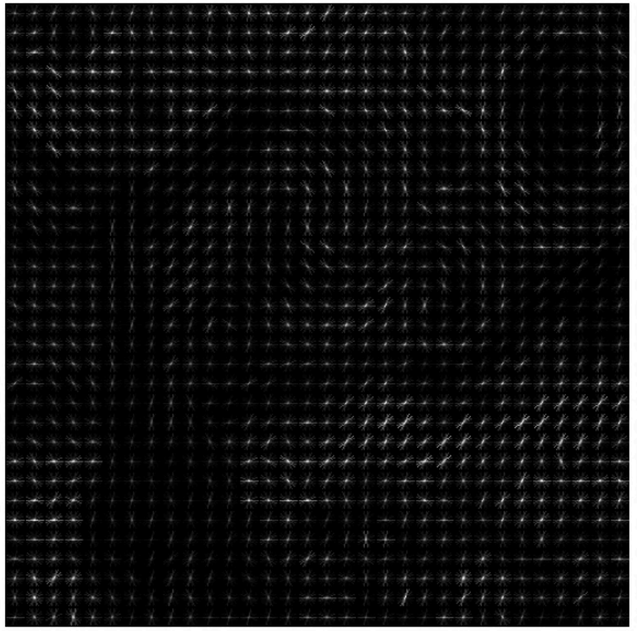
$$m(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \qquad \theta(x,y) = \tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right)$$

How does every pixel changes in brightness and color compared to the pixels around it?

