

Art Authentication

Combating Art Forgery With Technology

Which is the authentic Van Gogh?



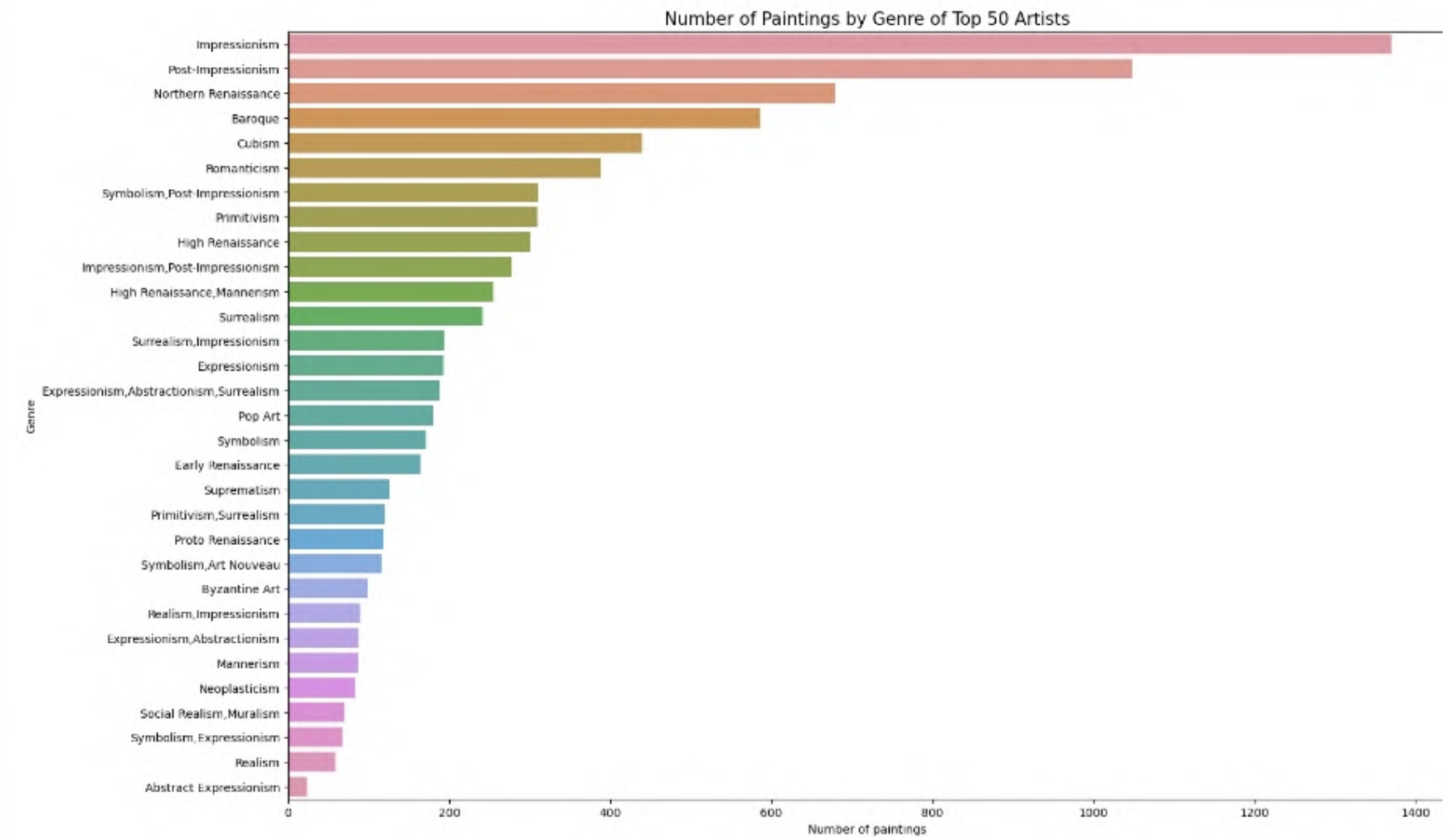
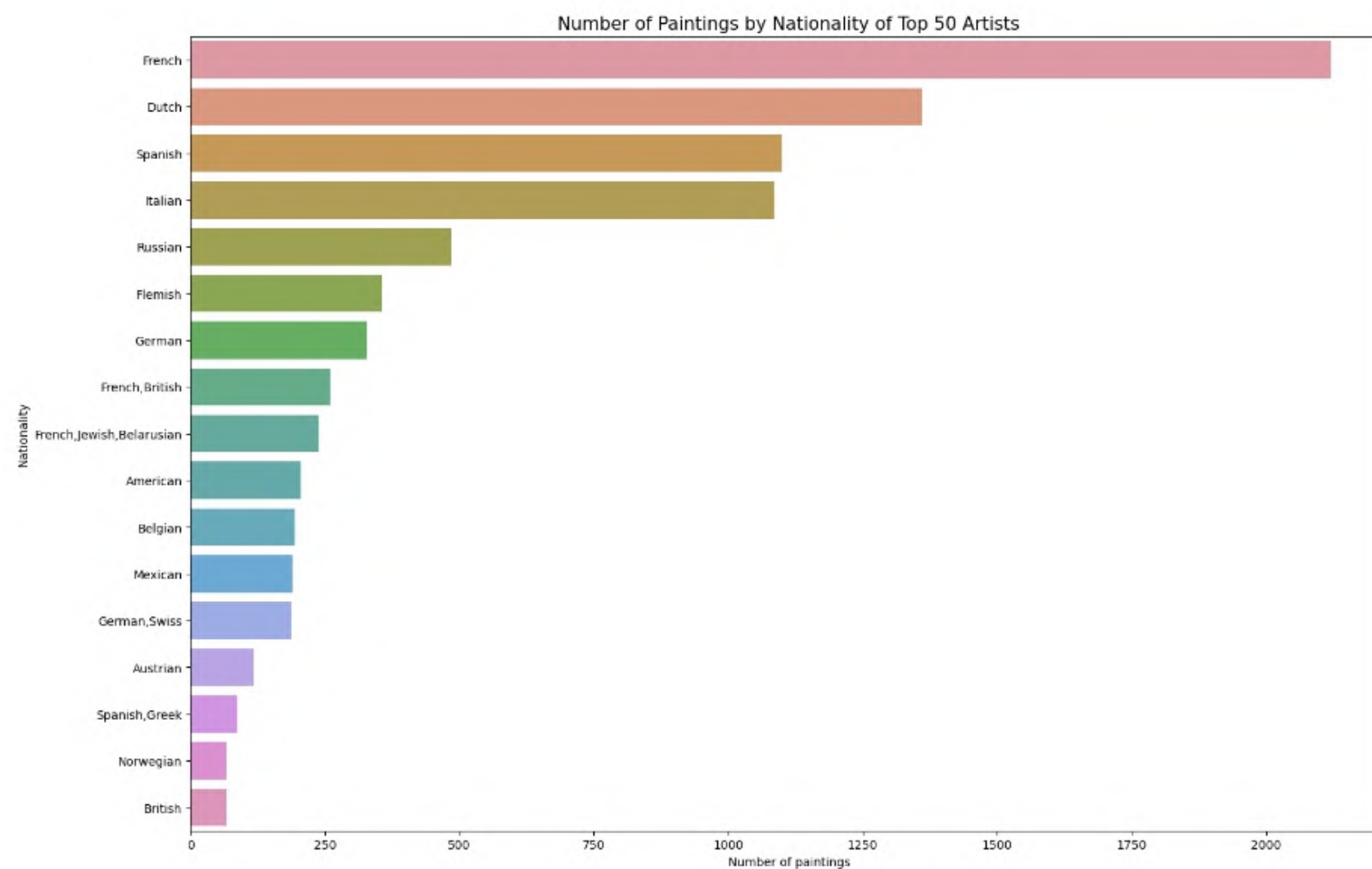
Business Problem

