Tracking Natural Events through Social Media and Computer Vision

Jingya Wang Mohammed Korayem* Saúl Blanco David J. Crandall
School of Informatics and Computing
Indiana University
Bloomington, IN
{wang203,mkorayem,sblancor,djcran}@indiana.edu

ABSTRACT

Accurate, efficient, global observation of natural events is important for ecologists, meteorologists, governments, and the public. Satellites are effective but limited by their perspective and by atmospheric conditions. Public images on photo-sharing websites could provide crowd-sourced ground data to complement satellites, since photos contain evidence of the state of the natural world. In this work, we test the ability of computer vision to observe natural events in millions of geo-tagged Flickr photos, over nine years and an entire continent. We use satellites as (noisy) ground truth to train two types of classifiers, one that estimates if a Flickr photo has evidence of an event, and one that aggregates these estimates to produce an observation for given times and places. We present a web tool for visualizing the satellite and photo observations, allowing scientists to explore this novel combination of data sources.

Keywords

Computer vision; social media; event detection; ecology

1. INTRODUCTION

Monitoring the state of the natural world over time and space is crucial for a variety of scientific fields. Satellites can observe at a large scale but only for phenomena that can be seen from far above, and are affected by clouds and atmospheric conditions. Even a seemingly simple task such as monitoring global ground snow cover is difficult. The MODIS instruments on NASA's Terra satellite, for instance, do not produce useful observations for regions obscured by clouds (e.g., ironically, during snow storms!) and can be misled by materials like sand [23]: is the "snow" on that tropical island a freak event, or a noisy observation?

Ground stations can of course verify and fill in missing data, but they are expensive to install in remote areas. Citizen science [1, 2] uses the public to contribute observations, but requires clever design and significant incentives to derive accurate data from untrained observers. A potentially rich alternative is to mine public social media for evidence of natural events, in effect turning billions of users into citizen scientists without any explicit effort on their part. This idea is motivated by the growing body of work that mines social media to predict and observe properties of the world, including stock markets [3], elections [27], tourism [26], and so on.

Most work has used textual data like Twitter feeds, but social images are potentially a richer source of information. Everyday consumer photos often include incidental evidence about the natural world, e.g., a family portrait might show flowering plants in the background. In addition, unlike textual data, photos record visual documentation that can be analyzed and inspected; the danger of text analysis and importance of validation were recently illustrated by Google Flu Trends, which showed initial promise in tracking the spread of influenza from web search queries [7] but later proved largely inaccurate [15]. However, mining useful semantic information from unstructured image collections is a significant challenge.

In this paper, we test the feasibility of using noisy image collections to observe nature, using modern deep learning-based computer vision to recognize visual content automatically. As a case study, we investigate two particular phenomena: continental-scale snowfall and vegetation coverage. Although not as dramatic as events like earthquakes or tsunamis, these are nonetheless important properties of the environment that are key indicators of climate change, for instance. From a practical perspective, they also are relatively easy to recognize, occur frequently in social images, and have (noisy) satellite ground truth available to let us test at a large scale (over an entire continent, daily, for nine years) instead of just on occasional occurrences. This last property lets us measure statistically meaningful results on how a system may perform in practice, and this insight could be applied to other events in the future.

We first collect millions of geo-tagged, timestamped, public photos from Flickr, and daily snow and weekly vegetation satellite maps for North America. By cross referencing the photo geo-tags and timestamps with the maps, we automatically label each image with whether or not it was taken in a place with actual snow or green vegetation. We then train state-of-the-art Convolutional Neural Networks and Support Vector Machines to recognize these phenomena in individual images. Of course, these classifiers are imperfect, in part because social image data is noisy with inaccurate timestamps and geo-tags, and the satellite data is also incomplete. We thus train an additional classifier that aggregates evidence from multiple images taken at a given time and place, yielding more accurate observations. We evaluate at a large scale, training and testing on millions of Flickr images and quantitatively evaluating the performance at hundreds of thousands of places and times. Finally, we present a tool to visualize the combination of satellite and social photo-derived observations. The tool is general and can be applied to a wide range of phenomena with minimal additional effort.

