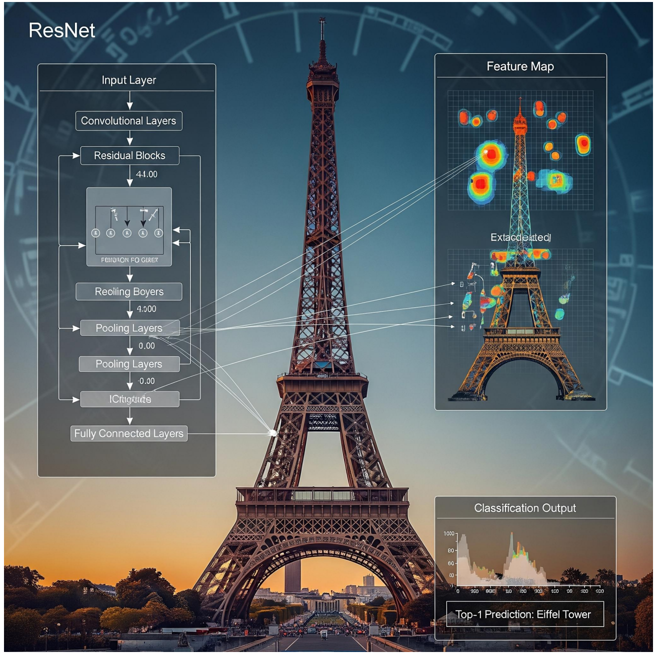
Geographic Landmark-Based Visual Localization

Brandon Tzou, Bosen Chia,Hildah Ngondoki, Rick Pereira
W281-Computer Vision Project

Landmark Recognition and Goals

- Automated landmark recognition predicts labels from image properties
- Useful for photo organization, image search, and geospatial intelligence
- Developed a classifier using the top 8 Google Landmark Recognition categories
- Explored HOG, Color Histograms, ViT, and ResNET feature embeddings
- Classified with Logistic Regression, KNN, SVM, and Resnet CNN
- Primary Metrics - Accuracy and Precision



EDA

- Original GLD v2 contains more than 5M images and 200k labels
- Examine top 30 most frequent categories
- Manually select 8 categories for classification

Figure 2.1. Example category that is not a landmark

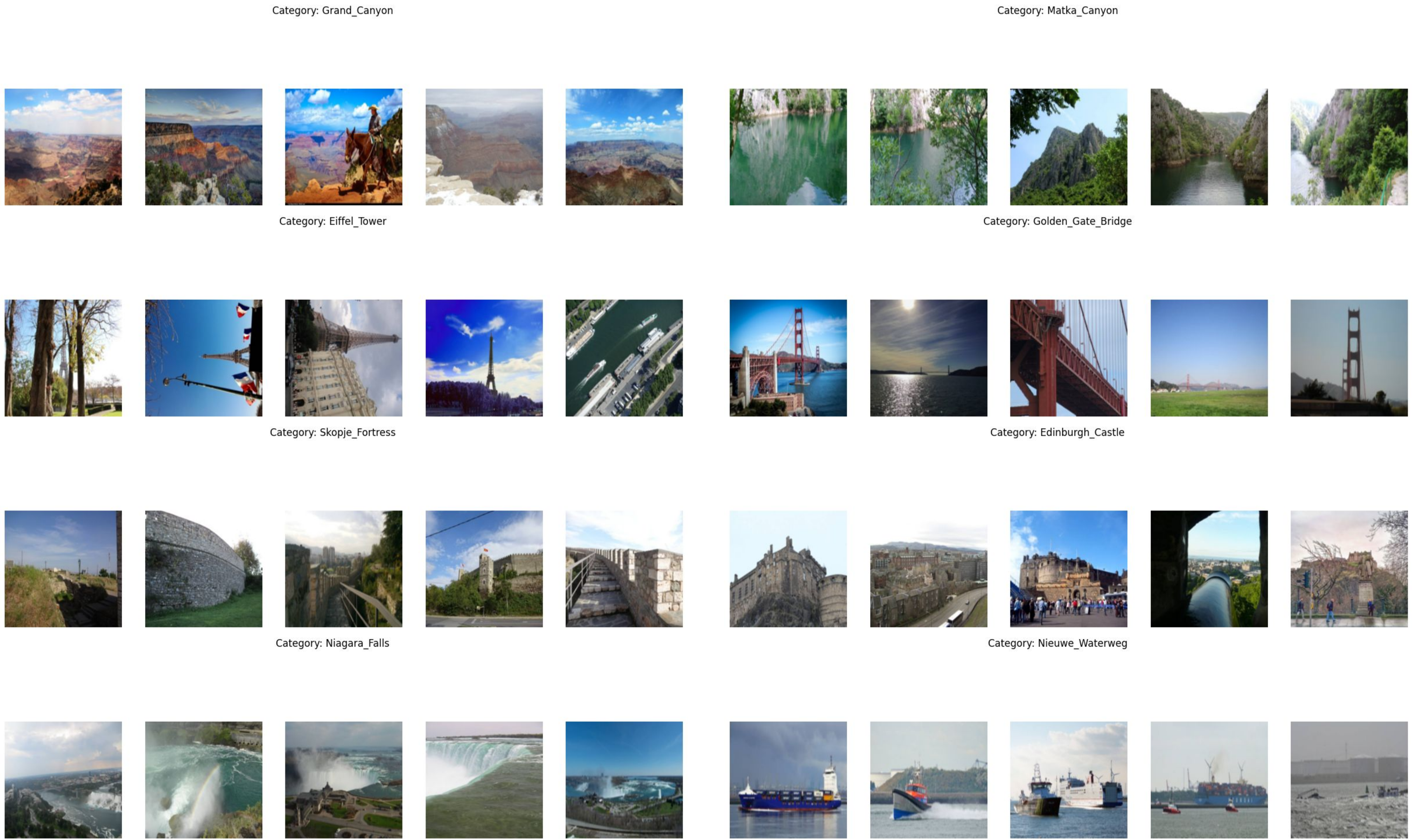
Category: Corktown,_Toronto (Label: 22)