About half of all fine art in global circulation is suspected to be fraudulent. The art industry is estimated to generate \$60+ billion annually, implying that the forgery industry is worth around \$30+ billion. Art forgery is the third-highest grossing criminal enterprise of the last 40 years.

Van Gogh is one of the most popular Post-Impressionist painters and a prime target for forgeries. Even with expert analysis, it is hard for people to distinguish fakes, as seen when nine paintings previously designated as fakes have been determined to be real Van Gogh works.

I aim to use machine learning image classification models to better analyze and predict true Van Gogh paintings from forgeries, imitations, and even AI generated version of his art.

The original dataset contained 8,400+ artworks from the top 50 artists in art history, of which 877 were Van Gogh paintings. The model was also trained on AI generated Van Gogh art, Van Gogh replica paintings and Van Gogh style imitation paintings. The primary nationality of the artists was French and most painted in Impressionist or Post-Impressionism styles. The dataset was sampled to only keep a similar amount of non-Van Gogh artworks.

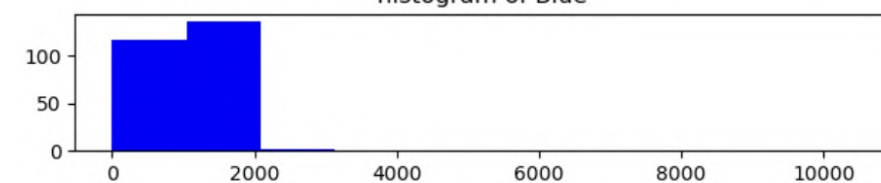


Van Gogh Painting