*MK is now with CareerBuilder, LLC.

MM '16, October 15-19, 2016, Amsterdam, Netherlands

DOI: http://dx.doi.org/10.1145/2964284.2984067

2. RELATED WORK

Automatically crowd-sourcing data from public social media has been investigated for a variety of applications, from predicting election outcomes [27], to quantifying tourism patterns [18,26], to predicting the stock market [3], to estimating land use [24]. The vast majority of this work is based on textual analysis, even for photo collections [24, 26, 27]. For example, Zhang et al. [28] analyze Flickr photos to estimate ecological phenomena (including snow) but using text tags, which (as they point out) is limited by how accurately and precisely photographers tag photos. We explore the more difficult but potentially more accurate approach of using visual analysis to extract semantics.

A few papers have applied computer vision to recognize environmental properties in images. Most of these use video (e.g. from static webcams) so that changes over time can be easily detected. For example, Laffont et al. [13] investigate detecting transient attributes of scenes over time, Glasner et al. [8] predict temperature, Murdock et al. [20, 21] estimate cloud cover, Li et al. [19] estimate smog, and Fedorov et al. [5, 6] detect snow on mountain peaks. Compared to webcams, public photos give greater coverage: whenever a user uploads a photo to Flickr, they are contributing a potentially useful observation about the world at that time and place. Most work with photos has only estimated static properties of places like land use [17] and demographics [16, 29], and typically over limited spatial areas, in contrast to estimating time-varying events on a daily basis on a continental scale as we do.

The closest paper to our work is Wang et al. [25], which like us tries to recognize snowfall in images. Their results were quite preliminary, however, and used simple visual features like color histograms. Here we apply cutting-edge deep learning classifiers, and evaluate at a large scale with millions of images at thousands of times and places. Our web-based tool also allows users to navigate and visualize the results, not only letting people validate data from the satellite and the photos and vice-versa, but also giving greater insight into the situations in which crowd-sourced observation of the natural world is likely to succeed and when it is likely to fail.

3. OUR APPROACH

A major goal in this paper is to investigate the extent to which modern image classification could be used to accurately predict environmental conditions at a given time and place, given a collection of social images taken then and there. We investigate two specific types of conditions: (1) whether there was snow on the ground, and (2) whether there was green vegetation. Both of these properties change over time and over geospatial location on Earth. To do this we require two key steps: deciding whether or not there is evidence of snow or greenery in an individual image, and then integrating this (very noisy) evidence across multiple images to estimate the actual real-world natural state at that time and place.

Data. We collected images geo-tagged in North America and time-stamped between 2007–2015 using Flickr's public API (similar to [4]). We removed photos with inaccurate geo-tags (thresholding at 12 on Flickr's GPS precision score) and suspicious time-stamps (e.g. time taken after time uploaded), yielding 77.6 million images. We otherwise did not filter images in any way, so our set includes much noisy and confusing image content (e.g. indoor images). Throughout our experiments, we used the 2007–2010 data for training and reserved 2011–2015 as a separate test set.

For the ground truth for training and testing, we used public data from NASA's Terra satellite [9, 14, 23], which gives daily snow and bi-weekly vegetation cover maps gridded into $0.05^\circ \times 0.05^\circ$ latitude-longitude bins (roughly 5km×5km at the middle latitudes). Unfortunately, this data is neither complete nor fully accurate, primarily because many satellites cannot make accurate observations through clouds. For each day and each bin (which we call a "day-geobin"), the satellite data records the percentage of the bin that was visible, the percentage of the visible area that was covered by snow or greenery, and confidence scores. To identify day-geobins with reliable ground truth, we excluded low-confidence bins, computed a probability as a function of the snow (or greenery) and visibility percentages, and labeled those below 0.15 as non-snow (or greenery) day-geobins, and over 0.85 as snow (or greenery) day-geobins. (This is similar to what was done in [28] except that they coarsened to 1° bins, and used unspecified separate thresholds on visibility and coverage). The remaining day-geobins were ignored.

