Bare Advanced Demo of IEEEtran.cls for Computer Society Journals

Michael Shell, *Member, IEEE,* John Doe, *Fellow, OSA,* and Jane Doe, *Life Fellow, IEEE*

Abstract—

workshop

The billions of public photos on online social media sites contain a vast amount of latent visual information about the world. In this paper, we study the feasibility of observing the state of the natural world by recognizing specific types of scenes and objects in large-scale social image collections. More specifically, we study whether we can recreate satellite maps of snowfall by automatically recognizing snowy scenes in geo-tagged, timestamped images from Flickr. Snow recognition turns out to be a surprisingly difficult and under-studied problem, so we test a variety of modern scene recognition techniques on this problem and introduce a large-scale, realistic dataset of images with ground truth annotations. As an additional proof-of-concept, we test the ability of recognition algorithms to detect a particular species of flower, the California Poppy, which could be used to give biologists a new source of data on its geospatial distribution over time.

===================================

www

The popularity of social media websites like Flickr and Twitter has created enormous collections of user-generated content online. Latent in these content collections are observations of the world: each photo is a visual snapshot of what the world looked like at a particular point in time and space, for example, while each tweet is a textual expression of the state of a person and his or her environment. Aggregating these observations across millions of social sharing users could lead to new techniques for large-scale monitoring of the state of the world and how it is changing over time. In this paper we step towards that goal, showing that by analyzing the tags and image features of geo-tagged, time-stamped photos we can measure and quantify the occurrence of ecological phenomena including ground snow cover, snow fall and vegetation density. We compare several techniques for dealing with the large degree of noise in the dataset, and show how machine learning can be used to reduce errors caused by misleading tags and ambiguous visual content. We evaluate the accuracy of these techniques by comparing to ground truth data collected both by surface stations and by Earth-observing satellites. Besides the immediate application to ecology, our study gives insight into how to accurately crowd-source other types of information from large, noisy social sharing datasets.

Index Terms—Ecology, photo mining, deep learning, image classification.

1 Introduction

latest writing

Research has used Twitter and textual social media to track what is going on with people and the world. But images are better – more information, more faithful, harder to forge, don't have to rely on textual reports. Very little work has used images because they're hard to process automatically. Even the textual work doesn't really consider things at large scale or doesn't measure performance objectively. Here we use images and estimate at continental scale.

We particularly study ecologically related phenomena. Current data is imperfect and ecologists need better data sources, and Flickr could provide that. Current observations are good enough for us to use as ground truth but not perfect. So we can use it to measure our performance but our predictions would still be useful. We choose phenomena that is obvious enough so that scene classification techniques could detect it (as opposed to fine-grained tasks like tracking particular bird species or something).

This is a hard problem. We investigate several techniques to make work better. Metadata (text, timestamps, geotags) eases the problem so we don't have to rely on vision alone. Aggregating information across multiple users means we can make mistakes.

- *M. Shell is with the Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30332.*
  E-mail: see http://www.michaelshell.org/contact.html
- *J. Doe and J. Doe are with Anonymous University.*

Manuscript received April 19, 2005; revised September 17, 2014.

Fig. 1. Many Flickr images contain evidence about the state of the natural world, including that there is snow on the ground at a particular place and time, that a particular species of bird or animal is present, and that particular species of plants are flowering.

Probabilistic confidence interval models the noise explicitly, integrating weak information together. Finally we use deep learning techniques for the vision which are state-of-the-art on classification problems.

===================================

workshop

Digital cameras and camera-enabled smartphones are now ubiquitous, with a large fraction of the population taking photos regularly and sharing them online. These millions of people taking pictures form a massive social sensor network that is (in aggregate) observing and capturing the visual world across time and space. Modern phones and cameras record metadata like geo-tags and time-stamps in addition to the images themselves, giving (noisy) calibration information about how this ad-hoc sensor network is arranged. Social media sites like Flickr and Facebook thus contain a large amount of latent visual information about the the world and how it is changing over time.

For instance, many (if not most) outdoor images contain some information about the state of the natural world, such as the weather conditions and the presence or absence of plants and animals (Figure 1). The billions of images on social media sites could be analyzed to recognize these natural objects and phenomena, creating a new source of data to biologists and ecologists. Where are marigolds blooming today, and how is this geospatial distribution different from a year ago? Are honeybees less populous this year than last year? Which day do leaves reach their peak color in each county of the northeastern U.S.? These questions can be addressed to some extent by traditional data collection techniques like satellite instruments, aerial surveys, or longitudal manual surveys of small patches of land, but none of these techniques allows scientists to collect fine-grained data at continental scales: satellites can monitor huge areas of land but cannot detect fine-grained features like blooming flowers, while manual surveys can collect high-quality and fine-grained data only in a small plot of land. Large-scale analysis of photos on social media sites could provide an entirely new source of data at a fraction of the cost of launching a satellite or hiring teams of biologist observers.