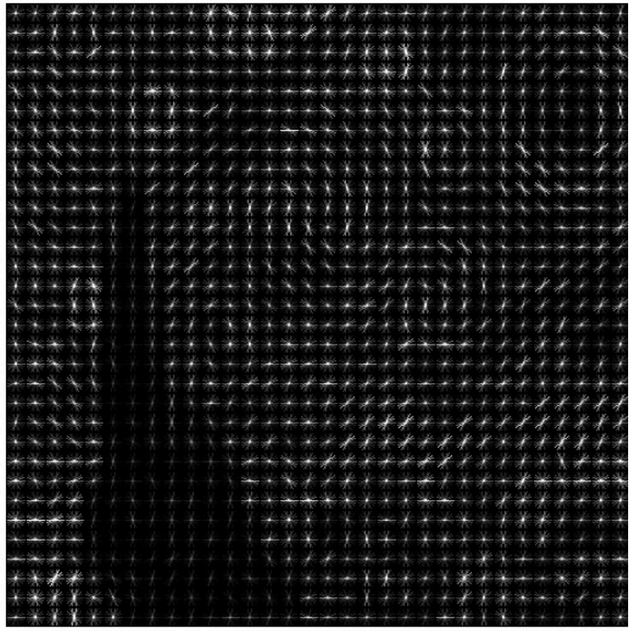
HOG on Lightness channel



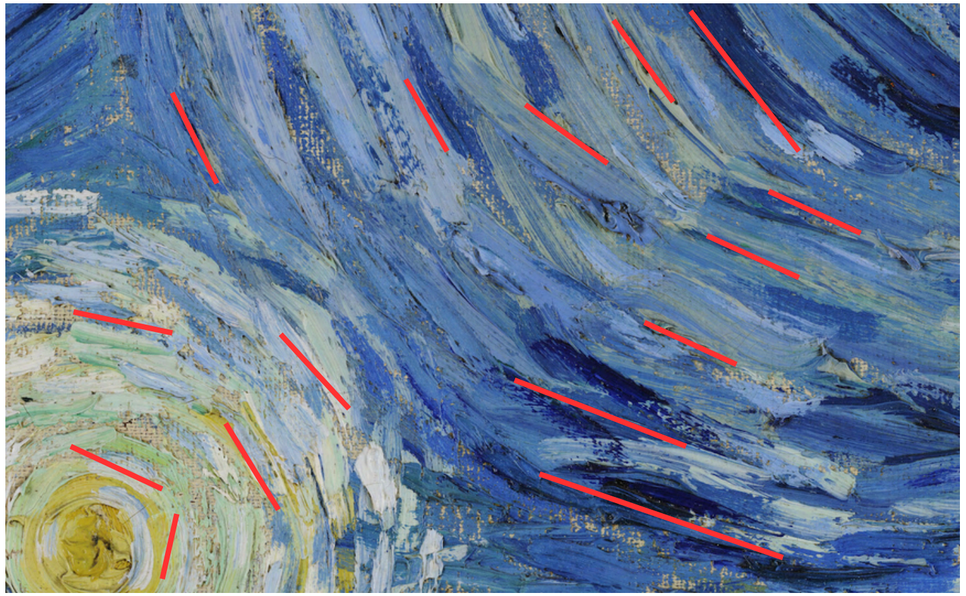
HOG on Red/Green channel

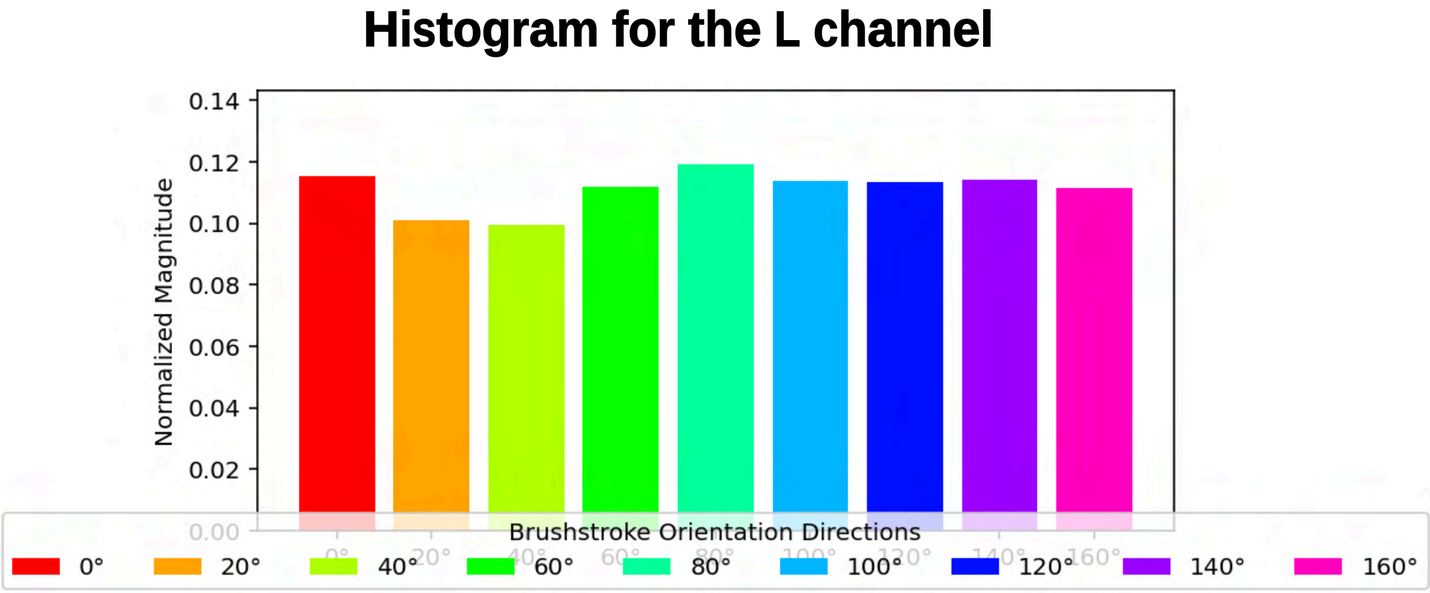


HOG on Blue/Yellow channel



4. Go through each “area” and make a histogram out of their magnitude and orientations.





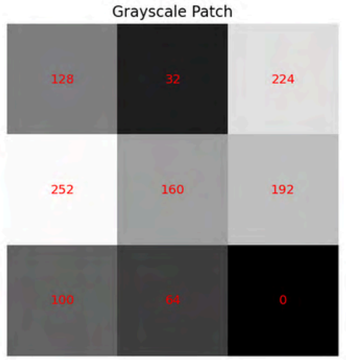
Grouping the gradients allow us to quantify the motion and rhythm of the artists’ hand.

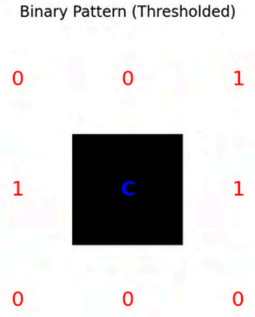
5. Each area would have 9 values corresponding to each bin. Concatenate them into a feature vector that has 9 entries.