Image classification. We take a machine learning approach to image classification. In training, we consult the satellite data to find all day-geobins where there is a high confidence of the event occurring or not occurring, and label all these images as positive or negative exemplars, respectively. The disadvantage to this approach is that it is very noisy: many images are taken indoors and have no evidence of the natural world, for instance, and many images have incorrect geo-tags and timestamps. The advantage is that it permits cheap, scalable training with little human effort.

We consider two types of features: text tags and visual content. For text tags, we built a vocabulary consisting of the 1,000 most frequent tags in the training set and represented each image as a 1000-d binary vector indicating presence or absence of each tag. We then trained a linear Support Vector Machine [11] to predict whether or not the tags have evidence of the event. For visual features, we learned a model using Convolutional Neural Networks (CNNs), which are the state-of-the-art in image classification [12]. We used the AlexNet network architecture and the Caffe open-source software framework [10], and followed the popular procedure of initializing CNN weights based on a network trained on ImageNet, and then fine-tuning using our training set [22].

Aggregating evidence. The classifications on individual images are not perfect, and mislabeled geo-tags and time-stamps would yield misleading evidence even if they were. To mitigate this, we combine classification results from multiple images taken at the same time and place, taking into account the image classifier's confidence. In particular, for each day-geobin, we build a histogram of quantized confidence scores, recording how many of the photos were classified as snow and non-snow (or green/non-green) at 20 quantized confidence levels. While this improves results compared to considering single images, it suffers from the problem that users with many photos have a disproportionate influence. We thus build a histogram over *users* instead of *photos*, so that each of the 20 histogram bins counts how many users took at least one photo at that confidence level. We then trained an SVM to estimate environmental state from these histograms.

4. EXPERIMENTAL RESULTS

To evaluate the potential of user-contributed social photographs for estimating properties of the natural world, we trained classifiers using data from North America for the years 2007–2010. The training data consisted of any photos taken in any day-geobin in which the probability of the event according to the satellite was below 15% or above 85%, calculated as described above. To make results more easily interpretable and to prevent problems with unbalanced classes, we randomly sampled from the larger class to yield a roughly equal number of positive and negative exemplars for each event. For snow, there were 626,522 such photos taken by 49,462 distinct users in 87,586 distinct day-geobins; for vegetation, there were 645,694 photos by 35,510 users in 84,921 day-geobins. We