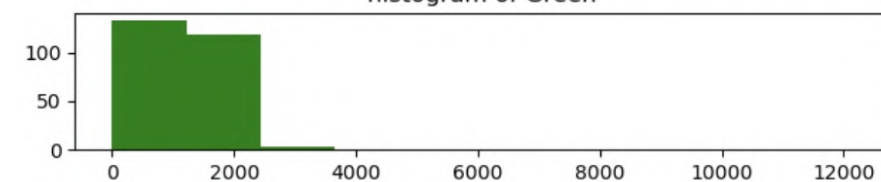
Original Image



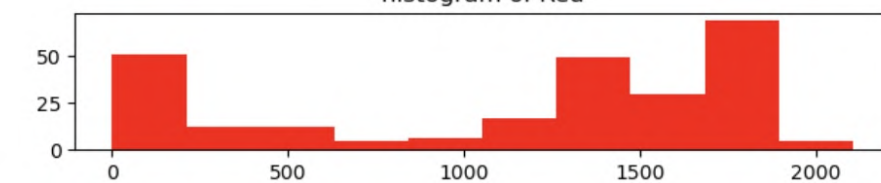
histogram of Blue



histogram of Green



histogram of Red

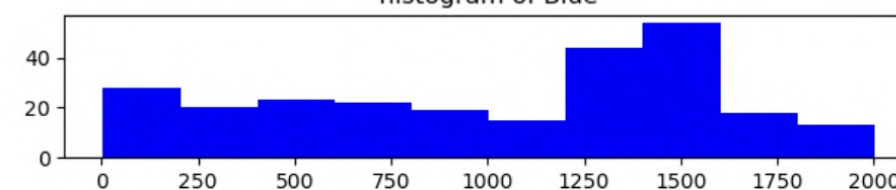


AI Generated

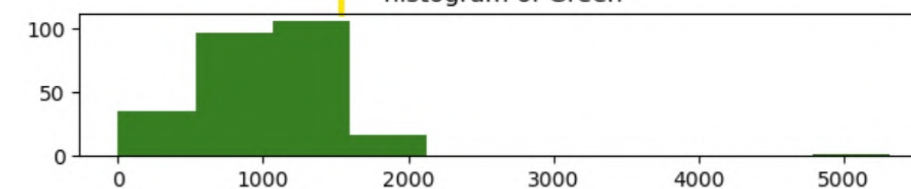
Original Image



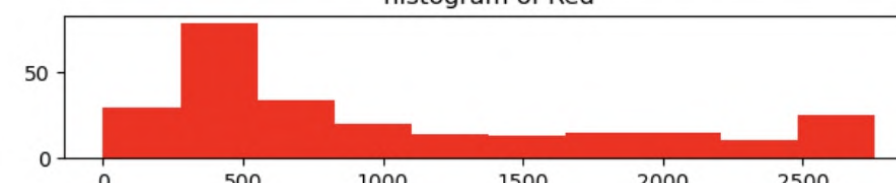
histogram of Blue



histogram of Green



histogram of Red

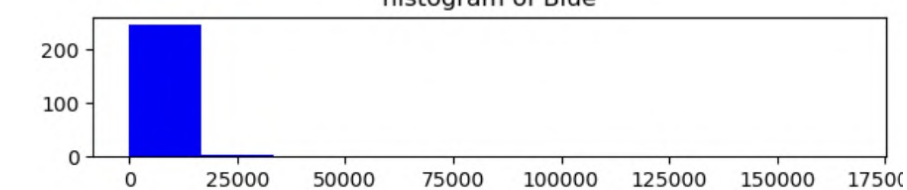


Caravaggio

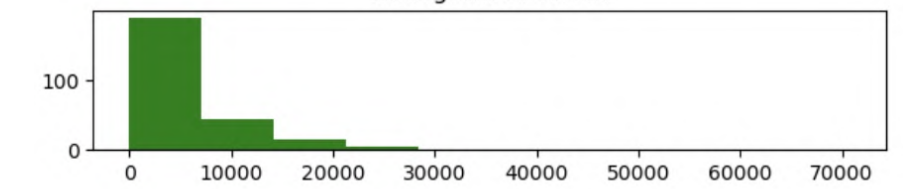
Original Image



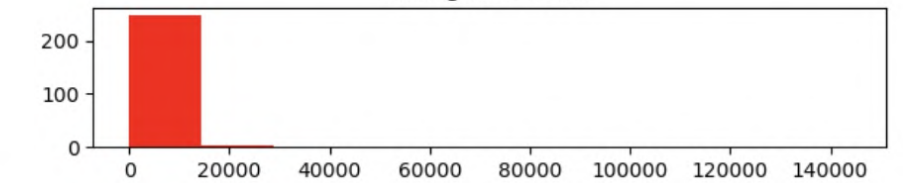
histogram of Blue



histogram of Green



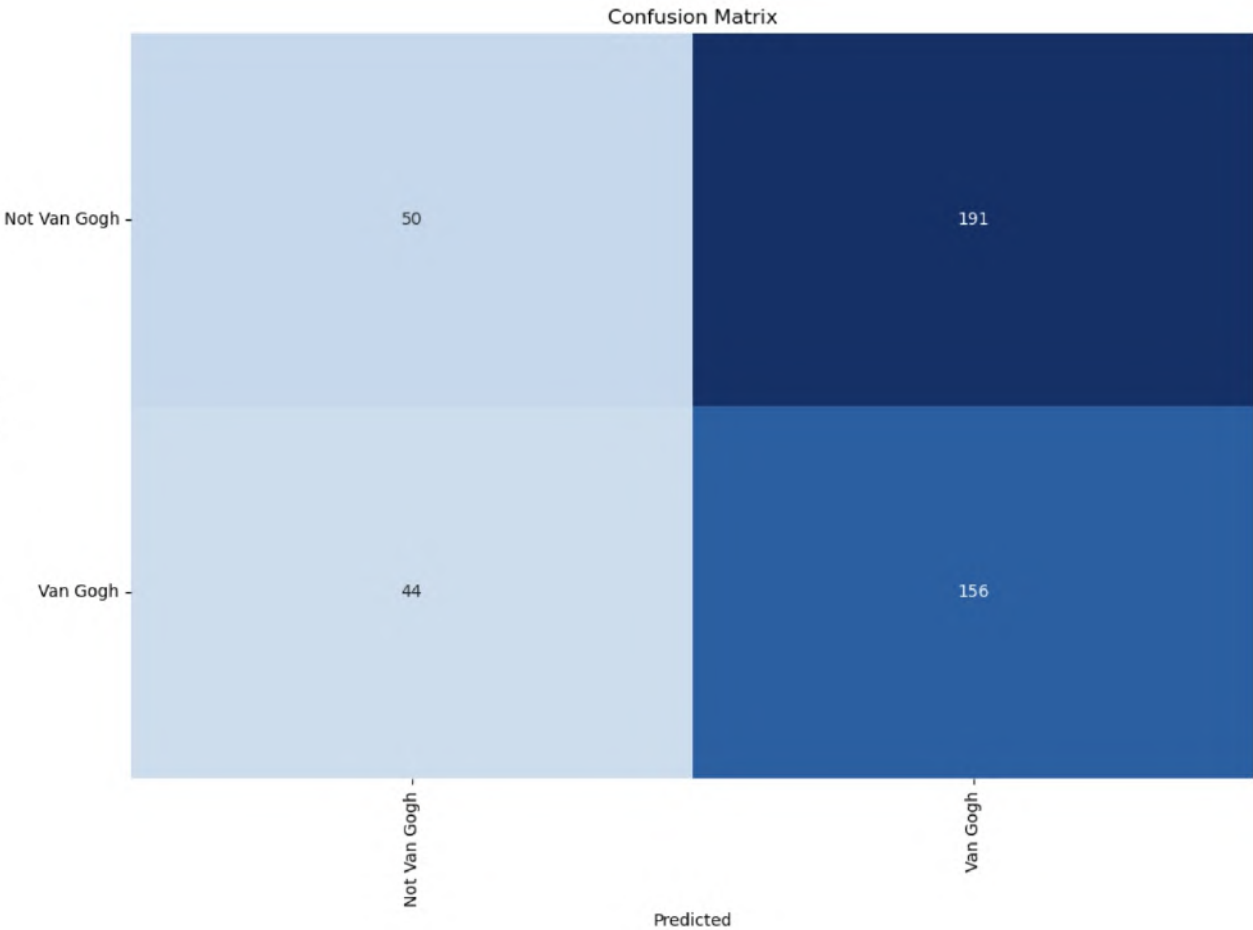
histogram of Red



The best Image Classification model was 53% accurate when determining non-Van Gogh paintings and 45% accurate when determining Van Gogh paintings.

According to the confusion matrix, the model primarily had trouble overclassifying non-Van Gogh art as being one of his paintings.

Layer (type)	Output Shape	Param #
=====		
conv2d (Conv2D)	(None, None, None, 32)	416
activation (Activation)	(None, None, None, 32)	0
max_pooling2d (MaxPooling2D)	(None, None, None, 32)	0
conv2d_1 (Conv2D)	(None, None, None, 64)	8256
activation_1 (Activation)	(None, None, None, 64)	0
max_pooling2d_1 (MaxPooling2D)	(None, None, None, 64)	0