Figure 2.2. Example images of a city

Category: Golden_Gate_Bridge (Label: 19)

Figure 2.3. Example of ideal category

UC Berkeley


Category: Grand_Canyon
Category: Matka_Canyon
Category: Eiffel_Tower
Category: Golden_Gate_Bridge
Category: Skopje_Fortress
Category: Edinburgh_Castle
Category: Niagara_Falls
Category: Nieuwe_Waterweg

Methodology

UC Berkeley

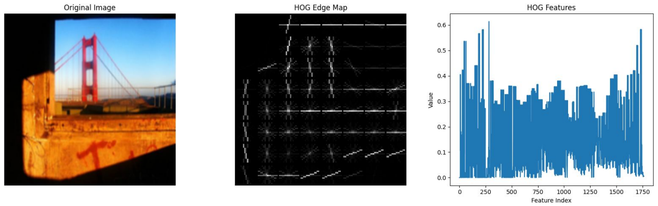
Feature Extraction

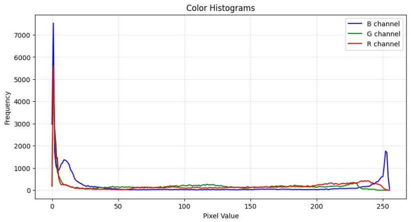
1. Simple Features

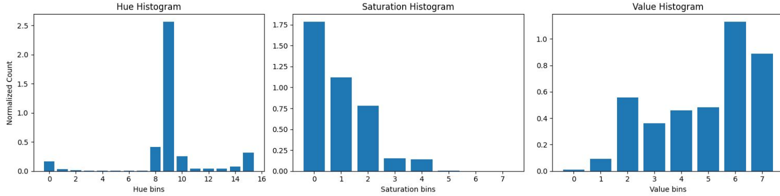
HOG

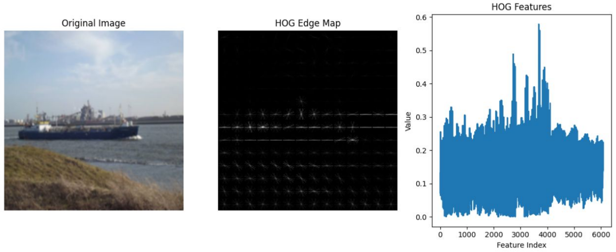
Color(RGB)

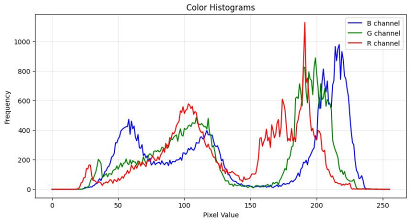
HSV

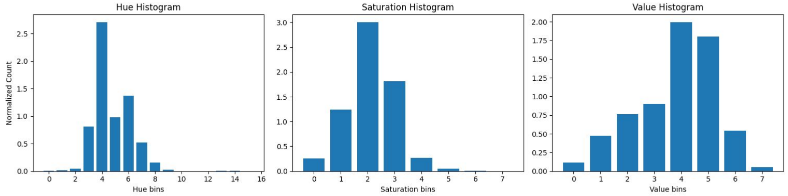










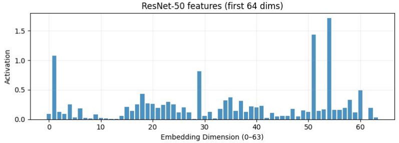


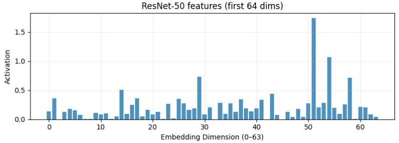
Representation of features between images with distinct characteristics.
HOG descriptors are useful for mages where color information is less distinctive. Color Histograms were able to capture the chromatic characteristics, while HSV provided more color separability between landmarks that share similar characteristics