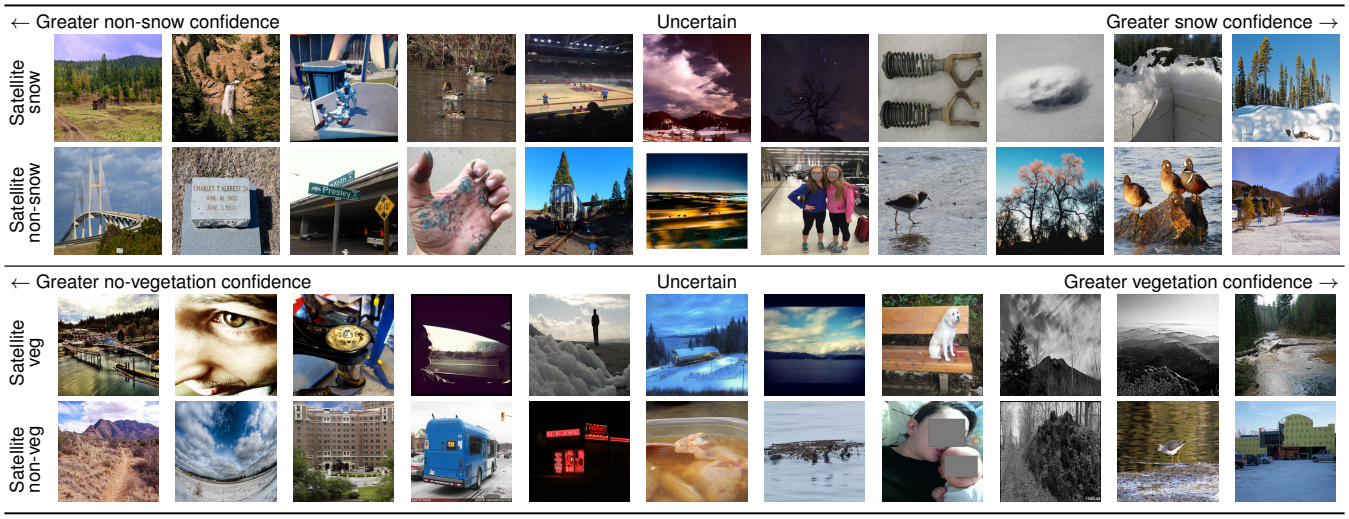

Figure 1: Classification results on random images from times and places where satellites reported snow (top), no snow (second row), high vegetation (third row) and low vegetation coverage (bottom). Images are ordered according to the classifier's confidence, from highly certain of absence (left), to uncertainty either way (middle), to highly certain of presence (right). *Faces obscured for privacy.*

tested using data from 2011–2015, again balancing the classes, for a total of 577,186 test images for snow and 769,992 for vegetation.

Individual image classifier. We first tested accuracy on the individual image classification problem. This task is extremely difficult, even for a human, because many images are taken inside or otherwise do not have useful information about the natural world, and many images have incorrect timestamps or geo-tags. The tag features achieve 63.0% accuracy for snow and 67.5% for vegetation, compared to random baselines of 50.0%. Among the vocabulary of 1,000 tags, the SVM found that *snow*, *snowshoeing*, *blizzard*, *ski*, and *sledding* were most positively correlated with snow, while *july*, *florida*, *sandiego*, *baseball*, and *bikes* were most negatively correlated; for vegetation, top positive tags were *ferns*, *redwoods*, *fawn*, *woods*, and *forest*, and top negative tags were *lasvegas*, *newmexico*, *skyscraper*, *tucson*, and *desert*. Although these tags are intuitive, they also reveal a problem with tag-based features: the classifier can easily learn biases in the data. For instance, while the tag *snow* may be a strong indicator of a snowy scene, the tag *july* is simply exploiting the bias that relatively few places in North America have snow in summer. This bias means that the classifier is unlikely to detect a highly unusual event (e.g., unprecedented summer blizzard), reminiscent of the problems discovered with Google Flu [15]. Also, tag-based analysis places the classifier at the mercy of the quality and completeness of user-supplied tags.

Visual features, in contrast, are always present and less ambiguous. We saw this reflected in the results, where visual features performed at 69.2% accuracy for snow and 80.5% for vegetation. A visualization of some sample visual classification results along with the classifier's confidences are shown in Figure 1 (see caption for details). We see that the classifier can generally separate snow images from non-snow images, although some scenes such as beaches (second row, eighth column) are similar enough to snow to cause confusion. The two most extreme "errors" (first row first column, and second row last column) illustrate cases where the CNN actually classified the image correctly; here either the satellite data was erroneous or the photo geotags or timestamps were incorrect.

Day-geobin classifier. Having classified individual images, we next test performance of these estimates in accurately classifying individual day-geobins (e.g. deciding if there was snow on the ground on a given day and place). Our accuracy on this task for snow was