The idea of using crowd-sourced data for science and other purposes is of course not new. Citizen science projects have trained groups of volunteers to recognize and report natural phenomena (like bee counts [1], bird sightings [2], and snowfall [3]) near their homes. Data mining work has shown that social networking sites like Twitter can monitor political opinions [4], [5], predict financial markets [6], track the spread of disease [7], detect earthquakes [8], and monitor weather conditions [9]. However, the vast majority of this work has used textual data from micro-blogging sites like Twitter; very few papers have tried to do this with images, despite the fact that images offer evidence that is richer, less ambiguous, and much more difficult to fabricate. This is of course because it is much easier to scan for keywords in Twitter feeds than to automatically recognize semantic content in huge collections of images.

In this paper, we test the feasibility of observing the natural world by recognizing specific types of scenes and objects in large-scale image collections from social media. We consider a well-defined but nevertheless interesting problem: deciding whether there was snow on the ground at a particular place and on a particular day, given the set of publicly-available Flickr photos that were geo-tagged and time-stamped at that place and time. This builds on our early work in Zhang *et al* [10] which considered a similar problem, but used only tag information (essentially scanning for photos that had the tag "snow" with some very simple image processing to remove obvious outliers). Here, we explicitly test whether large-scale recognition of the image content itself could be used to do this task. Of course, snow cover can already be monitored through satellites and weather stations (although neither of these data sources is perfect: weather stations are sparse in rural areas and satellites typically cannot estimate snow cover when it is cloudy [11]), so this is not a transformative application for ecologists in and of itself. Instead, this is an interesting application for us precisely because fine-grained ground truth is available, so that we can test the accuracy of crowd-sourced observations of the natural world, and judge the feasibility of observing other natural phenomena for which there are no other possible sources of data.

We initially expected snowy scene recognition to be an easy problem, in which just looking for large white regions would work reasonably well. Surprisingly, amongst the hundreds of papers on object and scene classification in the literature, we were surprised to find very few that have explicitly considered detecting snow. A few papers on scene classification include snow-related categories [12], [13], [14], while a few older papers on natural materials detection [15], [16] consider it along with other categories. We test a variety of recognition techniques on this problem, using a new realistic dataset of several thousand images from Flickr with labeled ground truth. We find that snow detection in consumer images is surprisingly difficult, and we hope this paper and our dataset will help spark interest in this somewhat overlooked vision problem. We also consider an ecology application where reliable data does not exist and Flickr image analysis could be potentially quite valuable: estimating the geo-temporal flowering distribution of the California Poppy.

===================================

www

The popularity of social networking websites has grown dramatically over the last few years, creating enormous collections of user-generated content online. Photo-sharing sites have become particularly popular: Flickr and Facebook alone have amassed an estimated 100 billion images, with over 100 million new images uploaded every day [17]. People use these sites to share photos with family and friends, but in the process they are creating immense public archives of information about the world: each photo is a record of what the world looked like at a particular point in time and space. When combined together, the billions of photos on these sites combined with metadata including timestamps, geotags, and captions are a rich untapped source of information about the state of the world and how it is changing over time.

Recent work has studied how to mine passively-collected data from social networking and microblogging websites to make estimates and predictions about world events, including tracking the spread of disease [7], monitoring for fires and emergencies [18], predicting product adoption rates and election outcomes [4], and estimating aggregate public mood [19], [6]. In most of these studies, however, there is either little ground truth available to judge the quality of the estimates and predictions, or the available ground truth is an indirect proxy (e.g. since no aggregate public mood data exists, [19] evaluates against opinion polls, while [6] compares to stock market indices). While these studies have demonstrated promising results, it is not yet clear when crowd-sourcing data from social media sites can yield reliable estimates, or how to deal with the substantial noise and bias in these datasets. Moreover, these studies have largely focused on textual content and have not taken advantage of the vast amount of visual content online.

In this paper, we study the particular problem of estimating geo-temporal distributions of ecological phenomena using geo-tagged, time-stamped photos from Flickr. Our motivations to study this particular problem are three-fold. First, biological and ecological phenomena frequently appear in images, both because photographers take photos of them purposely (e.g. close-ups of plants and animals) or incidentally (a bird in the background of a family portrait, or the snow in the action shot of children sledding). Second, for the two phenomena we study here, snowfall and vegetation cover, large-scale (albeit imperfect) ground truth is available in the form of observations from satellites and ground-based weather stations. Thus we can explicitly evaluate the accuracy of various techniques for extracting semantic information from large-scale social media collections.