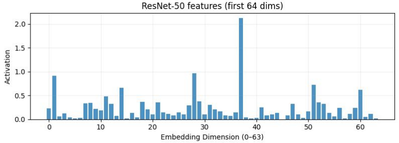
UC Berkeley

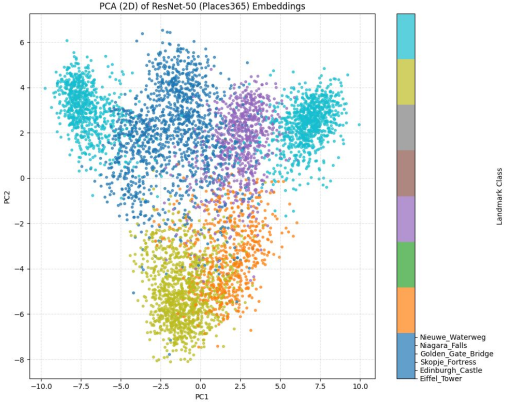
Feature Extraction

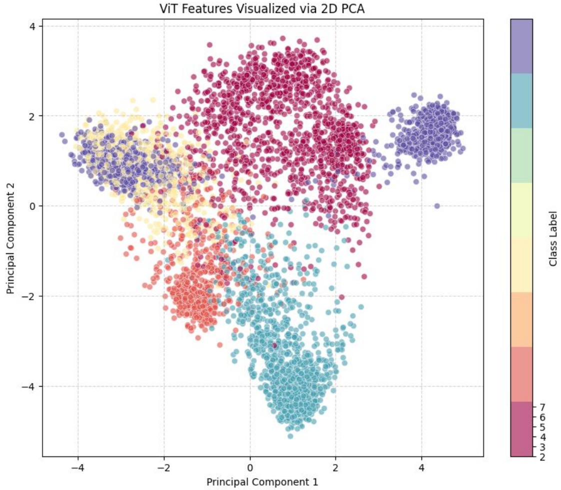
2. Complex Features











Resnet Embeddings : learned strong, distinct representations for each class when compressed down to 2D using PCA however, overlaps still exist.

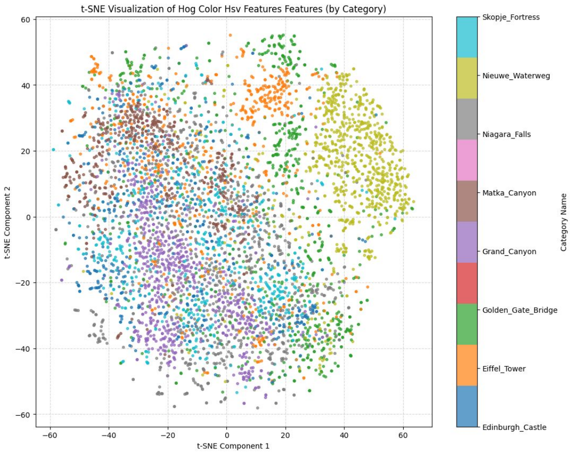
ViT Embeddings: The model captures meaningful structure however model's learned features aren't perfectly separable using just two PCA components