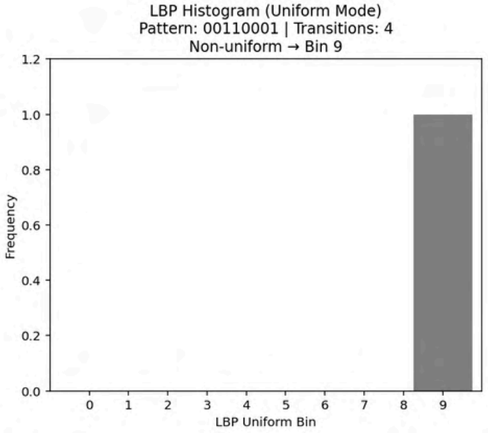
$$\mathrm{HOG}_{final} = 0.7\,\mathrm{HOG}_L + 0.15\,\mathrm{HOG}_A + 0.15\,\mathrm{HOG}_B$$

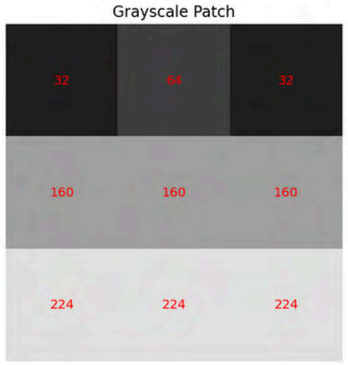
Texture is a key distinguishing feature among painters. It is difficult to replicate or fake this texture and the underlying canvas material.

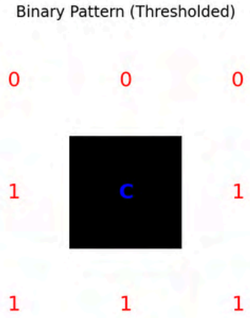
Feature Extraction #3: Local Binary Pattern