Third, while ground truth is available for these particular phenomena, for other important ecological phenomena (like the geo-temporal distribution of plants and animals) no such data is available, and social media could help fill this need. In fact, perhaps no community is in greater need of real-time, global-scale information on the state of the world than the scientists who study climate change. Recent work shows that global climate change is impacting a variety of flora and fauna at local, regional and continental scales: for example, species of high-elevation and cold-weather mammals have moved northward, some species of butterflies have become extinct, waterfowl are losing coastal wetland habitats as oceans rise, and certain fish populations are rapidly declining [20]. However monitoring these changes is surprisingly difficult: plot-based studies involving direct observation of small patches of land yield high-quality data but are costly and possible only at very small scales, while aerial surveillance gives data over large land areas but cloud cover, forests, atmospheric conditions and mountain shadows can interfere with the observations, and only certain types of ecological information can be collected from the air. To understand how biological phenomena are responding to both landscape changes and global climate change, ecologists need an efficient system for ground-based data collection to give detailed observations across the planet. A new approach for creating ground-level, continental-scale datasets is to use passive data-mining of the huge number of visual observations produced by millions of users worldwide, in the form of digital images uploaded to photo-sharing websites.

Challenges. There are two key challenges to unlocking the ecological information latent in these photo datasets. The first is how to recognize ecological phenomena appearing in photos and how to map these observations to specific places and times. Fortunately, modern photo-sharing sites collect a rich variety of non-visual information about photos, including metadata recorded by the digital camera — exposure settings and timestamps, for example — as well as information generated during social sharing — text tags, comments, and ratings, for example. Many sites also record the geographic coordinates of where on Earth a photo was taken, as reported either by a GPS-enabled camera or smartphone, or input manually by the user. Thus online photos include the ingredients necessary to produce geo-temporal data about the world, including information about content (images, tags and comments), and when (timestamp) and where (geotag) each photo was taken.

The second challenge is how to deal with the biases and noise inherent in online data. People do not photograph the Earth evenly, so there are disproportionate concentrations of activity near cities and tourist attractions. Photo metadata is often noisy or inaccurate; for example, users forget to set the clock on their camera, GPS units fail to find fixes, and users carelessly tag photos. Even photos without such errors might be misleading: the tag "snow" on an image might refer to a snow lily or a snowy owl, while snow appearing in an image might be artificial (as in an indoor zoo exhibit).

This paper. In this paper we study how to mine data from photo-sharing websites to produce crowd-sourced observations of ecological phenomena. As a first step towards the longer-term goal of mining for many types of phenomena, here we study two in particular: ground snow cover and vegetation cover ("green-up") data. Both are critical features for ecologists monitoring the earth's ecosystems. Importantly for our study, these two phenomena have accurate fine-grained ground truth available at a continental scale in the form of observations from aerial instruments like NASA's Terra earth-observing satellites [21], [11] or networks of ground-based observing stations run by the U.S. National Weather Service. This data allows us to evaluate the performance of our crowd-sourced data mining techniques at a very large scale, including thousands of days of data across an entire continent. Using a dataset of nearly 150 million geo-tagged Flickr photos, we study whether this data can potentially be a reliable resource for scientific research. An example comparing ground truth snow cover data with the estimates produced by our Flickr analysis on one particular day (December 21, 2009) is shown in Figure 2. Note that the Flickr analysis is sparse in places with few photographs,

