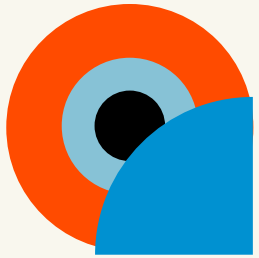
Utilizing Convolutional Neural Networks to Predict Style of Paintings.

Daniel Wang | Nicholas Nguyen | Steven Shields | Kaytlynn Skibo
General Assembly DSI Project 5 | 03/24/23

SLIDESMANIA.COM

Table of Contents.



Introduction & Background

Why might someone need this ML model? (The backstory) Vocabulary

Data Collection & Cleaning

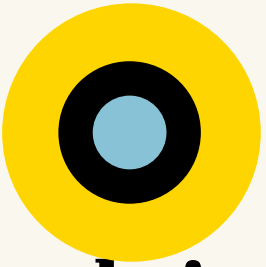
The origins of our dataset and our process for preparing it.

EDA & Visualization

Interesting snippets and statistics about our dataset.

Modeling Process

The different models we evaluated (& how they differed).



Conclusion & Next Steps

Results, explanations, steps forward, and the epilogue to our story.

BONUS: Audience interaction!

SLIDESMANIA.COM

Can a painting's art style be predicted by using the image in a CNN model?

Photo by Allef Vinicius on Unsplash

- Neural networks as a "magic bullet".
- So. Many. Art styles.
- Can be hard to distinguish between, even for art ~~experts~~ aficionados.
- CNN for predicting art style from image pixel data.

SLIDESMANIA.COM

Story Introduction / Problem Statement

- **Art thieves** who pulled off the heist of the century.
- Team's resident art expert was KIA.
- No idea how to fence the stolen art without **knowing the art styles**.
- Getting a **new art expert** is too risky.
- **Train a CNN machine learning model** instead!

SLIDESMANIA.COM

Neural Net Vocabulary.

- Convolutional Neural Network (CNN)
- RGB Pixel Arrays
- Convolution Filters
- Kernel Size
- Batch Normalization
- Regularization
- Dropout Layer
- Dense Layer

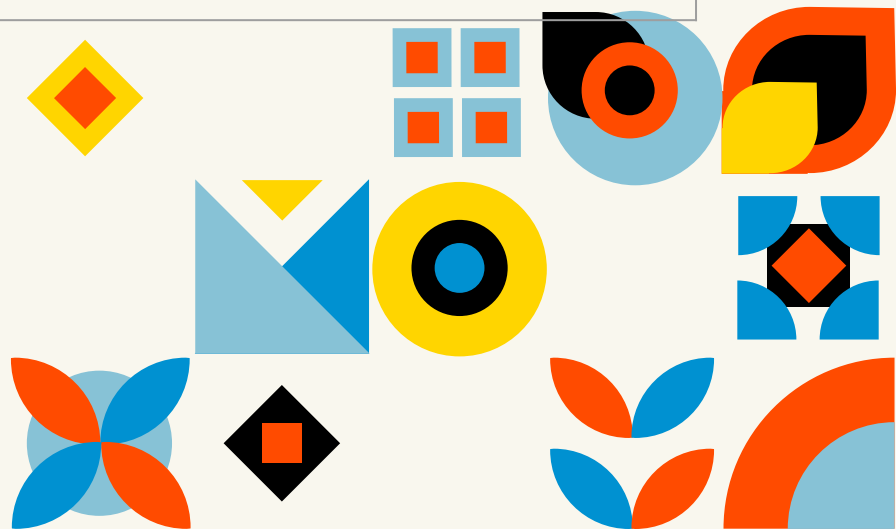
SLIDESMANIA.COM

Data Dictionary.

