

Rachel Dilley

LET'S TAKE A TRIP

A US ATTRACTION RECOMMENDER

TABLE OF CONTENTS

MOTIVATION
& USE CASES

DATA

MODEL

01

02

03

05

04

WEB APP DEMO

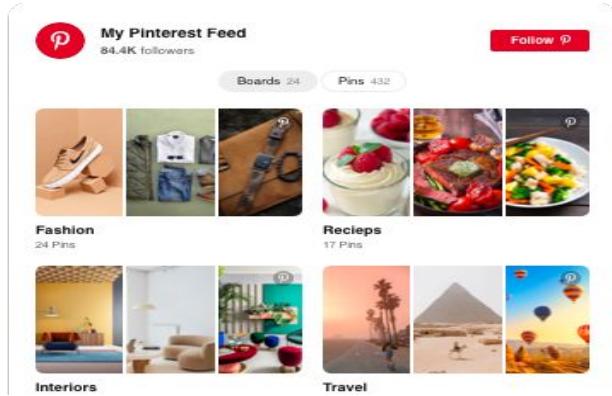
FINAL RESULTS



MOTIVATION

With so many options to choose from,
vacation planning can be stressful

- What if you want to recreate a past vacation experience in a different location?
- You came across a cool image online that looks like a good destination for your next vacation, but you're not sure where the image was taken



A screenshot of a Facebook profile for "Roxy". The news feed shows updates from friends like Kassie Beyer, Alex Wade, and Lauren Useman. It also shows a post from Alex Wade with new photos from Cornell. The right sidebar includes sections for Requests, Pokes, and My Status.

BUSINESS USES

- Targeted marketing for businesses
- Market attractions to consumers based on images on their social media accounts



DATA

- Tourist uploaded images scraped from  **Tripadvisor**
- Top 30 attractions from each of the continental US states
- 75,000 images from 1,500 attractions



MODEL



ATTRACTION LABELING

- Topic modeling on attraction names
- NMF
- TF-IDF vectorization



IMAGE CLASSIFICATION

- Neural network trained on features for image classification
- Transfer learning with Pre-trained CNN (VGG-16)



