Recognition of Tourist Attractions

Yuanfang Li (yli03@stanford.edu), Xin Li(xinli16@stanford.edu), Chang Yue(changyue@stanford.edu)

Motivation

The ability to recognize landmarks from images can be extremely useful both when choosing a travel destination and when trying to identify locations in a foreign place. Our project aims to be able to recognize ten famous tourist attractions in Beijing by applying transfer learning to a pretrained convolutional neural network, PlacesCNN.

Data and Features

The images in the dataset were obtained through Google image search, augmented with horizontal reflection and vertical crop. Two sets of features were extracted for each image:

1. 4096 features extracted from last fully-connect layer of PlacesCNN.
2. 64 features representing the pixel ratio.

Methodology

We trained separate classifiers for day and night, as the features extracted for day and night images by PlacesCNN are different. Some attractions also look similar at night, which we anticipated would lower our accuracy.

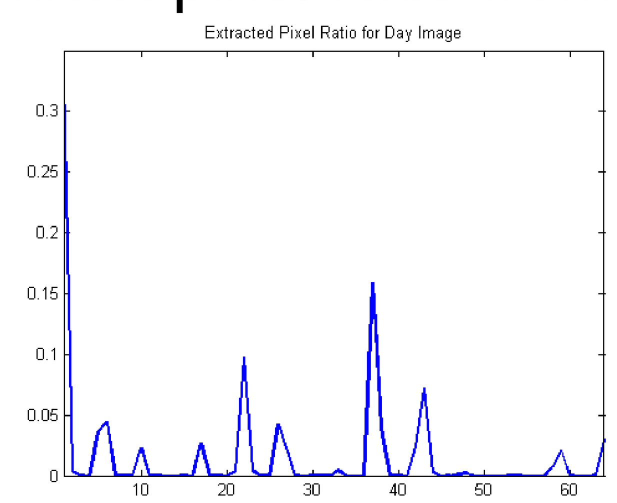
Results

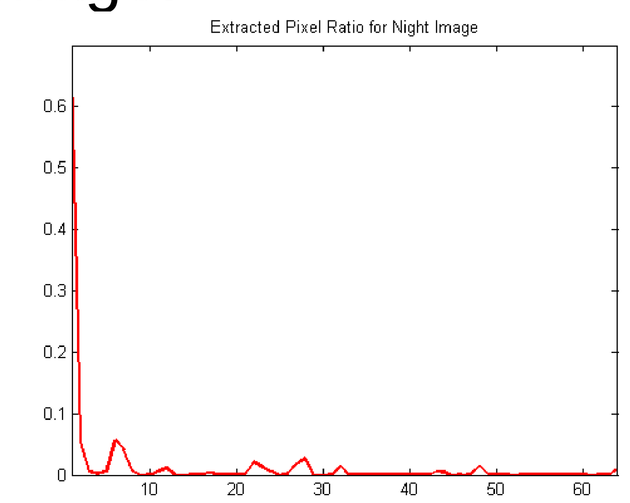
	Training data size	Training accuracy	Validation data size	Validation accuracy
Day vs. night	2800	96%	1200	94%
Day SVM	5600	93%	1400	83.5%
Night SVM	1400	96.81%	600	87.13%
Day Softmax	5600	52.34%	1400	46.6%
Night Softmax	1400	82.17%	600	75.17%