UC Berkeley

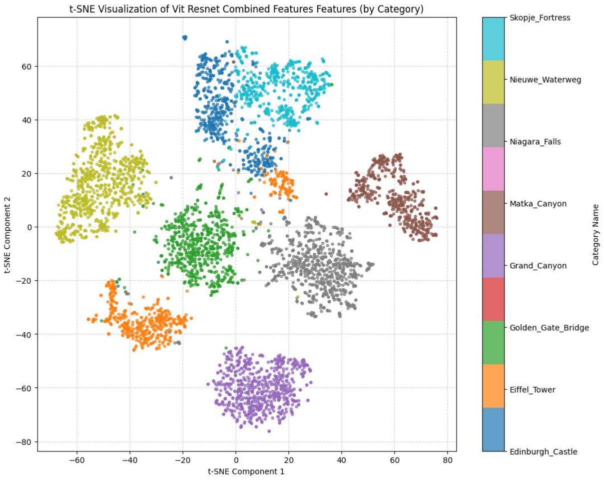
Feature Extraction

3. Combined Features

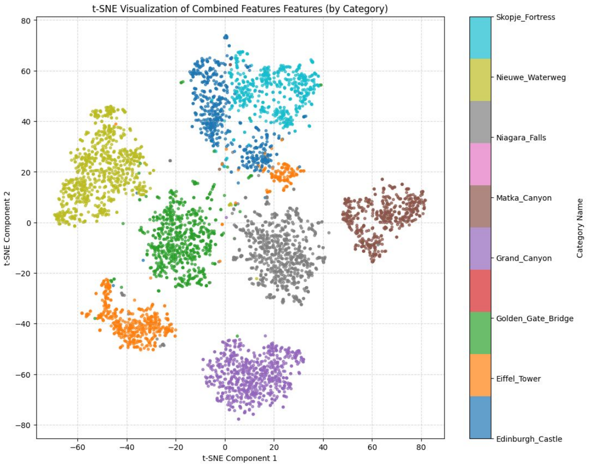
t-SNE Representations



Combined simple features



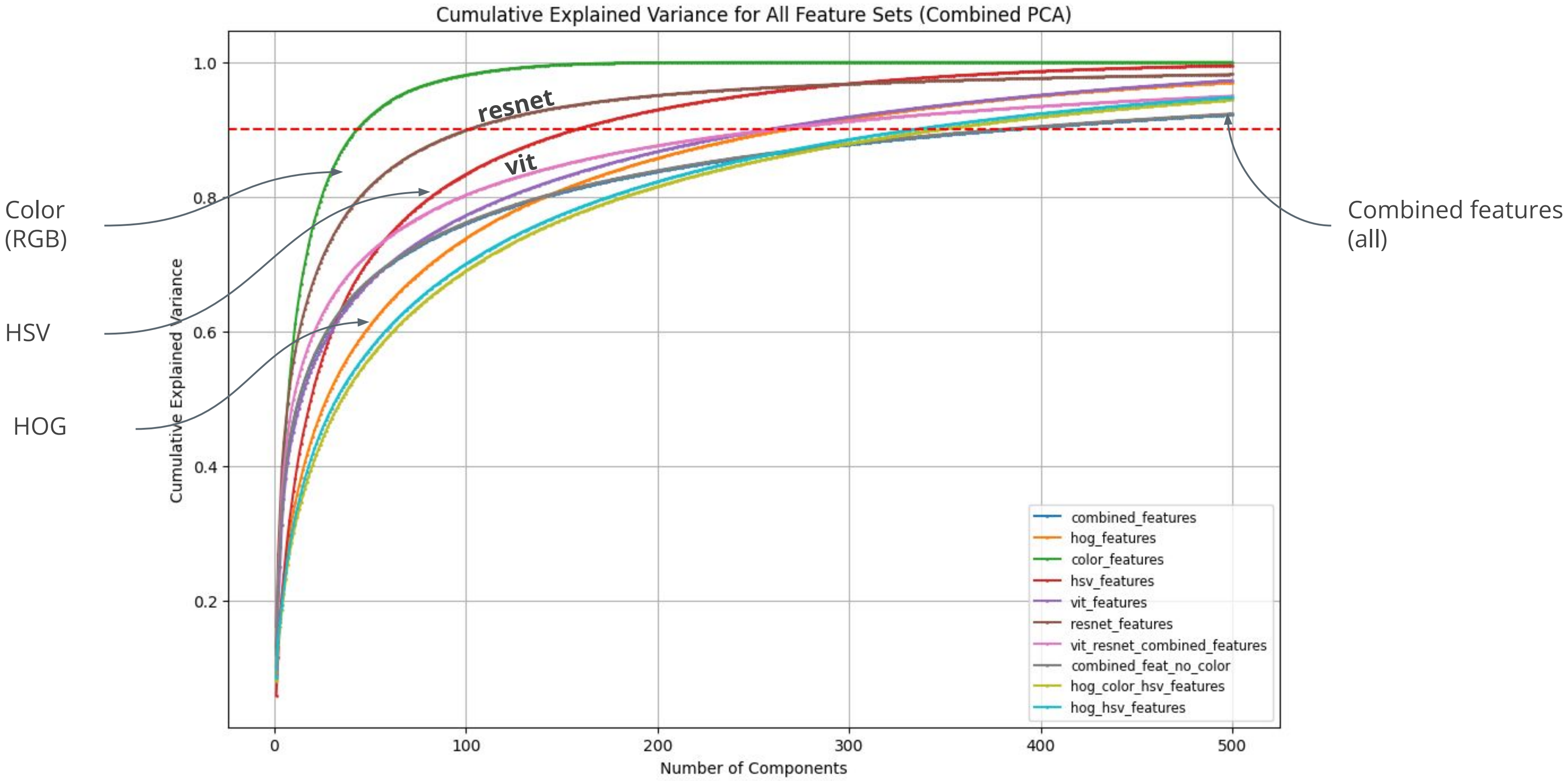
Combined complex features



Combined simple + complex features

UC Berkeley

PCA



Cumulative Explained Variance for All Feature Sets (Combined PCA)
Cumulative Explained Variance
Number of Components
1.0
0.8
0.6
0.4
0.2
0
100
200
300
400
500
resnet
vit
Color
(RGB)
HSV
HOG
Combined features
(all)
combined_features
hog_features
color_features
hsv_features
vit_features
resnet_features
vit_resnet_combined_features
combined_feat_no_color
hog_color_hsv_features
hog_hsv_features

Results

UC Berkeley

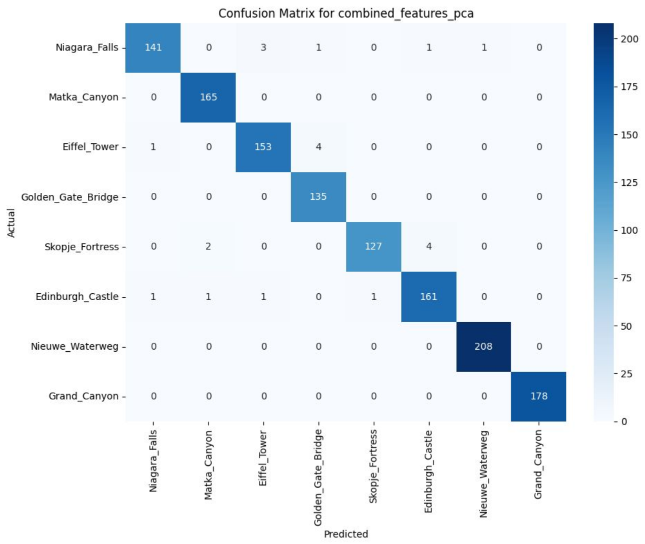
Logistic Regression

	Simple Features			Combined Simple Features		Complex Features		Combined Complex Features	Combined features (Simple & Complex)	
