Discovering Scenic Roads Using Neural Networks

Bogdan State

Abstract

Convolutional Neural Networks (CNNs) coupled with novel large image datasets offer ample hope for building a computational understanding of the relationship between humans and environment. CNNs are also particularly effective at learning complex sets of rules, such as the ones governing aesthetics. Here I present one potential application of computational aesthetics to the detection of scenic byways. To do so I use a dataset of Google Streetview images collected alongside California State Highways, with labels derived from the California Department of Transportation official "scenic byway" designation. A CNN is trained on a dataset of order-10^5 Streetview images, achieving moderate accuracy against a validation set. With suitable refinement the approach shows promise in quantifying the degree to which roadways are aesthetically pleasing, a measure with potential applications in tourism, transportation, and urban planning.

Motivation

- Aesthetics of locations important consideration for planning, land use.
- Internet data showing great potential for fine-grained understandings of human-environment interaction. [5, 4]
- Deep learning using Streetview imagery [1] and satellite data [3] = potential revolution in understanding the spatial dimension of social data.
- Computational aesthetics of natural landscapes under-researched, despite evidence of important psychological impacts.[8]

6, 2

Data

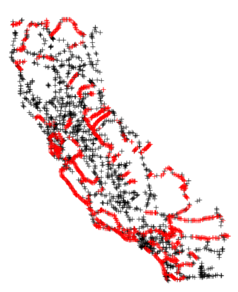

Figure: Scenic (red) and non-scenic (black) sample points.

- Labels: eligible scenic routes according to CA DoT vs. all state highways.
- Images: balanced sample of Google Streetview images from scenic, non-scenic roads.

Methods

Deep convolutional neural network to predict scenicness given image pixels:

- 3x3 conv. layer, depth 10, stride 1.
- Leaky ReLu activation layer (p=0.01).
- 2d batch normalization layer. [2]
- 2d max-pooling layer, kernel size 7 and stride 2.
- 7x7 conv. layer, depth 3, stride 1 and dilation=3.
- 2d dropout layer. [6]
- 2d max-pooling layer, kernel size 7, stride 2.
- Leaky ReLu activation layer (p=0.01).
- 2d max-pool layer, kernel size 7 and stride 2.
- linear layer w/ 128 output neurons.
- linear layer w/ 2 output neurons.

Training done in PyTorch on machine w/ 1080Ti GPU.

Data Augmentation

- Flickr YFCC100M [7] image features + approximate geotagging = average embedding of area near streetview image.
- Use most / least discriminative 30 dimensions as inputs to model predicting baseline probability

Figure: Scenic and non-Scenic Flickr 100m tags.