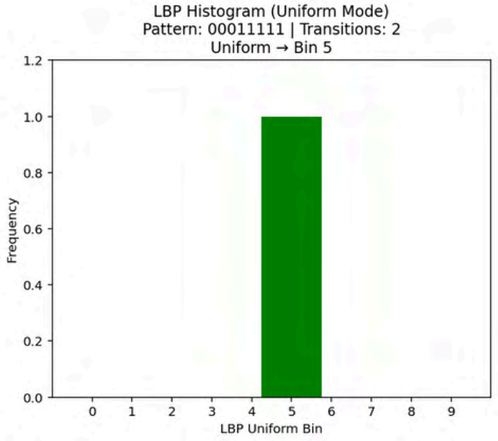










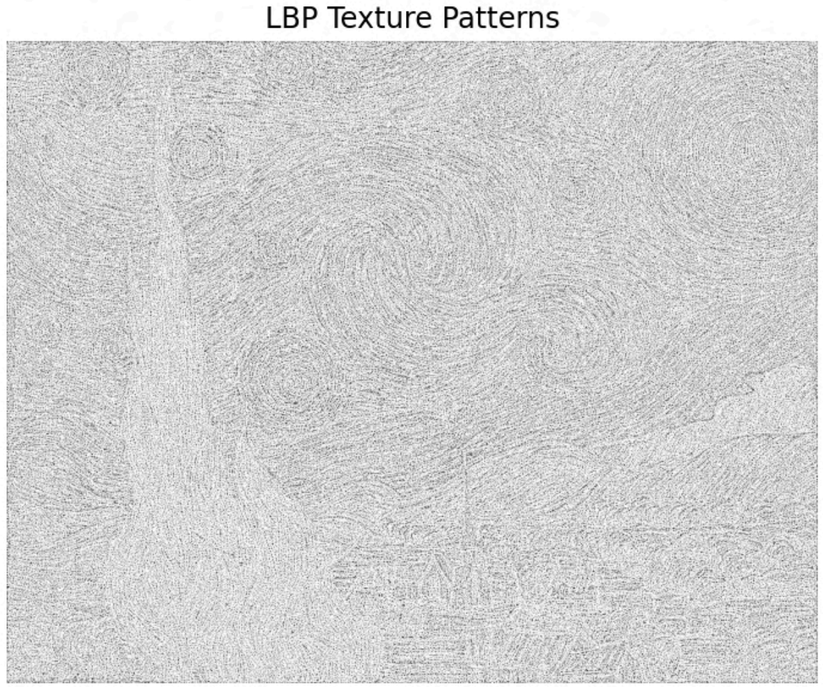


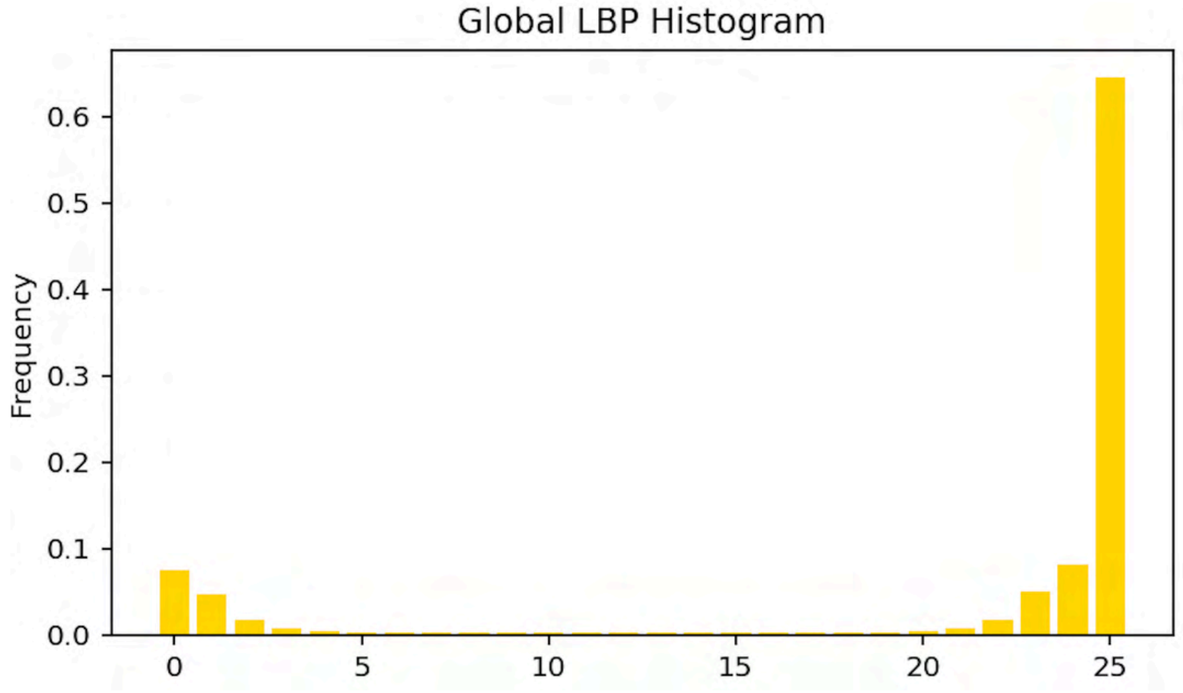
1.Find the pixel values of the 8 blocks that surrounds it

2. Assign 0 to the blocks darker than the center block and assign 1 otherwise

3. group the binary combinations into bins.
If no. transition <= 2 → put into the bin with the number of 1s
if not → put into bin 9

If neighbors ≈ center pixel → smooth area
If neighbors ≠ center pixel → edges or texture





Van Gogh's works show texture from thick brushstrokes, even in uniform color areas.

High binary variation = rough surface
Feature vector: 26 values capturing local texture variation