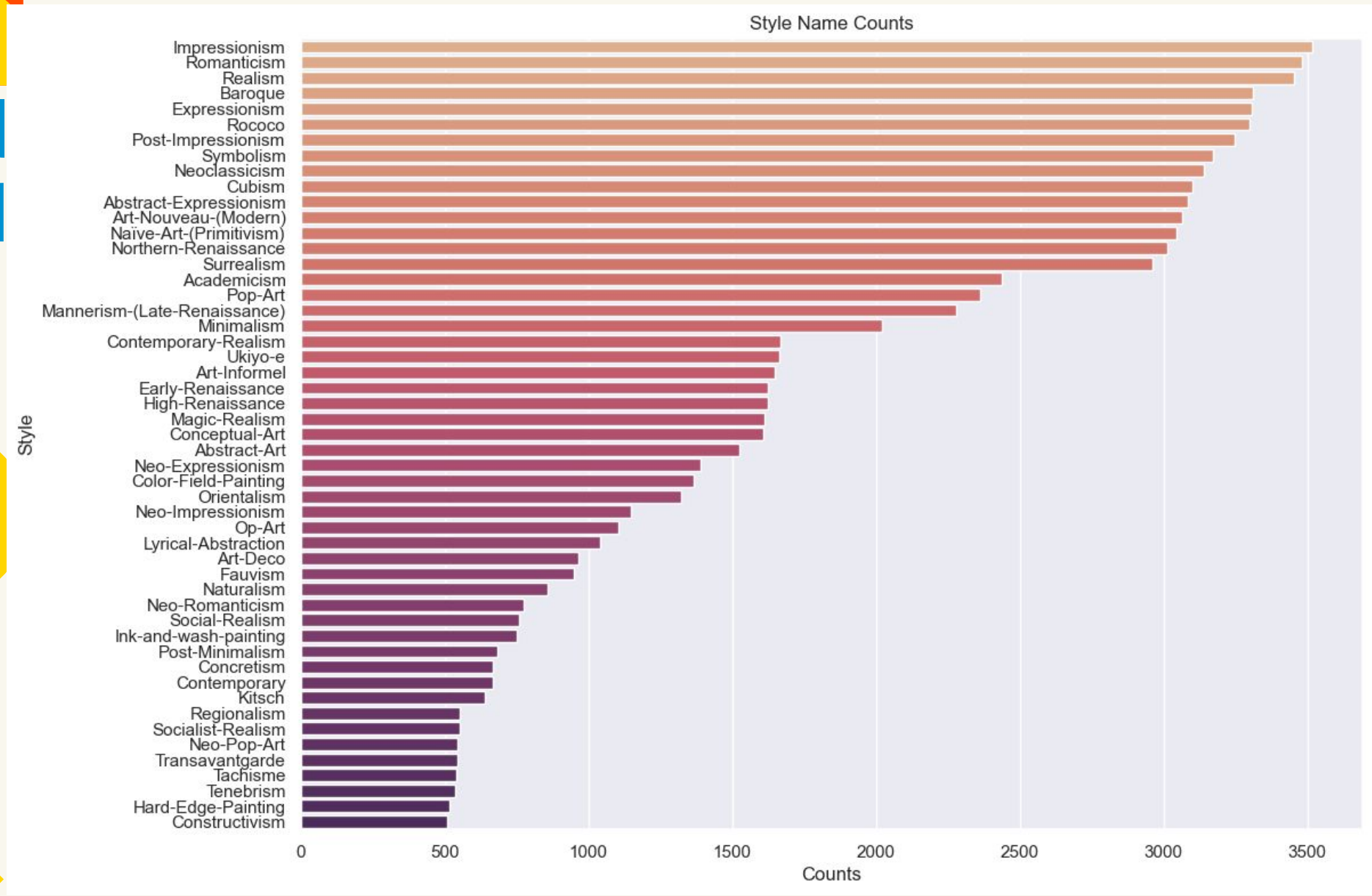
Style -artwork style	**Link** - url for image	**V_Sent** - sentiment analysis
Artwork - art name	**Language** - origin language	**Hex** - color code
Artist	**Translated** - translated art name	**Color** - main color of work
Date	**Style Num** - number for style	

Data Description:

- Data Source : Kaggle
- Data Size : ~120k x 5 columns elements before cleaning
- Data Shape : 89,545 rows x 11 columns after cleaning
-