Results

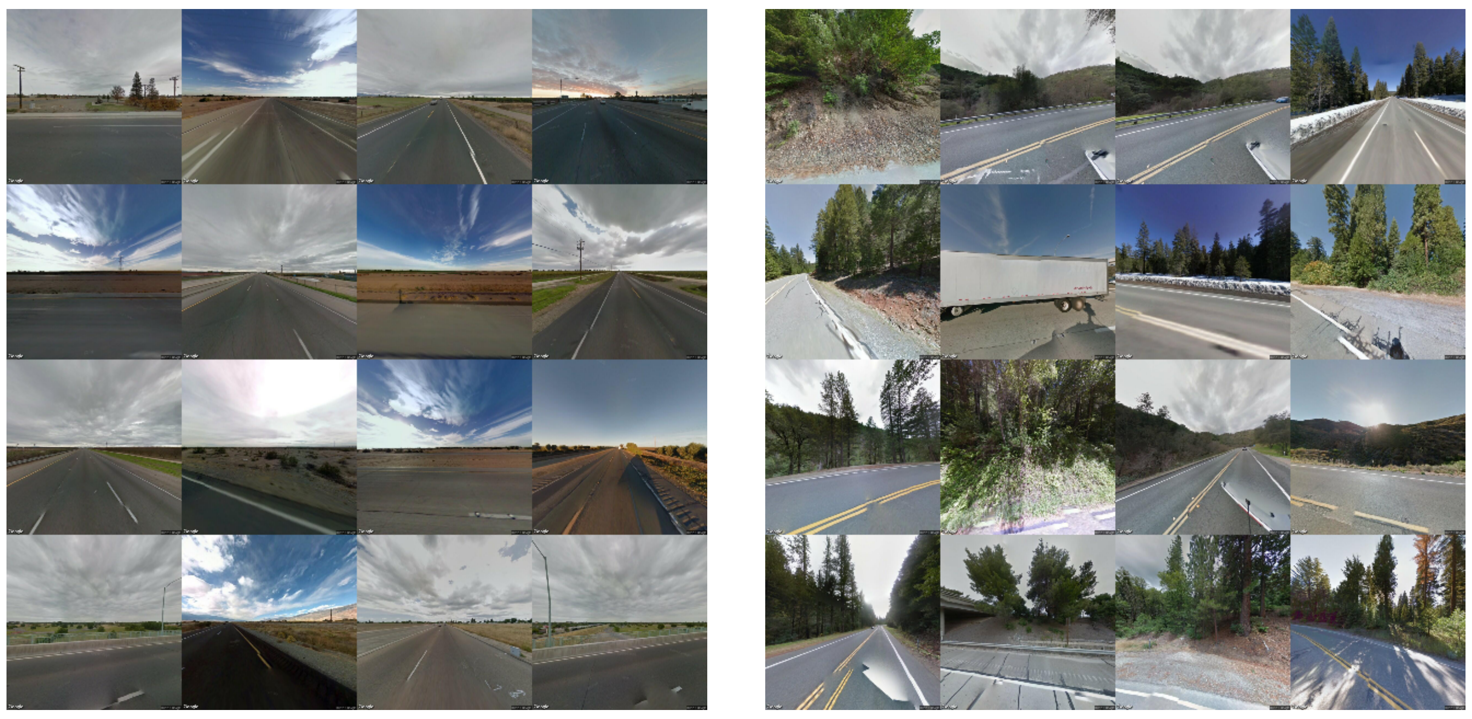

Figure: Least and Most Scenic Images

- Accuracy hovers at 60%.
- Problem: not all streetview pictures on scenic roads truly scenic.
- Out-of-sample predictions: county-maintained roads in Santa Cruz County.
- Reasonable results, but note issue with highway 1 (few trees, ocean scenic too).

