Car context heat map for fast exploration

Rogerio Bonatti (rbonatti) and Jaskaran Grover (jaskarag)

Abstract

- Exploration: find all cars in a terrain as fast as possible
- Problem: what to do when the object is not in sight?
- Trained car context classifier with 272 aerial images with hand-labeled cars
- Obtained 65.4% and 79.5% accuracy with Neural Net classifier using low (2^{nd} layer) and high level (6^{th} layer) features extracted from AlexNet

Motivation

Figure 1: Exploration of objects such as cars or animals in an open environment, or search-and-rescue missions

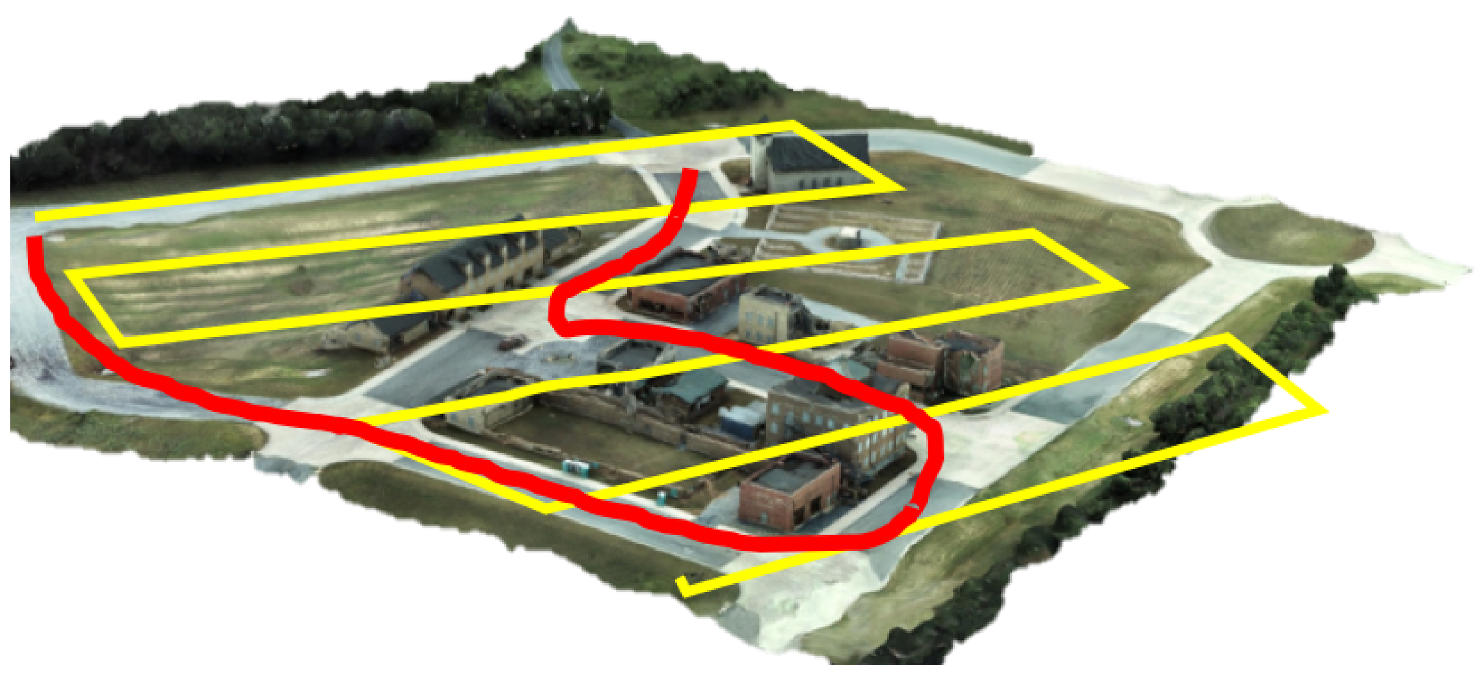

Figure 2: Focused exploration leads to more relevant information

Methods

We extracted features from regions close and far away from cars in 272 manually labeled examples.
We trained a binary classifier to distinguish between our two classes: positive for being close to a car and negative for being distant form a car. We implemented two approaches:

Step	**Description**
Region selection	Select patches in blobs close and far away to the labeled car
Feature extraction	**(1)** Extract features from the 2^{nd} layer of AlexNet — 100K samples with 96 features each **(2)** Extract features from the 6^{th} layer of AlexNet — 3K samples with 4096 features each
Training	Train a binary neural network classifier to output probability of patch being close to a car
Quantitative test	Evaluate performance of classifier on test set
Qualitative test	Break down an entire image in patches and evaluate qualitatively its heat map