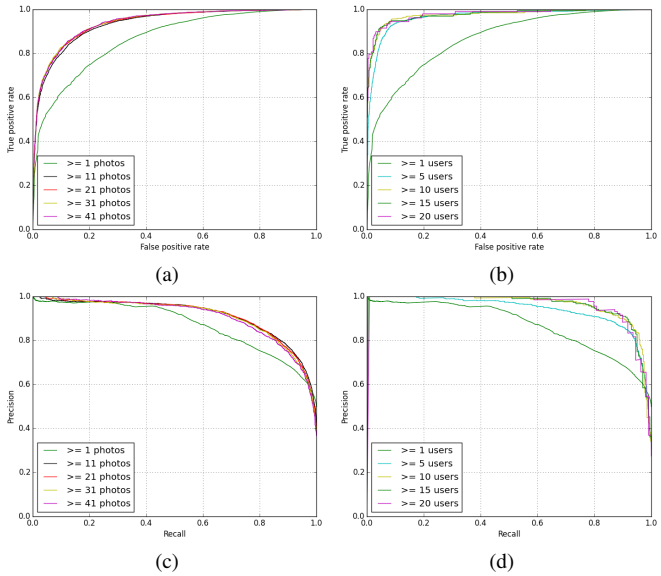

Figure 2: Performance on estimating snow presence for about 98,000 North American day-geobins from 2011–2015, in terms of *(top)* ROC and *(bottom)* Precision-Recall, as a function of number of *(left)* photos and *(right)* distinct users per bin.

about 60.8% for textual features alone, 69.3% for visual features, and 71.7% for the combination of visual features and textual features (in which we combined the two 20-d feature vectors to learn a single SVM on a 40-d feature space), compared to 50.0% random baseline; for vegetation, accuracies were 71.3% for tags, 79.4% for visual features, and 81.9% for the combination.

We have observed that most incorrectly detected day-geobins occur in places with very few observed photos contributed by few users (and often only a single photo), since in this scenario the classifier is basing its entire decision on very little evidence. Figures 2(a) and (b) plot ROC curves for snow as a function of the number of photos and number of distinct users in each day-geobin; vegetation curves are not shown due to space constraints, but the trend is similar. Accuracy increases when more than one photo is

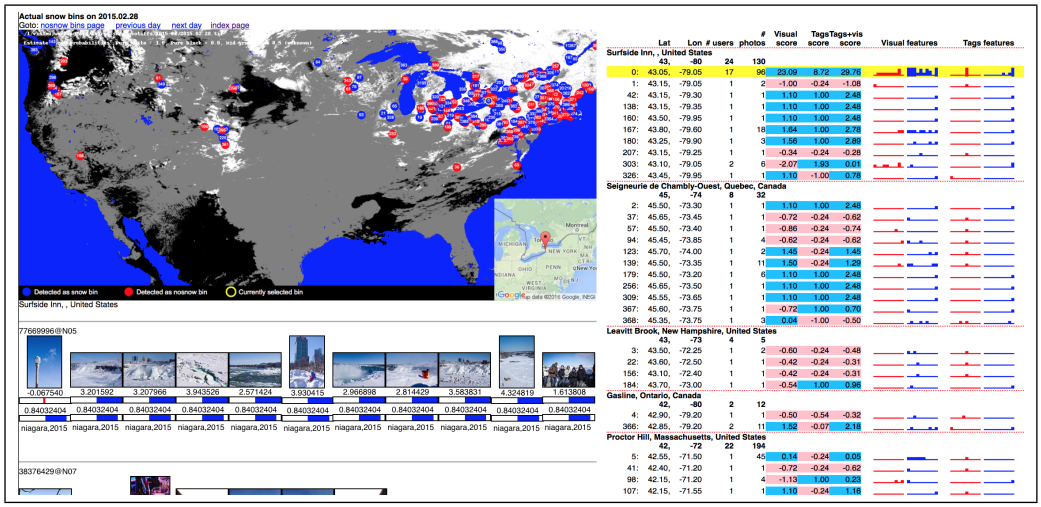

Figure 3: Screenshot of visualization tool, for snow coverage on February 28, 2015.

available, reaching about 85% for 40 photos (and eventually saturating at about 90% for 500 photos). Increasing the number of distinct users improves accuracy more dramatically, up to nearly 95% for 10 users and saturating at about 99% for 50 users. Presumably this boost is because evidence across multiple users is approximately conditionally independent given the event, as opposed to photos from any single photographer which are highly correlated. In many applications, it may be more important for scientists to retrieve places and times when specific events occurred, as opposed to accurately classifying at every place and time. Figures 2(c) and (d) shows precision-recall curves that adopt this retrieval view. At 60% recall, precision nears 90% even for day-geobins with single users, and reaches 99% for 20 users.