SLIDESMANIA.COM



SLIDESMANIA.COM

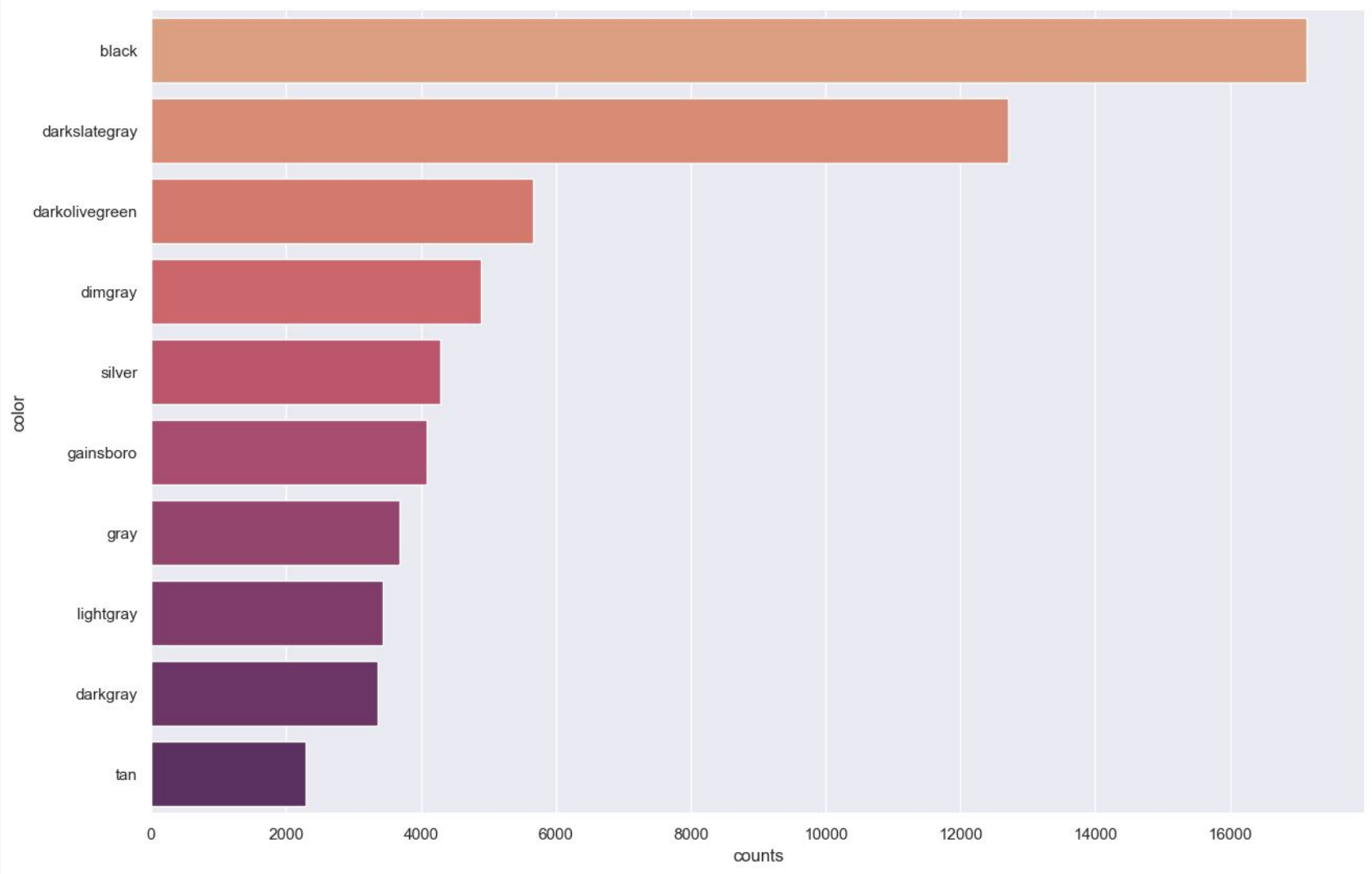
Data Cleaning

- Drop bad URL's
- Drop duplicate artwork
  - Sketches for final work
- Drop styles with less than 500 pieces
- Date column
  - Remove hyphens from date ranges
    - Keep first date
  - Change century dates to numeric
    - VII → 600 (ad)
    - Drop BC dates