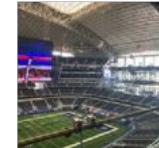
IDENTIFY CLOSEST ATTRACTIOnS

- Color and VGG-16 feature vectors
- Euclidean distance

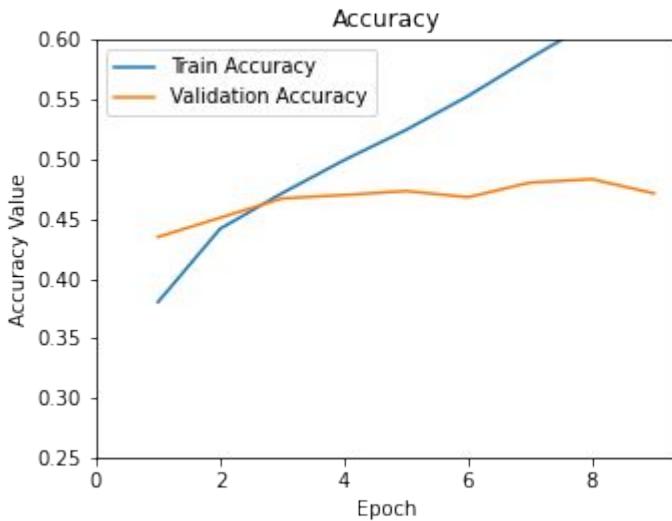
Attraction Labeling with NLP

Parks						
Beaches/Ocean						
Gardens/Zoos						
Art						

Attraction Labeling with NLP

Sports						
Entertainment						
Museums						
Landmarks						

Initial NN Results



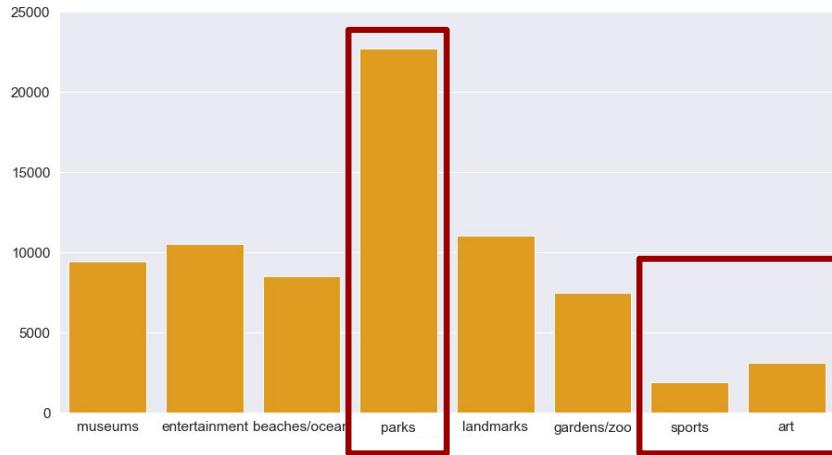
Overfitting!

Confusion matrix

	art	beaches/ocean	entertainment	gardens/zoo	landmarks	museums	parks	sports	
art	143	8	57	47	65	135	102	0	
beaches/ocean	2	656	108	39	83	90	610	12	
entertainment	23	69	567	119	110	421	417	19	
gardens/zoo	8	28	111	423	74	132	422	5	
landmarks	22	81	166	115	594	321	446	14	
museums	33	33	170	72	152	766	274	9	
parks	9	150	175	191	118	226	2400	10	
sports	2	12	50	27	11	74	65	107	

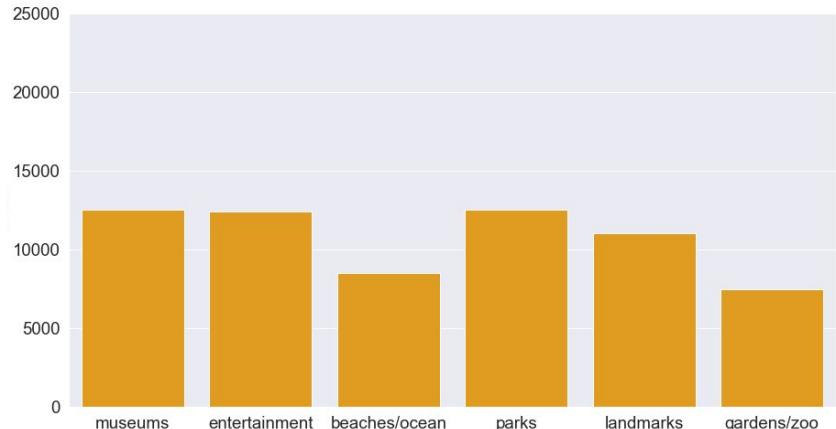
Class Imbalance!

Handling Class Imbalance



Random Undersampling

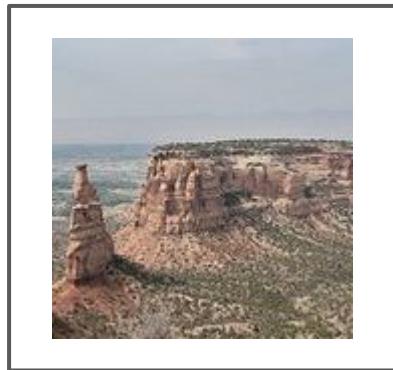
- 10,000 images removed from Parks class