Train/test csv files containing
Feature Extraction
with each row labled 0 or 1

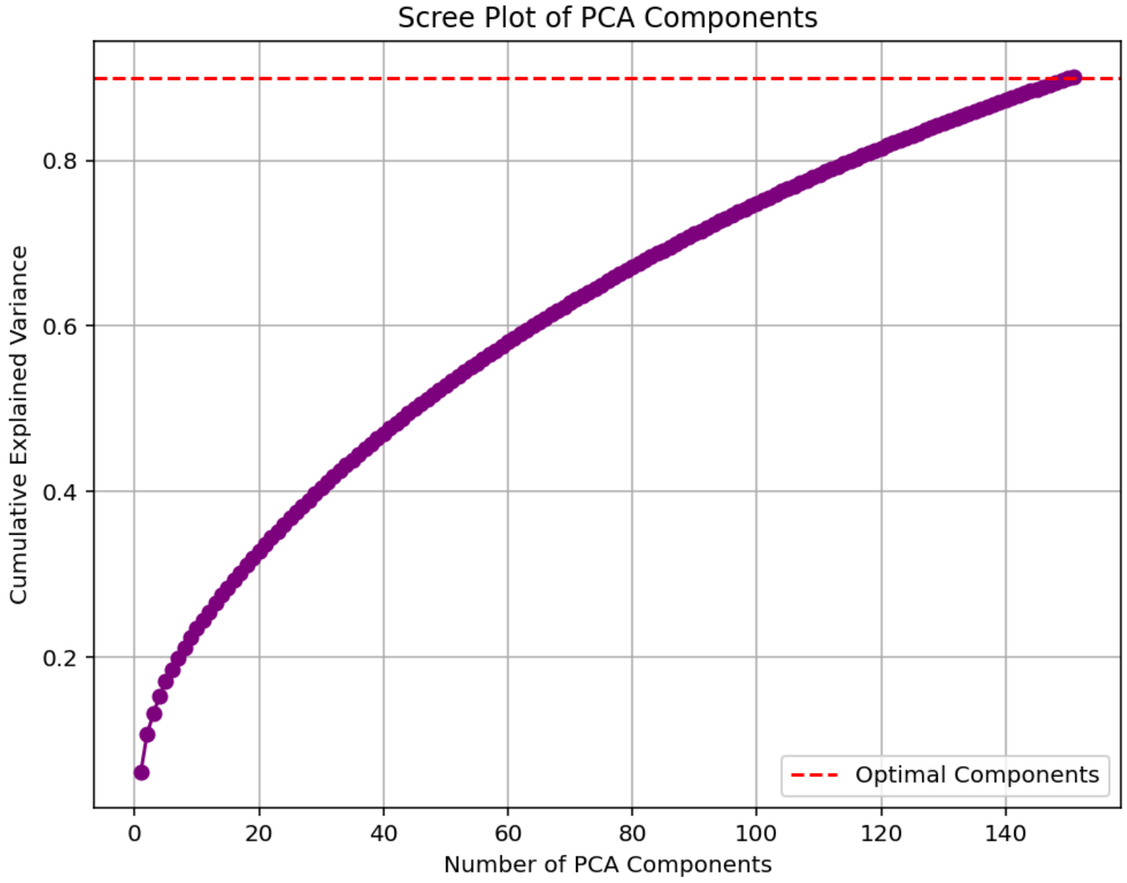
Preprocessing: standardization of features and PCA

Model: Support Vector Classifier
pca components: 0.9,
svm__C: 10.0,
svm__gamma: 'scale'

Evaluation metrics:

1. **ROC curve and AUC scores**
2. **Specificity**
3. **Classification report**

PCA Dimensionality Reduction

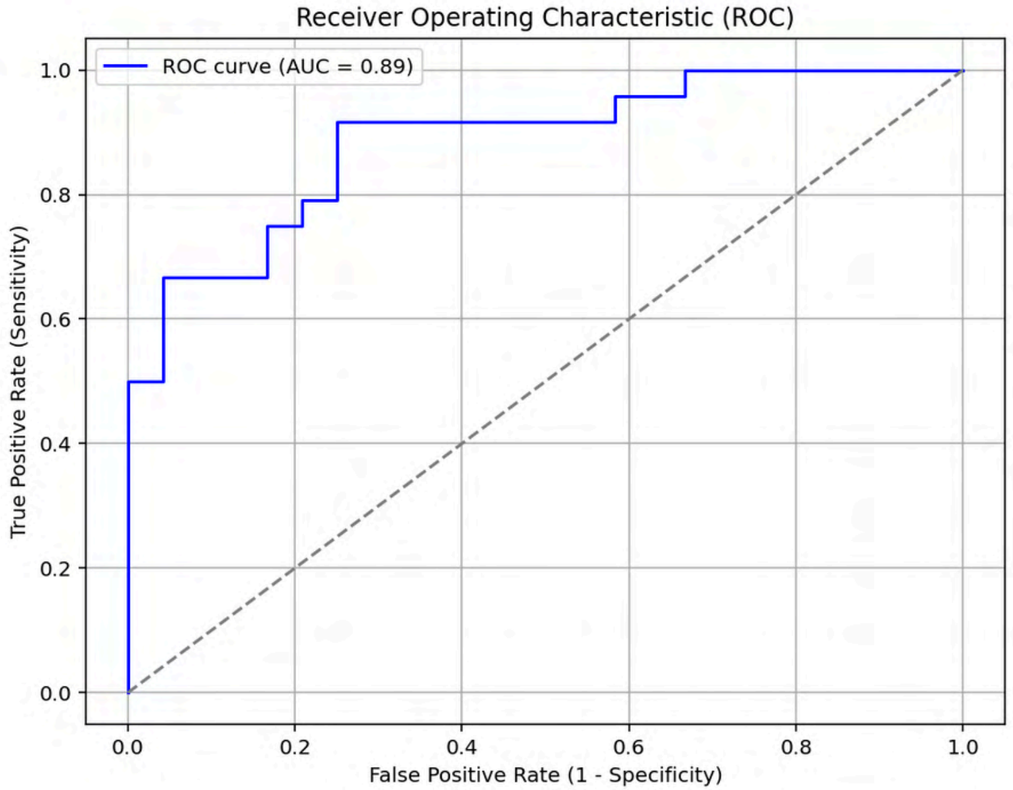


Dimensionality is reduced from 34718 features to 145.

Only 0.4% of the original dimensions are needed to explain 90% of the statistical variation in features.

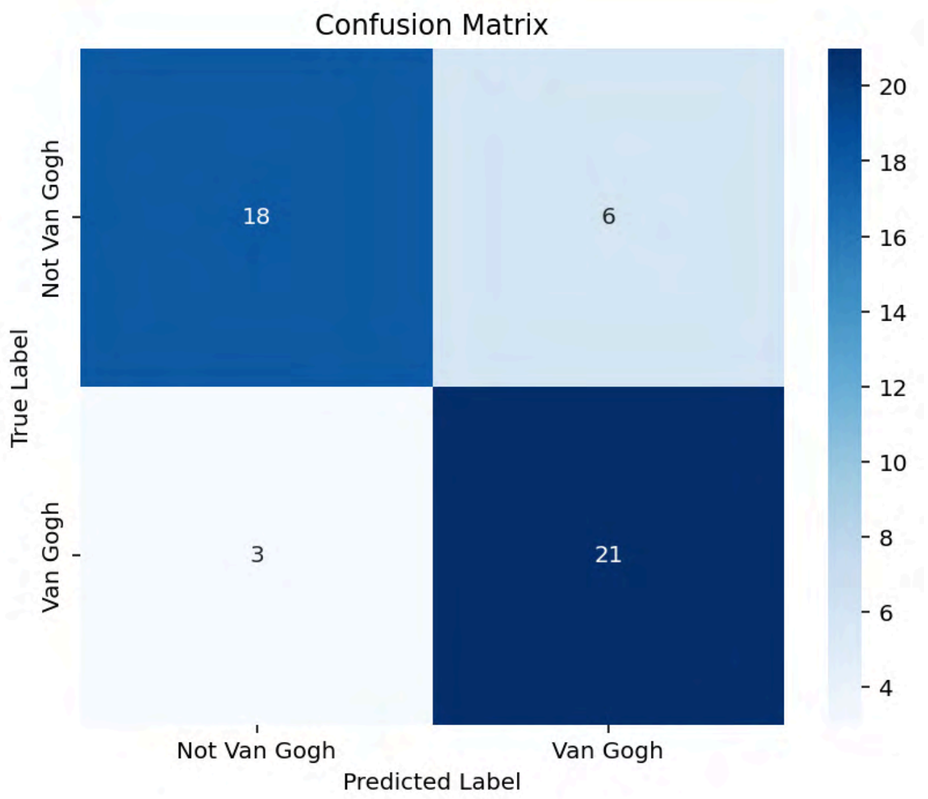
Many features are likely to be correlated and redundant.