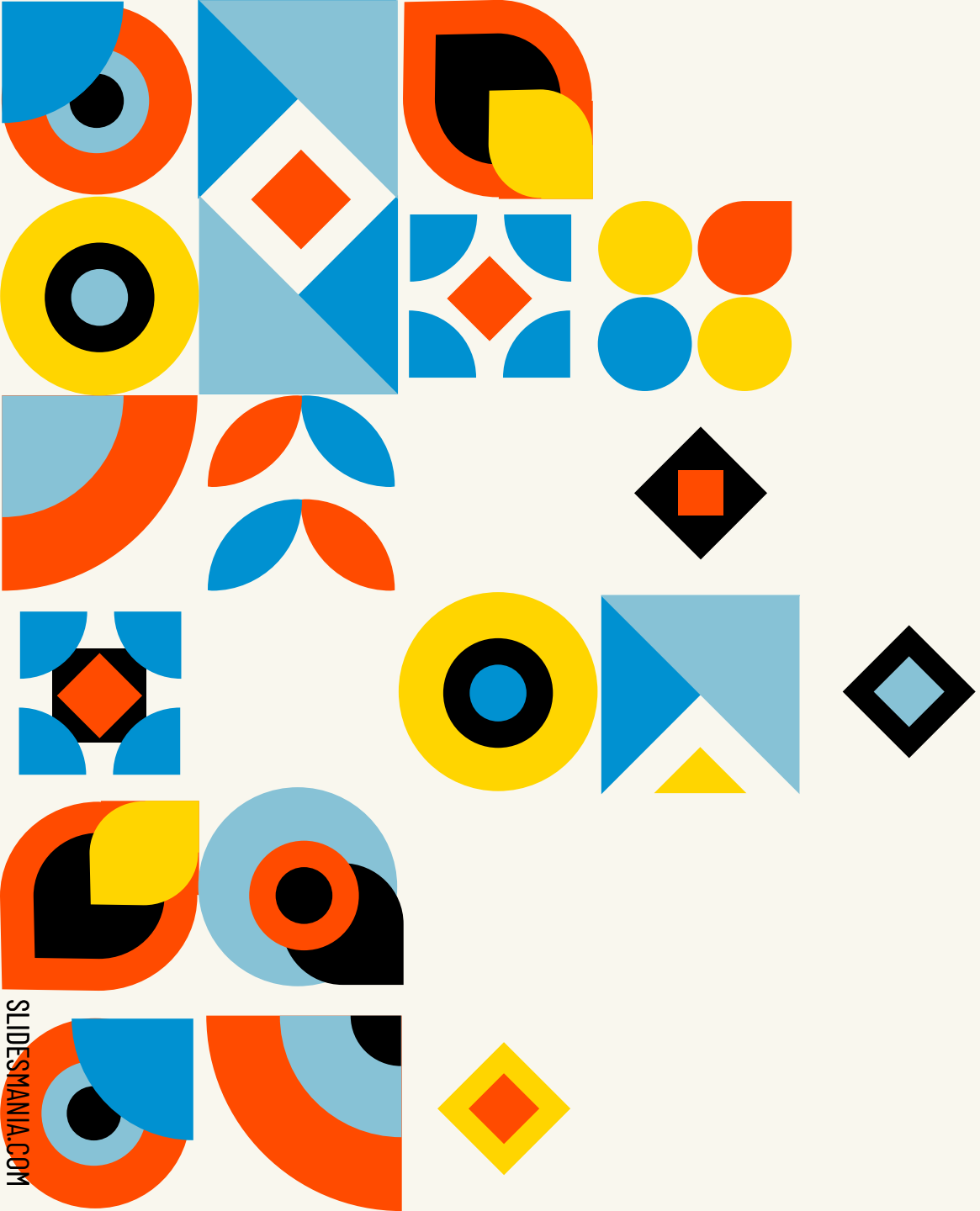
SLIDESMANIA.COM

Feature Engineering

- Translated
- V_sent
- Style Number
- Color
- Language
- Hex



SLIDESMANIA.COM



EDA!