Test data size	Test accuracy (using SVM)
1000	**81.17%**

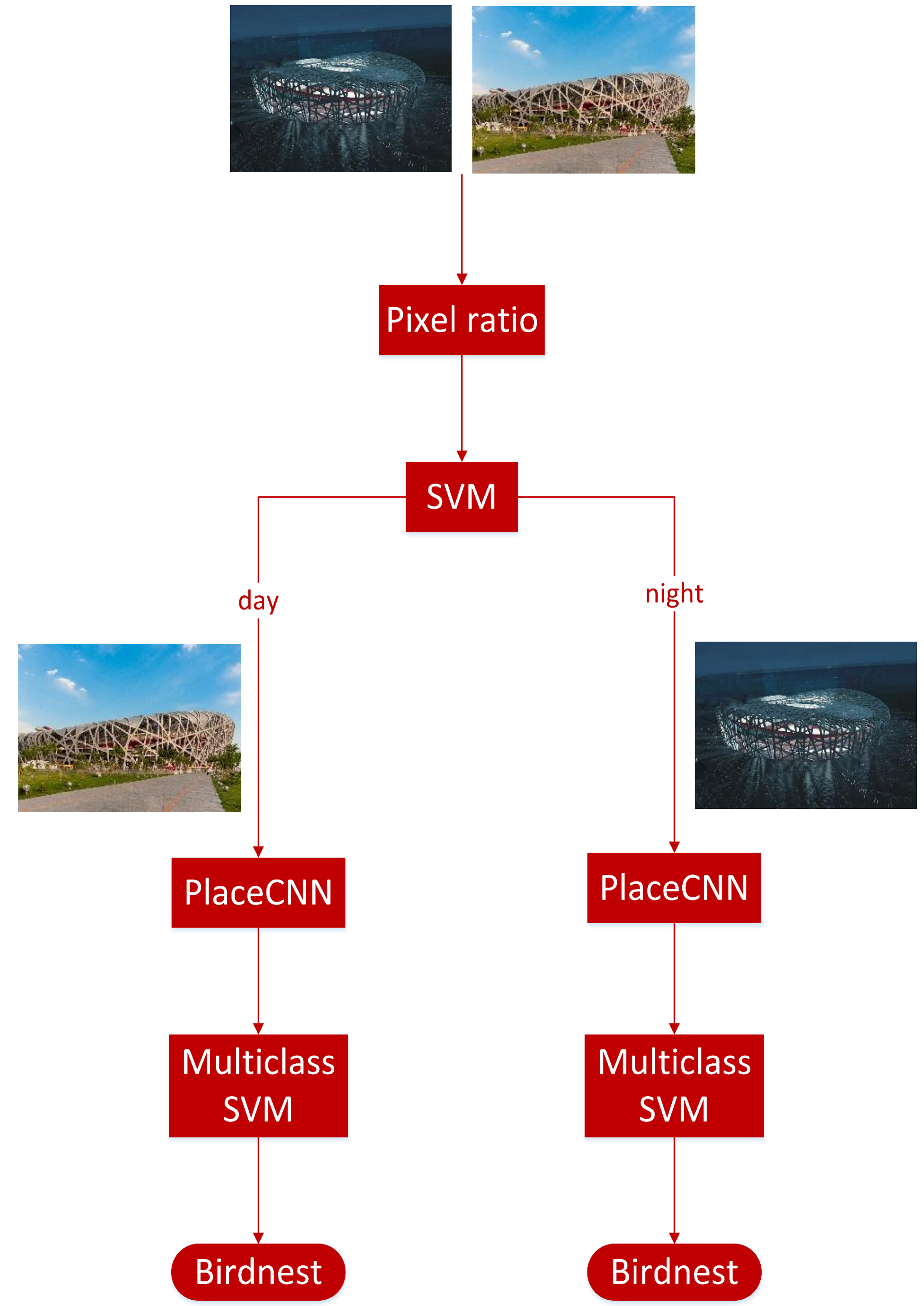
Classification Pipeline

1. Extract pixel ratio from image.



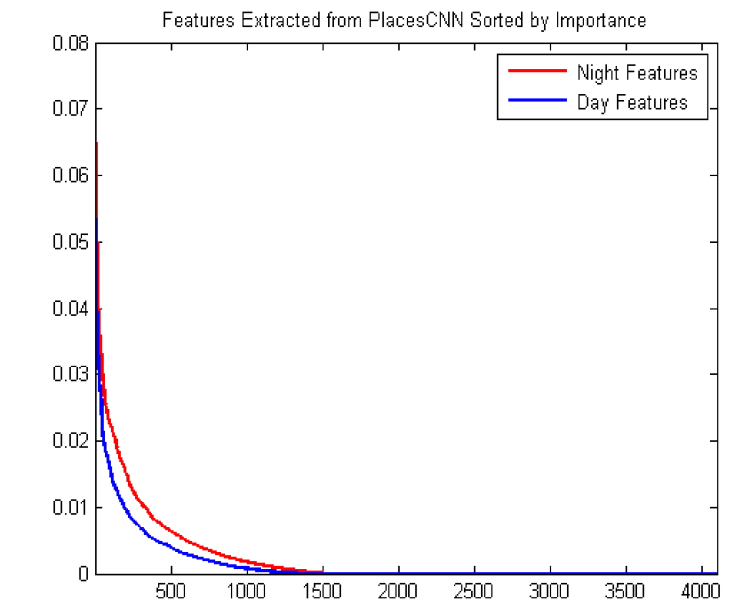


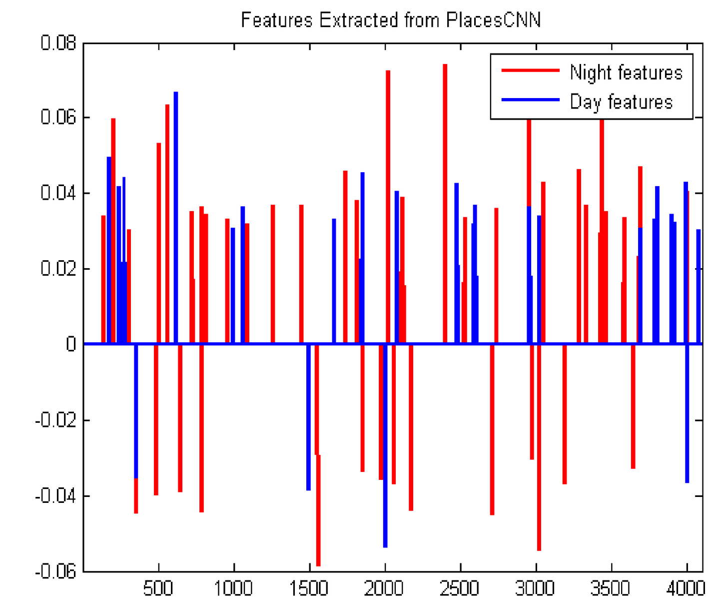
2. Classify image as either day or night using SVM with linear kernels.
3. Extract image features using PlacesCNN.
4. Classify image using either day- or night-trained multiclass SVM with linear kernels.



Analysis

- We expected a higher test accuracy for the final classifier since PlacesCNN was trained for scene recognition, which is very similar to our application. This may be because the majority of classes in PlacesCNN are natural whereas the attractions we chose were manmade. For the same reason, softmax gives lower accuracy than SVM classifier.
- Higher accuracy for night vs. day images is likely because noise from surroundings is not visible at night. Comparison of the feature vectors also shows that magnitude of night feature vectors is higher.





Future

Based on the results from Zhou et al[1], networks trained on different datasets will differ more in deeper layers. Ideally we would extract our features from an earlier convolution layer and train the rest of the network from scratch using our dataset.

References

[1] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in Advances in neural information processing systems, 2014, pp. 487-495.
[2] M. C. Science and A. I. Laboratory. Places CNN. [Online]. Available: http://places.csail.mit.edu/downloadCNN.html
[3] Giuseppe Cardone. "Day and night: an image classifier with scikit-learn". [Online]. Available: http://www.ippatsuman.com/