Train / Test Results	hog_features_pca	color_features_pca	hsv_features_pca	hog_color_hsv_features_pca	hog_hsv_features_pca	vit_features_pca	resnet_features_pca	vit_resnet_combined_features_pca	combined_features_pca	combined_feat_no_color_pca
Train	64.70%	57.12%	61.52%	75.66%	73.78%	97.66%	98.03%	98.97%	98.88%	99.16%
Test	61.99%	56.01%	58.81%	72.07%	69.82%	97.75%	96.51%	98.14%	98.37%	98.37%
Train Test Gap	2.71%	1.11%	2.71%	3.59%	3.96%	-0.09%	1.52%	0.83%	0.51%	0.79%

Class	Precision	Recall	F1-Score	Support
Niagara_Falls	0.975	0.968	0.971	158
Matka_Canyon	0.964	1	0.982	135
Eiffel_Tower	0.986	0.959	0.972	147
Golden_Gate_Bridge	0.982	1	0.991	165
Skopje_Fortress	0.992	0.955	0.973	133
Edinburgh_Castle	0.97	0.976	0.973	165
Nieuwe_Waterweg	0.995	1	0.998	208
Grand_Canyon	1	1	1	178
Accuracy			0.984	1289
Macro Avg	0.983	0.982	0.983	1289
Weighted Avg	0.984	0.984	0.984	1289