SLIDESMANIA.COM

Scatter of Style by Date
v_sent_trans
0.5
0.0
−0.5
Style
Abstract-Art
Abstract-Expressionism
Academicism
Art-Deco
Art-Informel
Art-Nouveau-(Modern)
Baroque
Color-Field-Painting
Conceptual-Art
Concretism
Constructivism
Contemporary
Contemporary-Realism
Cubism
Early-Renaissance
Expressionism
Fauvism
Hard-Edge-Painting
High-Renaissance
Impressionism
Ink-and-wash-painting
Kitsch
Lyrical-Abstraction
Magic-Realism
Mannerism-(Late-Renaissance)
Minimalism
Naturalism
Naïve-Art-(Primitivism)
Neo-Expressionism
Neo-Impressionism
Neo-Pop-Art
Neo-Romanticism
Neoclassicism
Northern-Renaissance
Op-Art
Orientalism
Pop-Art
Post-Impressionism
Post-Minimalism
Realism
Regionalism
Rococo
Romanticism
Social-Realism
Socialist-Realism
Surrealism
Symbolism
Tachisme
Tenebrism
Transavantgarde
Ukiyo-e
700
780
860
940
1,020
1,100
1,180
1,260
1,340
1,420
1,500
1,580
1,660
1,740
1,820
1,900
1,980
2,060
Date

Tableau Dashboard

SLIDESMANIA.COM

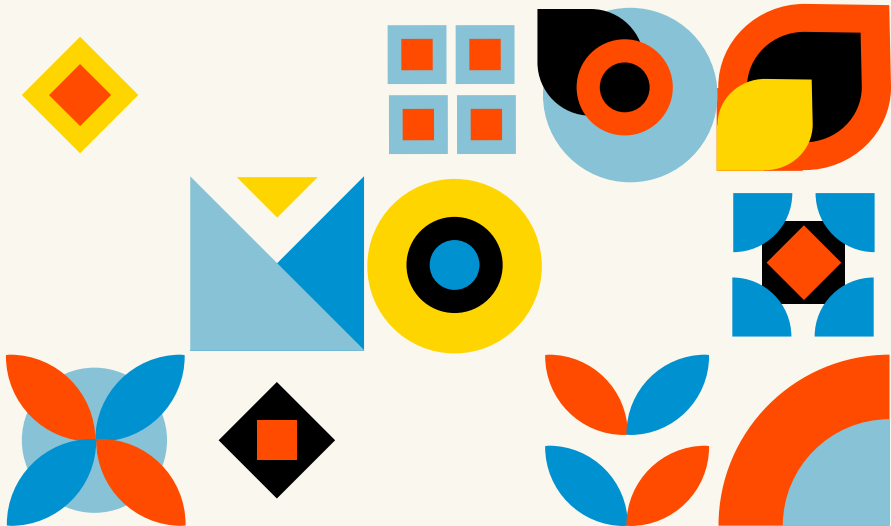
Photo by Steve Johnson on Unsplash

Preprocessing

Problems in preparing data for modeling

- Converting imgs into arrays
- Filepaths to imgs
- Creating DF of matrices

The solution?

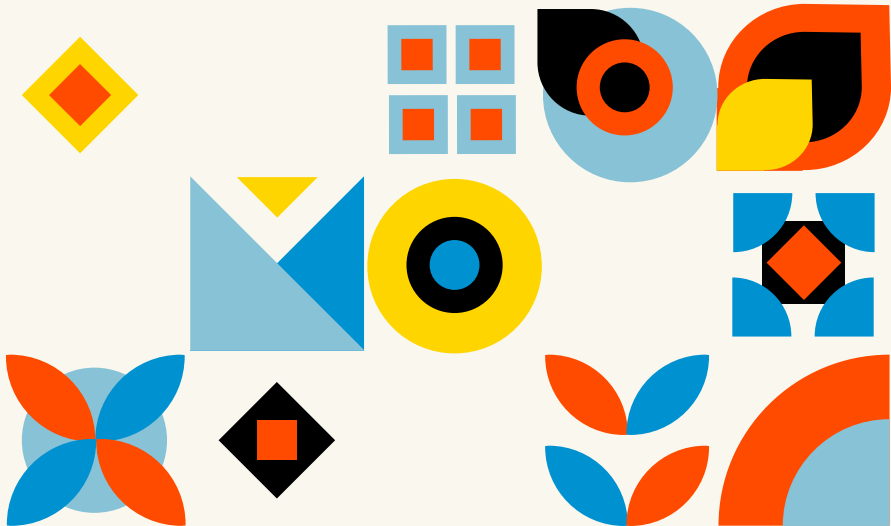


SLIDESMANIA.COM

Preprocessing Solution

Tensorflow Keras (*image_dataset_from_directory* function)