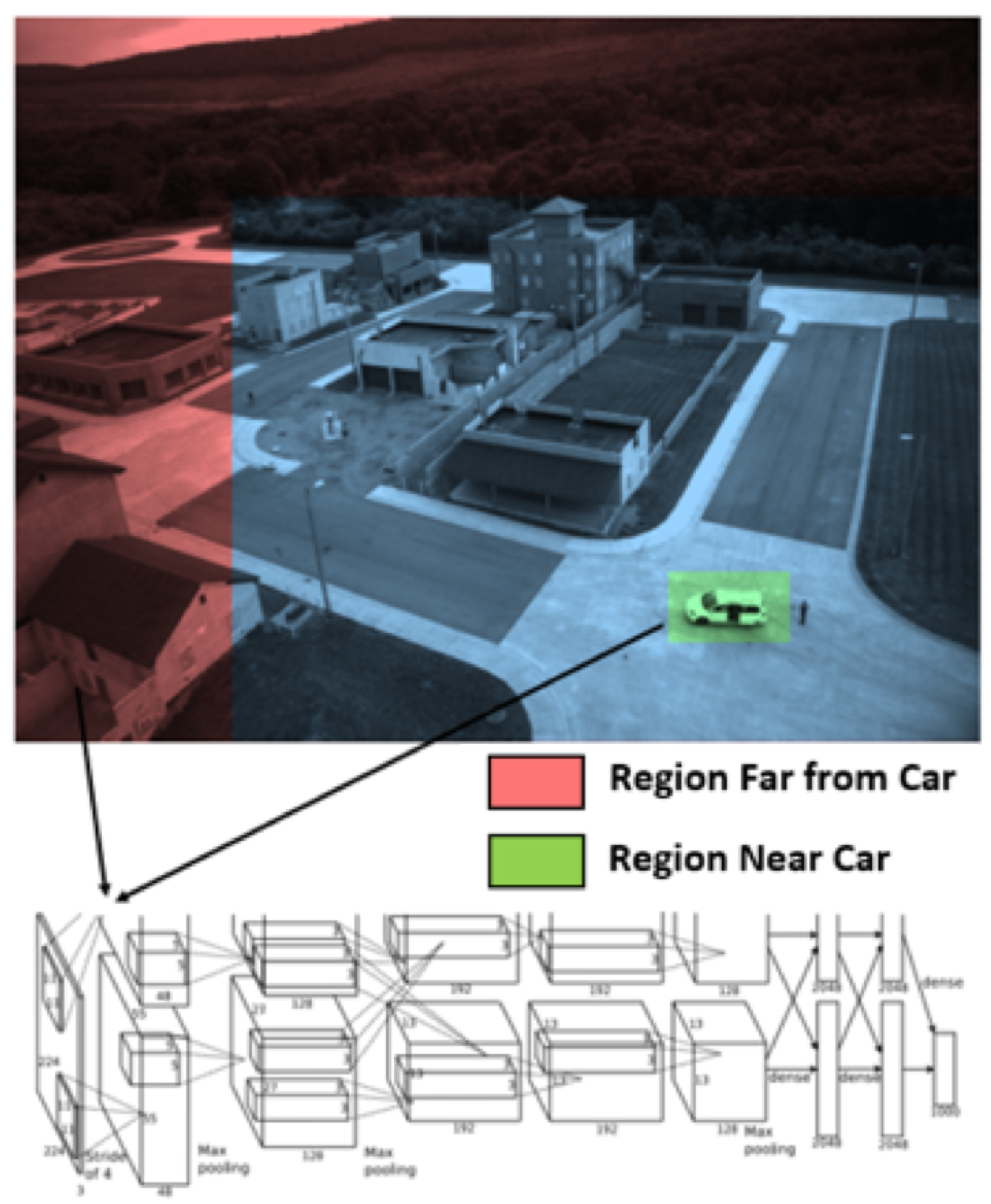

Figure 3: Feature extraction from candidate regions of an image

Heat maps of car context

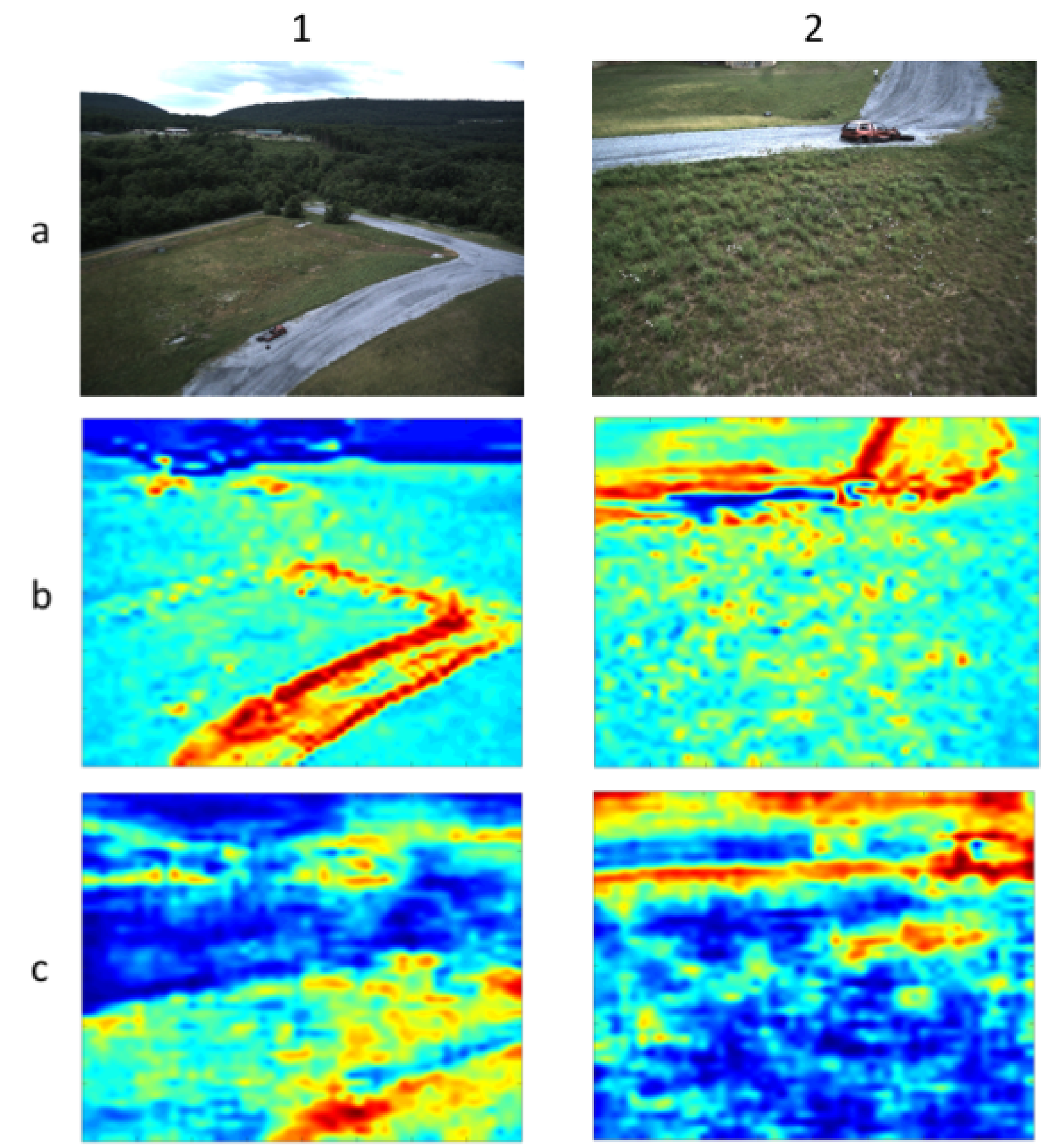

Figure 4: Heat maps generated for two sample images in column 1 and column 2. Row "a" contains the original image, row "b" has the heat map based on low-level features and row "c" has the heatmap based on high-level features from AlexNet

Qualitative evaluation of the images:

- Low-level features from AlexNet identify edges and texture well, as expected
- High-level features generate a more diffuse representation of context
- Images in extremes of scales are not well represented using either method

References

[1] Ben Athiwaratkun and Keegan Kang. Feature representation in convolutional neural networks. *arXiv preprint arXiv:1507.02313*, 2015.

[2] Joshua Gleason, Ara V Nefian, Xavier Bouyssounousse, Terry Fong, and George Bebis. Vehicle detection from aerial imagery. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 2065–2070. IEEE, 2011.

[3] Liang Zheng, Yali Zhao, Shengjin Wang, Jingdong Wang, and Qi Tian. Good practice in cnn feature transfer. *arXiv preprint arXiv:1604.00133*, 2016.

Future work

- Employ a combination of high and low level features to identify context (hypercolumns)
- Test different network architectures for both feature extraction and classification
- Use larger dataset and train a convolutional neural network for end-to-end classification

Acknowledgements

The authors would like to thank Daniel Maturana from CMU for providing the dataset of aerial images used in this project.