5. A VISUALIZATION TOOL

The quantitative results in the last section suggest that social media data could provide useful evidence about nature, but gave little insight into when the analysis would succeed or fail. We have developed a web-based tool that allows users to explore and compare satellite and social media data. Figure 3 shows a screenshot of the tool, visualizing snow coverage on one particular day. We briefly describe the main features here; please check our project website for more detailed information.[1] The map shows the satellite snow coverage, where black, gray, and white indicate no-snow, uncertain (cloud cover) and snow regions, respectively. Blue and red dots on the map indicate locations where the automatic photo-based classifier agrees and disagrees with the satellite, respectively, and the right panel of the interface lists details of these bins including position and output of the automatic day-geobin classifiers. Users can click on any geobin of interest to see photos taken at that time and place (lower left), organized by distinct user, and the visualization also shows the classification results estimated for each image. Clicking a photo shows it in detail, including text tags, geo-tags, timestamps, and other metadata. (In accordance with the Flickr Terms of Service, the images are not stored locally and clicking images leads to the photo page on Flickr.)

Although development of the tool is ongoing and we have not yet conducted a formal user study to test the tool with real users, we have informally found several interesting examples of use cases:

1. **Verifying suspicious satellite observations:** The satellite reported snow in West Virginia in Aug. 2012, but the classifier disagreed, and manual inspection of the photos taken there shows no evidence of snowfall.

2. **Complementing missing satellite data:** The satellite shows little evidence of a Jan. 2015 snowstorm in the northeast because of clouds, whereas the automatic classifier and Flickr photos confirm widespread snow coverage. The photo classifier also flags snow coverage near Blacksburg, VA on March 28, 2015, while the satellite does not; the images show a trace amount of snow that likely were not significant enough to be visible to the satellite.

3. **Debugging classification errors:** The classifier detected snow near Roanoke, VA on Jan. 1, 2014 while the satellite did not; multiple indoor scenes with white walls were incorrectly classified as containing snow. Meanwhile it also flags snow near Eugene, OR on the same day, because of photos of a distant snowy mountain peak that is in an adjacent geospatial bin.

6. CONCLUSION

We presented a technique and visualization tool for combining automatic image analysis of public Flickr photos with satellite maps for tracking natural events. We considered snow and vegetation as test cases, since continental-scale daily coverage data over nearly a decade is publicly available for these events, but the automatic classification techniques and visualization tools are general enough to be applied to a wider range of events. In ongoing work we are applying it to wildfires, flooding, and flowering of particular flower species, for example. We hope our work inspires further interest in using social photo collections and computer vision as a novel source for environmental data.

7. ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation through CAREER grant IIS-1253549 and the IU Data-to-Insight Center, and used compute facilities donated by NVidia. We thank Dennis Chen and Alex Seewald for assisting with initial data collection and system configuration.

[1] http://vision.soic.indiana.edu/snowexplorer/

8. REFERENCES

[1] Lost ladybug project. http://www.lostladybug.org.

[2] Project BudBurst. http://budburst.org/.

[3] J. Bollen, H. Mao, and X.-J. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.

[4] D. J. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg. Mapping the world's photos. In *International Conference on World Wide Web*, pages 761–770, 2009.

[5] R. Fedorov, P. Fraternali, C. Pasini, and M. Tagliasacchi. SnowWatch: snow monitoring through acquisition and analysis of user-generated content. *arXiv:1507.08958*, 2015.

[6] R. Fedorov, P. Fraternali, and M. Tagliasacchi. Snow phenomena modeling through online public media. In *IEEE International Conference on Image Processing*, pages 2174–2176, 2014.

[7] J. Ginsberg, M. Mohebbi, R. Patel, L. Brammer, M. Smolinski, and L. Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457:1012–1014, 2009.