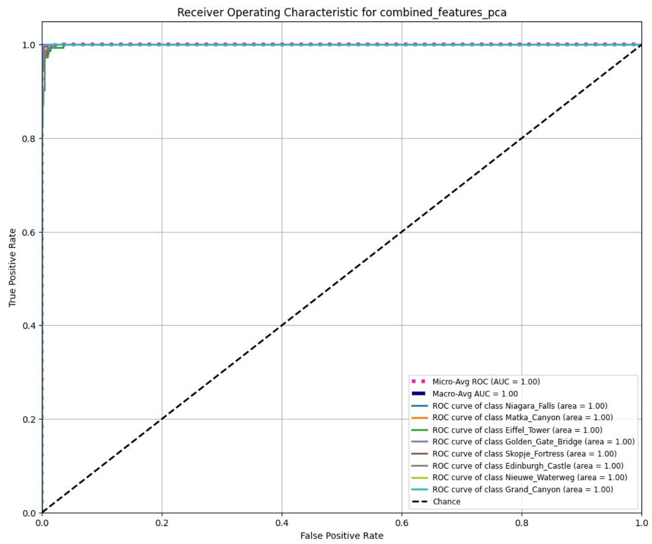
UC Berkeley

Logistic Regression



Actual \ Predicted	Niagara_Falls	Matka_Canyon	Eiffel_Tower	Golden_Gate_Bridge	Skopje_Fortress	Edinburgh_Castle	Nieuwe_Waterweg	Grand_Canyon
Niagara_Falls	141	0	3	1	0	1	1	0
Matka_Canyon	0	165	0	0	0	0	0	0
Eiffel_Tower	1	0	153	4	0	0	0	0
Golden_Gate_Bridge	0	0	0	135	0	0	0	0
Skopje_Fortress	0	2	0	0	127	4	0	0
Edinburgh_Castle	1	1	1	0	1	161	0	0
Nieuwe_Waterweg	0	0	0	0	0	0	208	0
Grand_Canyon	0	0	0	0	0	0	0	178

Confusion matrix for combined Resnet & ViT features



ROC curves for combined Resnet & ViT features

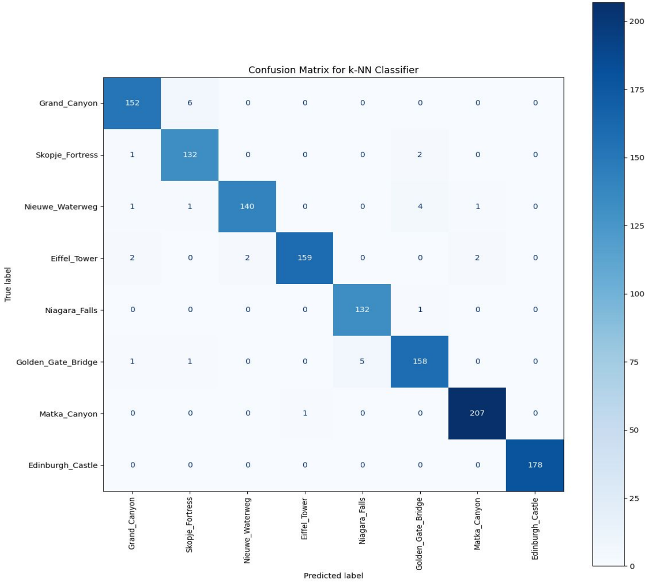
UC Berkeley

K Nearest Neighbor (KNN)

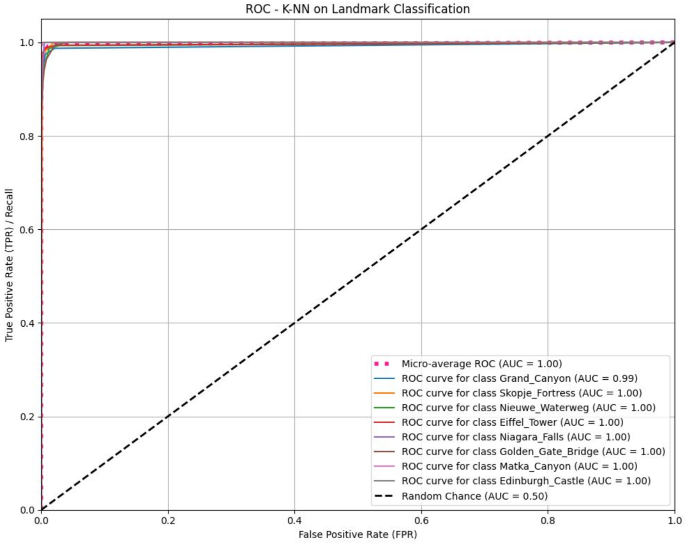
Sum of accuracy	*Features*									
Train / Test Results	color_features_pca	combined_feat_no_color_pca	combined_features_pca	hog_color_hsv_features_pca	hog_features_pca	hog_hsv_features_pca	hsv_features_pca	resnet_features_pca	vit_features_pca	vit_resnet_combined_features_pca
Train	0.6979	0.9848	0.9846	0.7671	0.7198	0.75	0.6593	0.9788	0.9766	**0.9858**
Test	0.6979	0.9848	0.9846	0.7671	0.7198	0.75	0.6593	0.9766	0.9766	**0.9858**
Train Test Gap	0	0	0	0	0	0	0	-0.0022	0	0

Class Label	Precision	Recall	F1-Score	Support
0 Grand_Canyon	0.99	0.99	0.99	786
1 Matka_Canyon	1.00	1.00	1.00	528
2 Eiffel_Tower	0.98	0.98	0.98	584
3 Edinburgh_Castle	1.00	1.00	0.98	723
4 Skopje_Fortress	0.96	0.96	0.98	571
5 Golden_Gate_Bridge	0.97	0.97	0.97	569
6 Niagara_Falls	0.99	0.99	0.99	865
7 Nieuwe_Waterweg	1.00	1.00	1.00	710