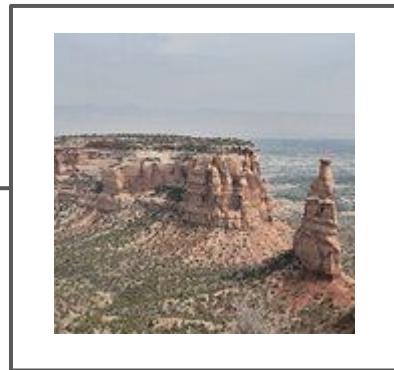
Classes Combined

- Sports → Entertainment
- Art → Museums

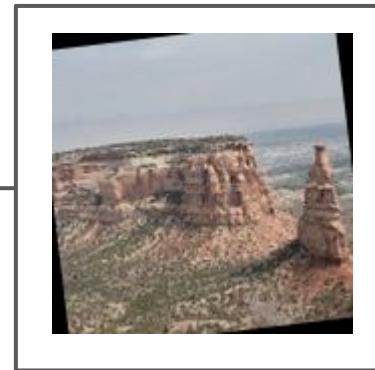
IMAGE AUGMENTATION



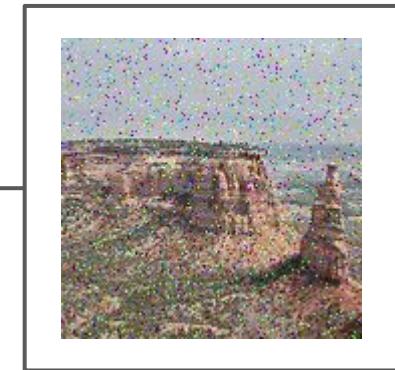
Original



Horizontal
Flip



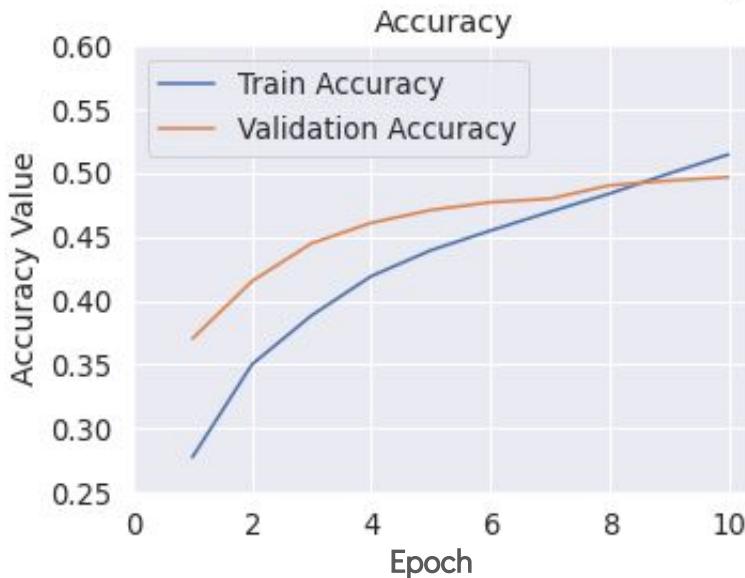
Random
Transform



Noise

Increase in training dataset: 38,000 → 150,000 images

Final NN Results



Confusion matrix

	beaches/ocean	entertainment	gardens/zoo	landmarks	museums	parks
beaches/ocean	993	165	40	151	71	312
entertainment	198	1194	123	374	381	179
gardens/zoo	45	229	585	183	136	339
landmarks	190	322	99	1040	317	214
museums	125	474	150	428	1145	151
parks	323	240	174	194	160	1446