Model Results



AUC: 0.89

Specificity: 75%



Recall (Van Gogh): 88%

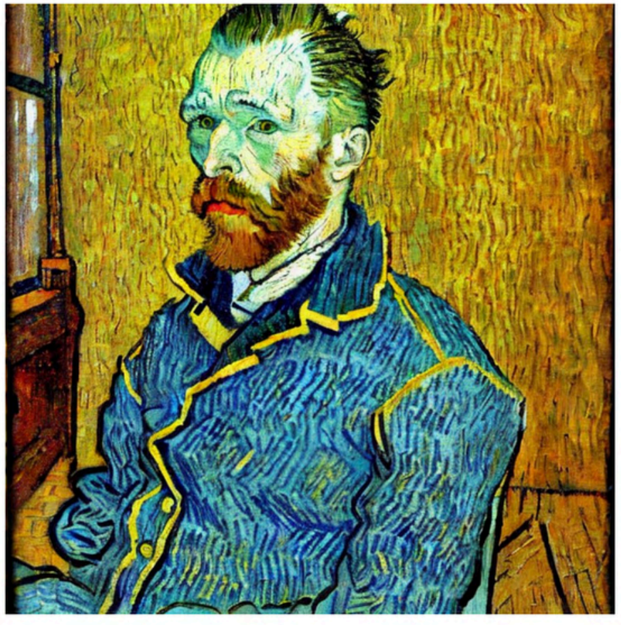
Precision (Van Gogh): 78%

False Negative
Real Van Gogh → Predicted Fake

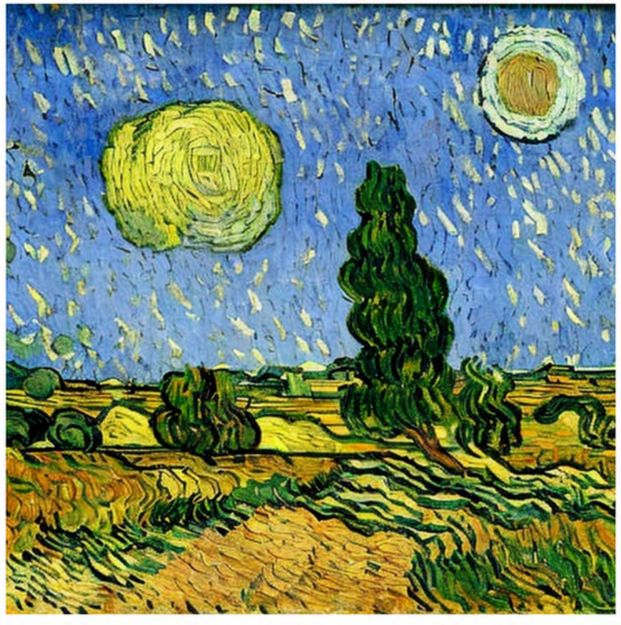
False Negative
Real Van Gogh → Predicted Fake