K Nearest Neighbor (KNN)



Confusion matrix for combined Resnet & ViT features



ROC curves for combined Resnet & ViT features

UC Berkeley

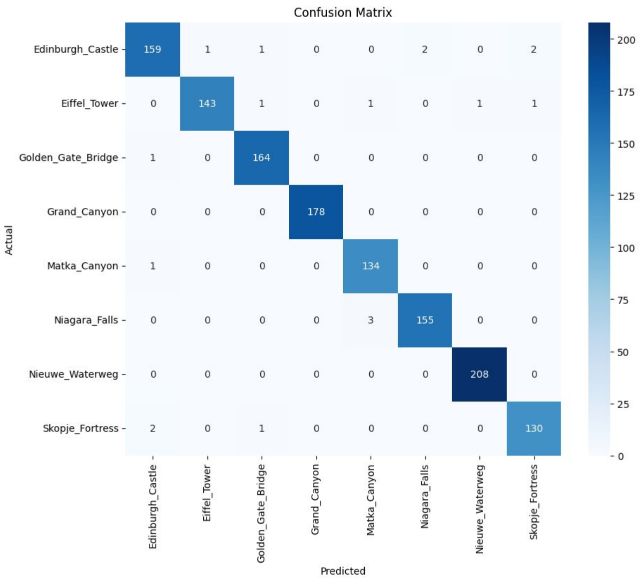
Support Vector Machines (SVM)

Features Used	Accuracy
combined_feat_no_color	0.986036
combined	0.986036
vit_resnet_combined	**0.986036**
vit	0.985260
resnet	0.979829
hog_color_hsv	0.795966
hog_hsv	0.775795
hog	0.716059
hsv	0.658650
color	0.626843

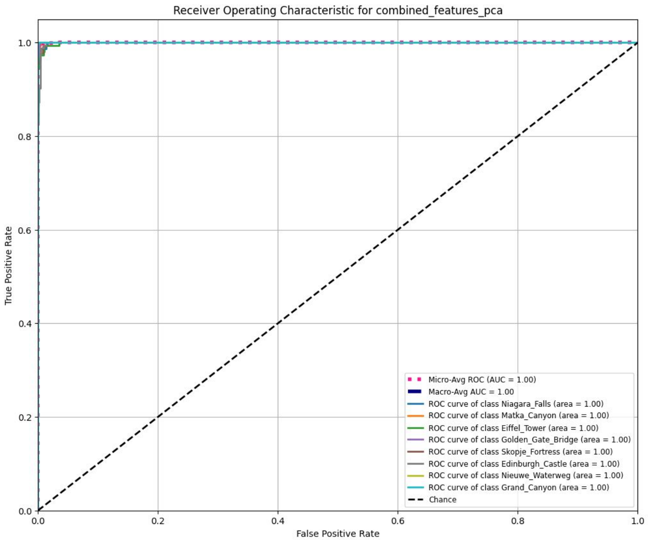
Class Label	Precision	Recall	F1-Score	Support
0 Grand_Canyon	1.00	1.00	1.00	178
1 Matka_Canyon	0.97	0.99	0.98	135
2 Eiffel_Tower	0.99	0.98	0.99	147
3 Edinburgh_Castle	0.98	0.96	0.97	165
4 Skopje_Fortress	0.98	0.97	0.97	133
5 Golden_Gate_Bridge	0.99	0.99	0.99	165
6 Niagara_Falls	0.97	0.98	0.98	158
7 Nieuwe_Waterweg	1.00	1.00	1.00	208

UC Berkeley

Support Vector Machines (SVM)



Confusion matrix for combined Resnet & ViT features



ROC curves for combined Resnet & ViT features

UC Berkeley

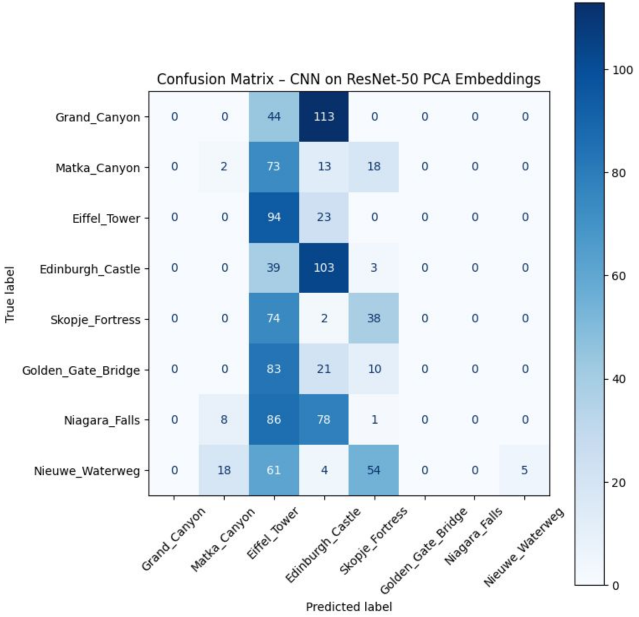
Convolution Neural Network (CNN)