conv2d_2 (Conv2D)	(None, None, None, 64)	16448
activation_2 (Activation)	(None, None, None, 64)	0
max_pooling2d_2 (MaxPooling2D)	(None, None, None, 64)	0
flatten (Flatten)	(None, None)	0
dense (Dense)	(None, 128)	1843328
dense_1 (Dense)	(None, 64)	8256
dense_2 (Dense)	(None, 2)	130
=====		
Total params: 1,876,834		
Trainable params: 1,876,834		
Non-trainable params: 0		
=====		



Classification Report:

	precision	recall	f1-score	support
Not Van Gogh	0.5319	0.2075	0.2985	241
Van Gogh	0.4496	0.7800	0.5704	200
accuracy			0.4671	441
macro avg	0.4907	0.4937	0.4344	441
weighted avg	0.4946	0.4671	0.4218	441

Conclusion

Currently the main forms of art identification are establishing provenance or highly invasive microscopy/mass spectrometry/x-ray/infrared technology studies. When it comes to machine learning algorithms, the analysis can be run on images alone and no transport of the original artwork is necessary. Furthermore, the method is non-invasive to the artwork, the results are available in a matter of hours or days.

Human authentication can be more fallible than people in the art world would want to believe. Machine learning algorithms can improve decision making when it comes to art forgeries and bring much needed objectivity to the authentication process.