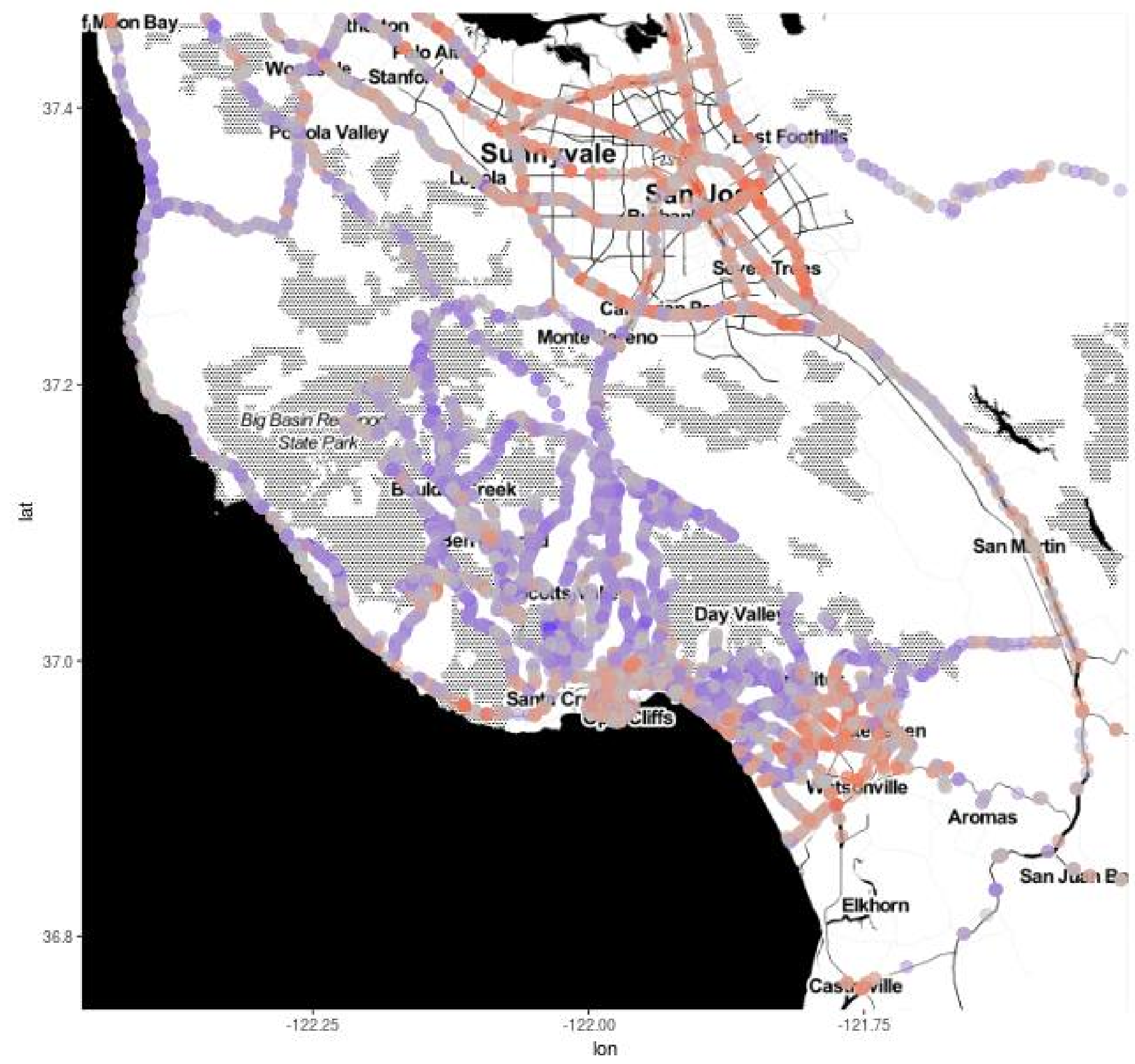

Figure: Predicted scenicness of roads in Santa Cruz County and environs.

Future Extensions

- add GIS features (proximity to water, altitude, etc.)
- Satellite features (e.g. LANDSAT data)
- out-of-sample predictions for entire U.S. and beyond.

Conclusion

- Streetview imagery can enhance computational understandings of landscape aesthetics.
- Potential to use expert-curated geographic datasets as sources of labels.
- Machine Learning + Sensors can help remove administrative biases in data collection (i.e., only state hwys eligible for scenic designation).

References

[1] Timnit Gebru, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, Erez Lieberman Aiden, and Li Fei-Fei. 2017. Using deep learning and google street view to estimate the demographic makeup of the us. *arXiv preprint arXiv:1702.06683* (2017).

[2] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167* (2015).

[3] Neal Jean, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. 2016. Combining satellite imagery and machine learning to predict poverty. *Science* 353, 6301 (2016), 790–794.

[4] Daniele Quercia, Neil Keith O'Hare, and Henriette Cramer. 2014. Aesthetic capital: what makes london look beautiful, quiet, and happy?. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 945–955.

[5] Daniele Quercia, Rossano Schifanella, and Luca Maria Aiello. 2014. The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In *Proceedings of the 25th ACM conference on Hypertext and social media*. ACM, 116–125.

[6] Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* 15, 1 (2014), 1929–1958.

[7] Bart Thomee, David A Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. 2016. Yfcc100m: The new data in multimedia research. *Commun. ACM* 59, 2 (2016), 64–73.

[8] Roger S Ulrich. 1986. Human responses to vegetation and landscapes. *Landscape and urban planning* 13 (1986), 29–44.