Experiment	Learning rate	Epochs	Batch size	Dropout	Training accuracy	Training time (s)
1	1e-3	50	Mini-batch	0.3	0.2234	67.2
2	1e-3	20	Mini-batch	0.3	0.2195	28.8
3	1e-2	20	Full	0.3	0.1924	5.4
4	5e-4	20	Full	0.3	0.2078	5.4
5	1e-3	40	Full	0.3	0.2141	10.2
6	1e-3	20	Full	0.3	**0.2266**	5.4

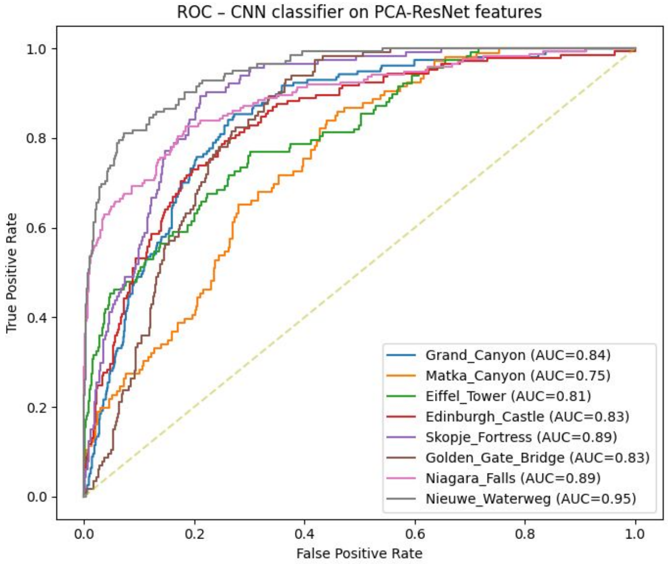
Label	precision	recall	f1-score	support
0 Grand_Canyon	0.00	0.00	0.00	157
1 Matka_Canyon	0.07	0.02	0.03	106
2 Eiffel_Tower	0.17	0.80	0.28	117
3 Edinburgh_Castle	0.29	0.71	0.41	145
4 Skopje_Fortress	0.31	0.33	0.32	114
5 Golden_Gate_Bridge	0.00	0.00	0.00	114
6 Niagara_Falls	0.00	0.00	0.00	173
7 Nieuwe_Waterweg	1.00	0.04	0.07	142

UC Berkeley

Convolution Neural Network (CNN)



Confusion matrix for Resnet features



ROC curves for Resnet features

UC Berkeley

Discussion

UC Berkeley

Model Generalizability and Features

- Top classifiers demonstrated strong generalization due to deep features.
- Small gap between training and testing accuracies confirms robustness.
- Combined deep features captured transferable, scene-centric patterns well.
- Handcrafted features like HOG caused overfitting and reduced generalization.
- Advanced features were essential for achieving high generalization accuracy.

UC Berkeley

Accuracy vs Efficiency Tradeoffs

Model Comparison	Optimal Deployment Strategy
Accuracy-Optimized (SVM)	The Best Trade-off: K-Nearest Neighbors (KNN)
Accuracy: 98.60% (Highest overall)	**Accuracy**: 98.58% (Near-maximal)
Inference Time: 0.77 seconds	**Inference Time**: 0.3 seconds (2.5x faster)
Cost: Moderate computational overhead.	**Conclusion**: KNN efficiently leverages high-quality deep-learning features to deliver robust, near-perfect accuracy with minimal delay.

UC Berkeley

Looking Ahead

Conclusion

- Feature representation is the dominant factor in model performance
- ViT & ResNet embeddings enable near-perfect generalization
- Handcrafted features (HOG, HSV, RGB) underperform
- Classifier–feature compatibility is critical
- KNN offers the best accuracy–efficiency trade-off for deployment

UC Berkeley

Future Work

- Scale from 8 classes to full dataset
- Address large-scale challenges: Class imbalance, Visually similar landmarks, Computational scalability
- Improve robustness to real-world conditions
- Fine-tune pretrained models
- Optimize inference for large-scale deployment

UC Berkeley

END

Classification Approach

Models

- Logistic Regression,
- K Nearest Neighbour (KNN),
- Support Vector Machines(SVM) and
- a Resnet Convolution Neural Network (CNN)

Metrics

- Accuracy
- Precision
- F1-score
- ROC Curve
- Training/inference time

UC Berkeley