False Negative
Real Van Gogh → Predicted Fake

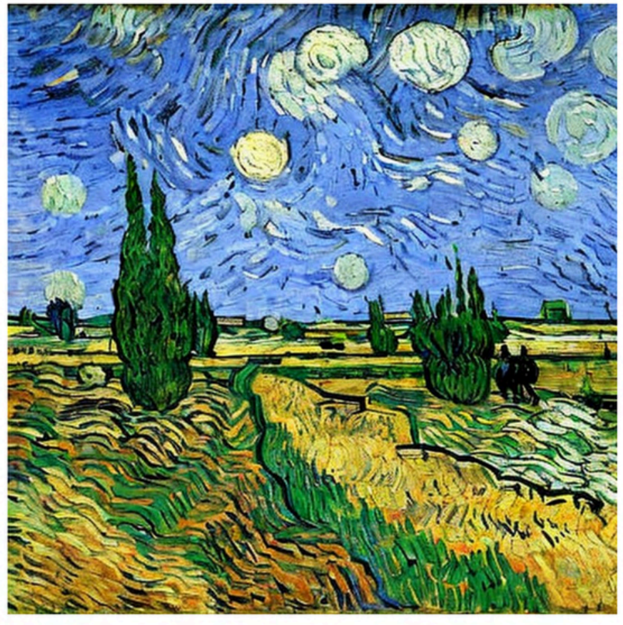
Not Van Gogh → Predicted Real



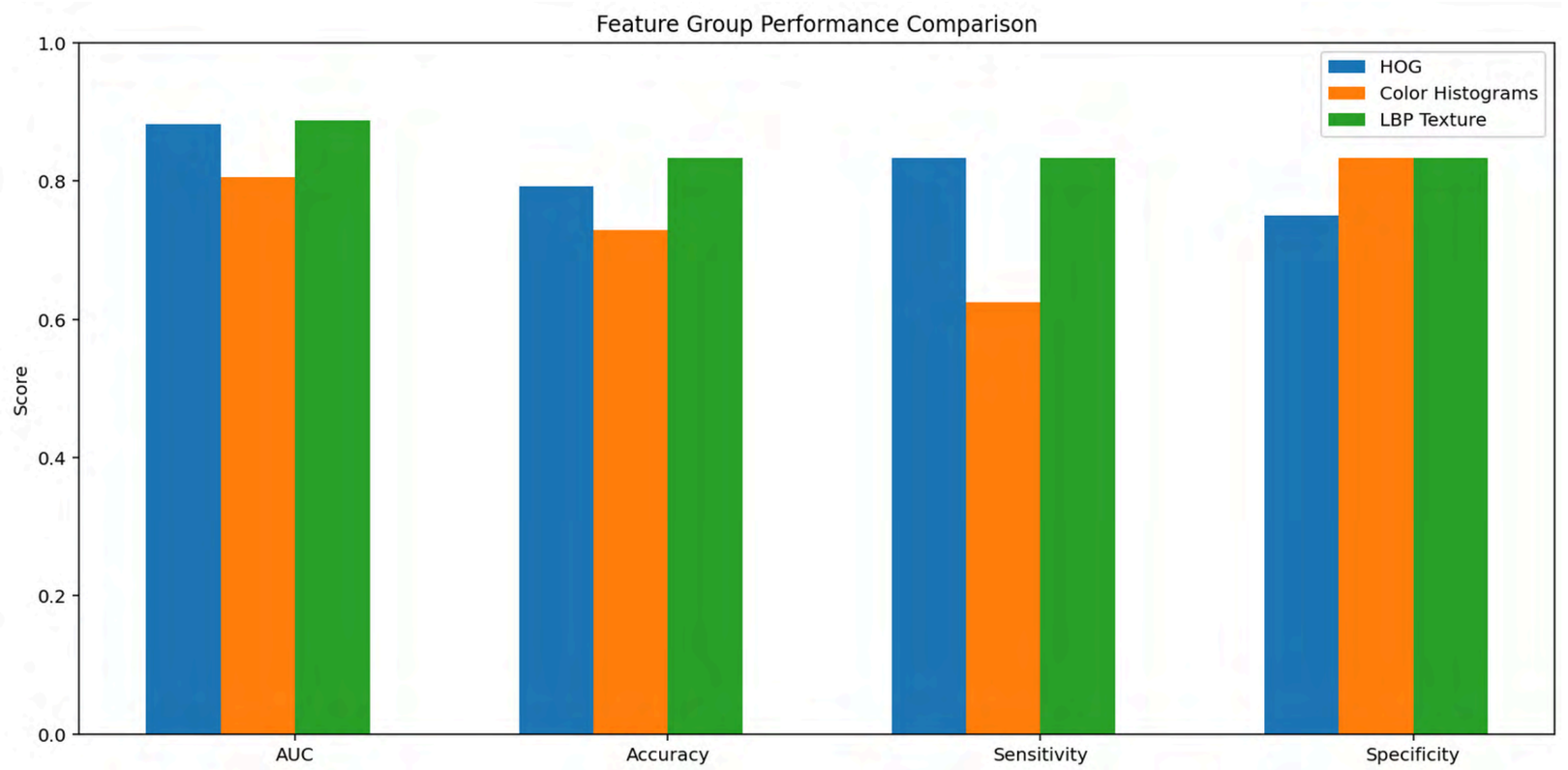
Not Van Gogh → Predicted Real



Not Van Gogh → Predicted Real



Which feature is the best?



Local Binary Pattern

Histogram of oriented gradients

Color histogram

Area of Improvements

01

Expand the dataset through additional data collection or data augmentation to reduce overfitting and improve generalization.

02

experiment with more feature extraction methods (CNN, hybrid approach) and find out which one yield the greatest accuracy.

03

Develop a multi-artist classification system, allowing the model to not only verify Van Gogh but also compare paintings across a broader artist database.


THANK YOU
FOR LISTENING