beaches/ocean

entertainment

gardens/zoo

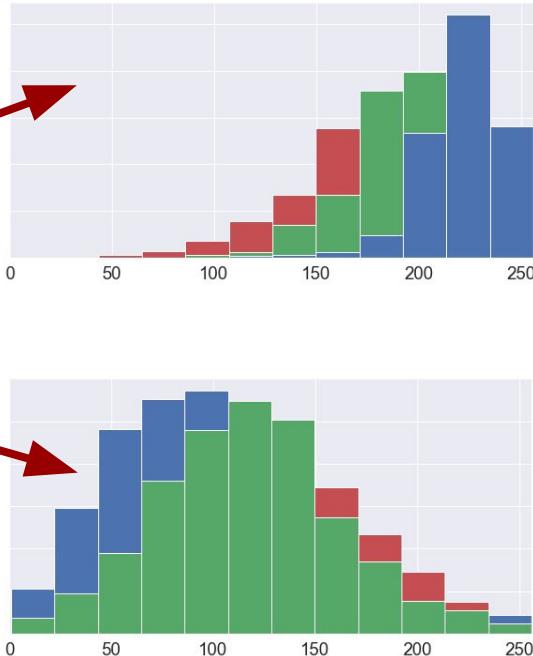
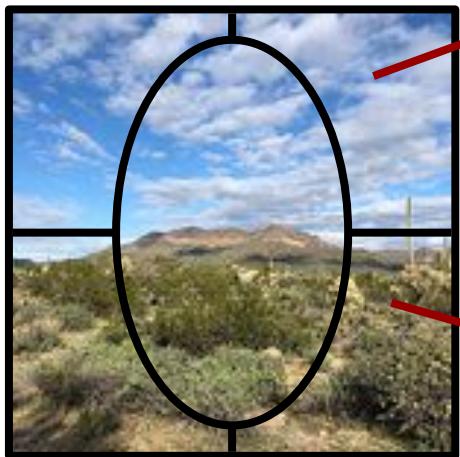
landmarks

museums

parks

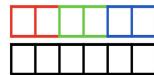
Predicted

Color Distribution Vector

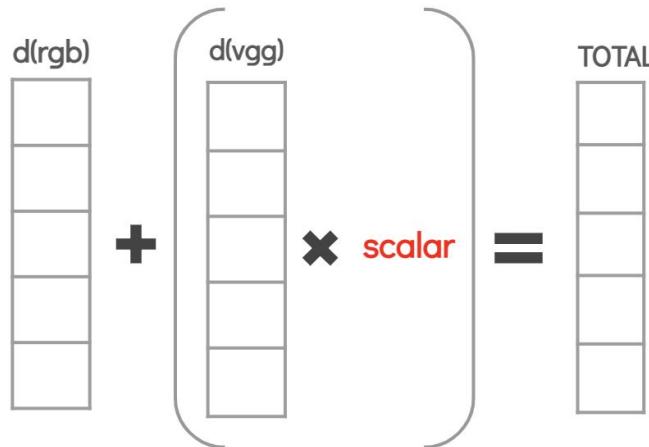


- RGB color measured in intensity from 0 to 256
- Images split into 5 sections
- For each section, 3 color distributions (OpenCV histogram) are added to a feature vector

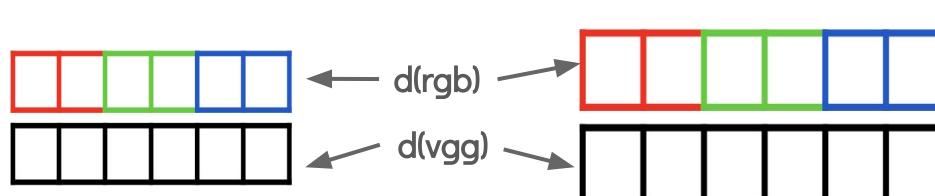
Dataset



Finding Closest Attractions



Cosine distance (d) between vectors



VGG-16 scalar dependent on class. Higher scalar for classes:

- Museums
- Landmarks
- Entertainment

Input Image



With so many vacation destinations and sights to see, planning your next vacation can be overwhelming and stressful.

Get some vacation planning help with this application, created using a neural network with transfer learning trained on over 70,000 tourist volunteer images scraped from TripAdvisor.

Recommending

A garden/green attraction

Anderson Japanese Gardens



Red Butte Garden



Portland Japanese Garden

The background of the slide is a photograph of a lush, colorful garden. In the foreground, there are large, well-maintained beds of tulips in shades of red, yellow, white, and blue. A wooden roller coaster track with red supports and blue cars is positioned in the center, partially obscured by the flower beds. The background features more greenery, trees, and a paved path where a few people are walking.

THANK YOU

Do you have any questions?

racheldilley11@gmail.com

www.linkedin.com/in/r-dilley/