- Faster, easier, higher quality
- BatchDataset
- 250x250 pixels
- Categorical Labels

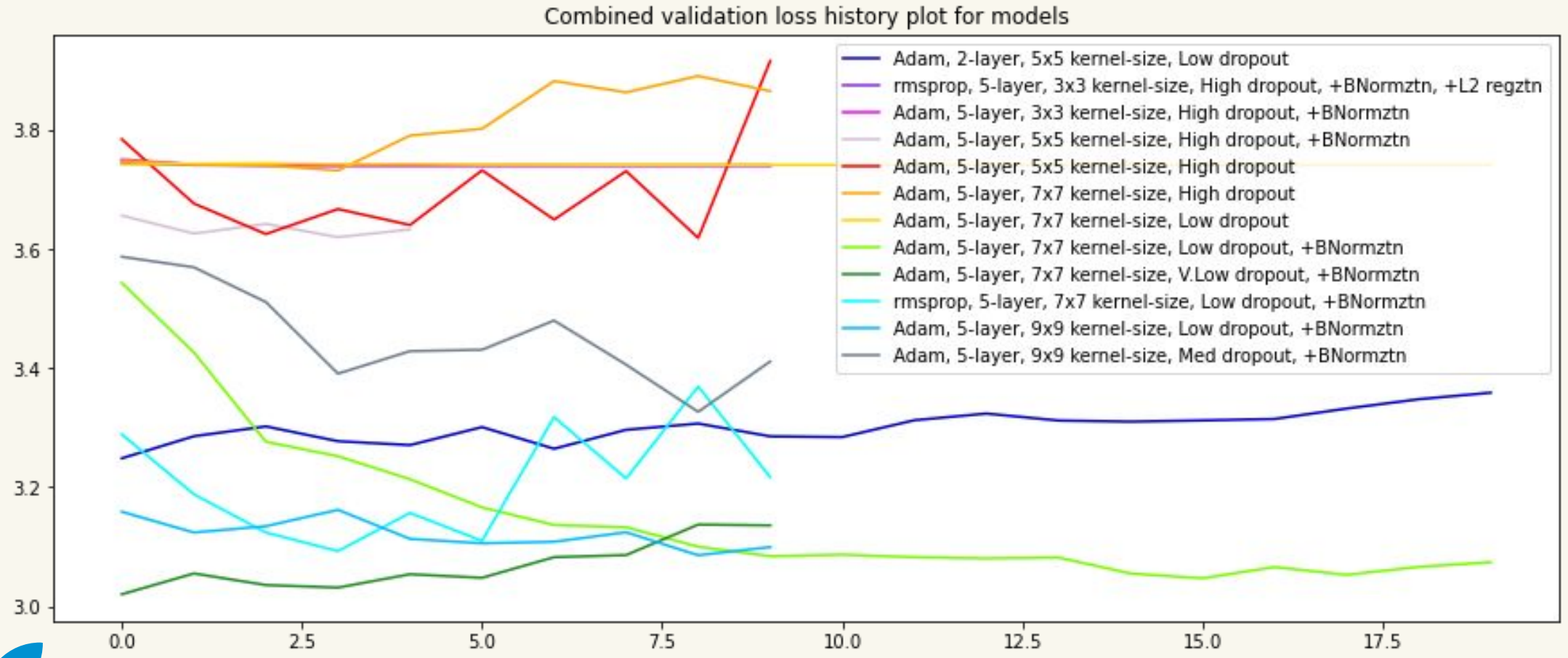


SLIDESMANIA.COM

Photo by Steve Johnson on Unsplash

SLIDESMANIA.COM

Sometimes, things don't always go to plan



SLIDESMANIA.COM

Model Summaries.