[8] D. Glasner, P. Fua, T. Zickler, and L. Zelnik-Manor. Hot or not: Exploring correlations between appearance and temperature. In *IEEE International Conference on Computer Vision*, pages 3997–4005, 2015.

[9] D. K. Hall, G. A. Riggs, and V. V. Salomonson. MODIS/Terra Snow Cover Daily L3 Global 0.05Deg CMG V004. Boulder, CO, USA: National Snow and Ice Data Center, 2011, updated daily.

[10] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *ACM International Conference on Multimedia*, pages 675–678. ACM, 2014.

[11] T. Joachims. Making large-scale SVM learning practical. In B. Schölkopf, C. J. C. Burges, and A. J. Smola, editors, *Advances in kernel methods – support vector learning*. MIT Press, 1999.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[13] P.-Y. Laffont, Z. Ren, X. Tao, C. Qian, and J. Hays. Transient attributes for high-level understanding and editing of outdoor scenes. *ACM Transactions on Graphics (TOG)*, 33(4):149, 2014.

[14] Land Processes Distributed Active Archive Center. MODIS/Terra Vegetation Indices 16-Day L3 Global 0.05Deg CMG V005. Sioux Falls, SD: U.S. Geological Survey, 2011.

[15] D. Lazer, R. Kennedy, G. King, and A. Vespignani. The parable of Google Flu: traps in big data analysis. *Science*, 343(14 March), 2014.

[16] S. Lee, H. Zhang, and D. Crandall. Predicting geo-informative attributes in large-scale image collections using convolutional neural networks. In *IEEE Winter Conference on Applications of Computer Vision*, 2015.

[17] D. Leung and S. Newsam. Proximate Sensing: Inferring What-Is-Where From Georeferenced Photo Collections. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[18] N. Levin, S. Kark, and D. Crandall. Where have all the people gone? Enhancing global conservation using night lights and social media. *Ecological Applications*, 25(8):2153–2167, December 2015.

[19] Y. Li, J. Huang, and J. Luo. Using user generated online photos to estimate and monitor air pollution in major cities. In *ACM International Conference on Internet Multimedia Computing and Service*, 2015.

[20] C. Murdock, N. Jacobs, and R. Pless. Webcam2satellite: Estimating cloud maps from webcam imagery. In *IEEE Winter Conference on Applications of Computer Vision*, pages 214–221, 2013.

[21] C. Murdock, N. Jacobs, and R. Pless. Building dynamic cloud maps from the ground up. In *IEEE International Conference on Computer Vision*, pages 684–692, 2015.

[22] M. Oquab, L. Bottou, I. Laptev, and J. Sivic. Learning and transferring mid-level image representations using convolutional neural networks. In *CVPR*, 2014.

[23] G. Riggs, H. D., and Salomonson. MODIS Snow Products User Guide. http://modis-snow-ice.gsfc.nasa.gov/uploads/sug_c5.pdf.

[24] M. Šećerov. Analysis of Panoramio photo tags in order to extract land use information. Master's thesis, Universidade Nova de Lisboa, 2015.

[25] J. Wang, M. Korayem, and D. Crandall. Observing the natural world with Flickr. In *IEEE International Conference on Computer Vision Workshops*, pages 452–459, 2013.

[26] S. A. Wood, A. D. Guerry, J. M. Silver, and M. Lacayo. Using social media to quantify nature-based tourism and recreation. *Scientific Reports*, 3, 2013.

[27] Q. You, L. Cao, Y. Cong, X. Zhang, and J. Luo. A multifaceted approach to social multimedia-based prediction of elections. *IEEE Transactions on Multimedia*, 17(12):2271–2280, 2015.

[28] H. Zhang, M. Korayem, D. Crandall, and G. LeBuhn. Mining Photo-sharing Websites to Study Ecological Phenomena. In *International Conference on World Wide Web*, pages 749–758, 2012.

[29] B. Zhou, L. Liu, A. Oliva, and A. Torralba. Recognizing city identity via attribute analysis of geo-tagged images. In *European Conference on Computer Vision*, pages 519–534, 2014.