baseline: .0397

Model	Train Loss	Validation Loss	Train Acc	Validation Acc
CNN with rmsprop	3.2450	3.2479	0.1456	0.1326
CNN with 3 Dense Layers	2.8841	3.079	0.1989	0.1672
CNN with 9x9 Kernel Size and Dropout (.1 - .2)	3.0910	3.0995	0.1628	0.1599
CNN with BatchNormalization and Dropout (.5-.1)	2.6489	3.159	0.2560	0.1797
CNN with 5 Filters, 7x7 Kernel Size, Dropout(.1 - .2), & BatchNormalization	3.0741	3.0341	0.1742	0.1783

SLIDESMANIA.COM

Recommendations

- Have some prior knowledge on what you're about to steal, helps a lot
- Don't steal 100,000 artworks at one time, or at least make sure they're all the same style
- Making a machine learning model can takes more time than the cops are giving us
- The model can't direct us much, so lets hope museums have a buy-back program

SLIDESMANIA.COM

Next Steps.

- More Time, More Processing Power
- Combine Classes or Reduce Them
- Less downsizing, Experiment with more Models

SLIDESMANIA.COM

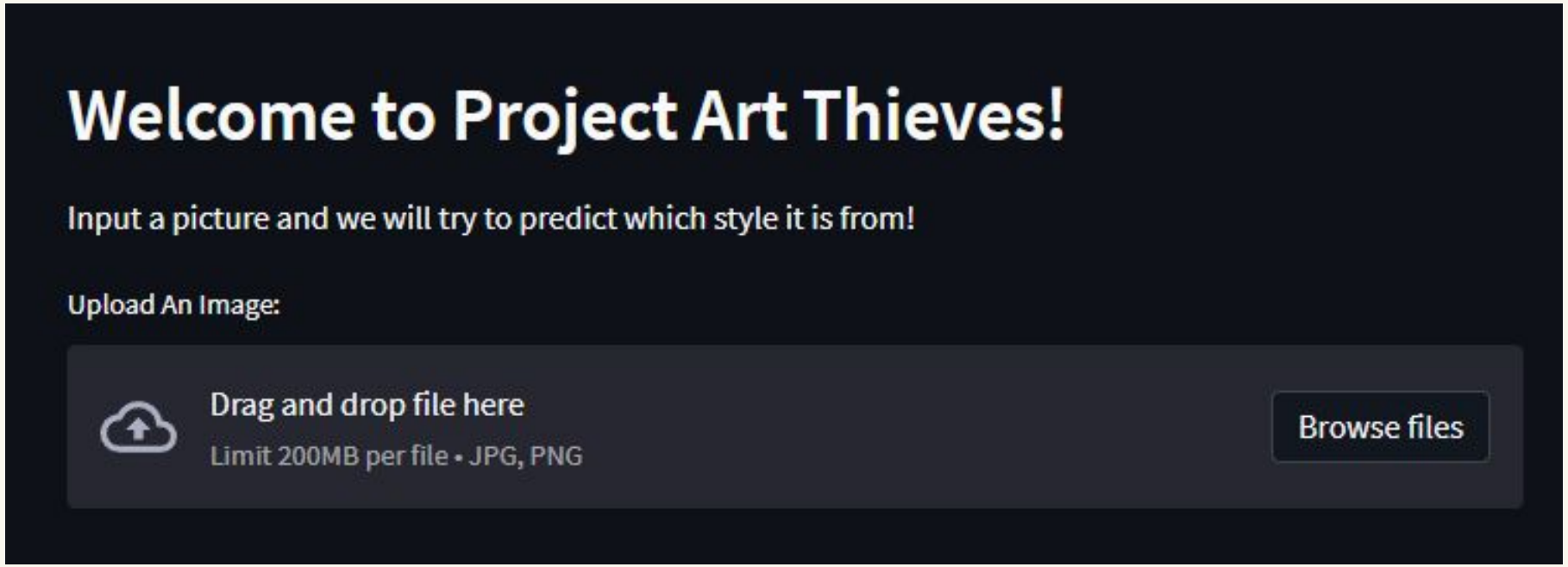
Expressionism or Romanticism?

SLIDESMANIA.COM

Conclusion: Failed Model but Successful Experience

SLIDESMANIA.COM

Streamlit Application!



SLIDESMANIA.COM

Questions ?

SLIDESMANIA.COM