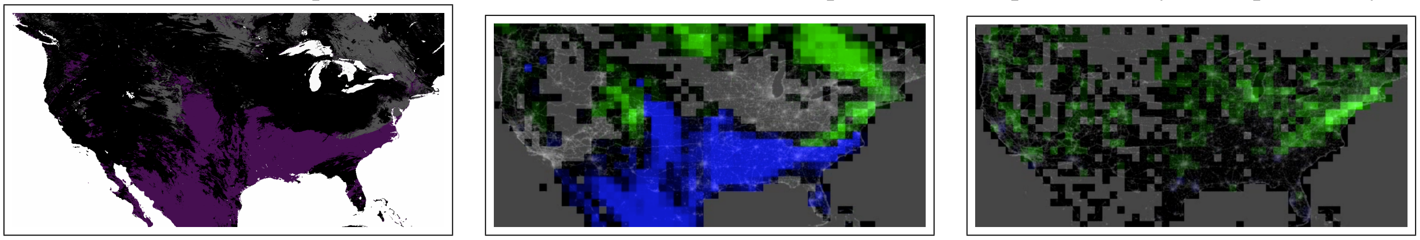


Fig. 2. Comparing MODIS satellite snow coverage data for North America on Dec 21, 2009 with estimates produced by analyzing Flickr tags (best viewed on screen in color). *Left:* Original MODIS snow data, where white corresponds with water, black is missing data because of cloud cover, grey indicates snow cover, and purple indicates no significant snow cover. *Middle:* Satellite data coarsened into 1 degree bins, where green indicates snow cover, blue indicates no snow, and grey indicates missing data. *Right:* Estimates produced by the Flickr photo analysis proposed in this paper, where green indicates high probability of snow cover, and grey and black indicate low-confidence areas (with few photos or ambiguous evidence).

while the satellite data is missing in areas with cloud cover, but they agree well in areas where both observations are present. This (and the much more extensive experimental results presented later in the paper) suggests that Flickr analysis may produce useful observations either on its own or as a complement other observational sources.

To summarize, the main contributions of this paper include:

— introducing the novel idea of mining photo-sharing sites for geo-temporal information about ecological phenomena,
— introducing several techniques for deriving crowd-sourced observations from noisy, biased data using both visual and textual tag analysis, and
— evaluating the ability of these techniques to accurately measure these phenomena, using dense large-scale ground truth.

2 RELATED WORK

A variety of recent work has studied how to apply computational techniques to analyze online social datasets in order to aid research in many disciplines [22]. Much of this work has studied questions in sociology and human interaction, such as how friendships form [23], how information flows through social networks [24], how people move through space [25], and how people influence their peers[26]. The goal of these projects is not to measure data about the physical world itself, but instead to discover interesting properties of human behavior using social networking sites as a convenient data source.

Crowd-sourced observational data. Other studies have shown the power of social networking sites as a source of observational data about the world itself. Bollen *et al* [6] use data from Twitter to try to measure the aggregated emotional state of humanity, computing mood across six dimensions according to a standard psychological test. Intriguingly, they find that these changing mood states correlate well with the Dow Jones Industrial Average, allowing stock market moves to be predicted up to 3 days in advance. However their test dataset is relatively small, consisting of only three weeks of trading data. Like us, Jin *et al* [4] use Flickr as a source of data for prediction, but they estimate the adoption rate of consumer photos by monitoring the frequency of tag use over time. They find that the volume of Flickr tags is correlated with with sales of two products, Macs and iPods. They also estimate geo-temporal distributions of these sales over time but do not compare to ground truth, so it is unclear how accurate these estimates are. In contrast, we evaluate our techniques against a large ground truth dataset, where the task is to accurately predict the distribution of a phenomenon (e.g. snow) across an entire continent each day for several years.

Crowd-sourcing from social media. Several recent studies have shown the power of social media for observing the world itself, as a special case of 'social sensing' [27]. This work includes using Twitter data to measure collective emotional state [28] (which, in turn, has found to be predictive of stock moves [6]), predicting product adoption rates and political election outcomes [4], and collecting data about earthquakes and other natural disasters [8].

Particularly striking examples include Ginsberg *et al* [7], who show that geo-temporal properties of web search queries can predict the spread of flu, and Sadilek *et al* [29] who show that Twitter feeds can predict when a given person will fall ill.

Crowd-sourced geo-temporal data. Other work has used online data to predict geo-temporal distributions, but again in domains other than ecology. Perhaps the most striking is the work of Ginsberg *et al* [7], who show that by monitoring the geospatial distribution of search engine queries related to flu symptoms, the spread of the H1N1 flu can be estimated several days before the official statistics produced by traditional means. DeLongueville *etal* [18] study tweets related to a major fire in France, but their analysis is at a very small scale (a few dozen tweets) and their focus is more on human reactions to the fire as opposed to using these tweets to estimate the fire's position and severity. In perhaps the most related existing work to ours, Singh *et al* [30] create geospatial heat maps (dubbed "social pixels") of various tags, including snow and greenery, but their focus is on developing a formal database-style algebra for describing queries on these systems and for creating visualizations. They do not consider how to produce accurate predictions from these visualizations, nor do they compare to any ground truth.

Accuracy of geo and temporal data on Flickr. Over a sample of 10 million images on Flickr.com, 37% (only a small subset) of them probably have incorrect timestamp [**?**]. The accuracy of geo-location is limited due to the camera device, and GPS precision.

Meanwhile, a lot of works are trying to correct estimate or correct geo-location of Flickr images. [31]estimates where images are taken for those missing geo-tags. They are optimizing a graph clustering problem. Attributes in their graph include textual tags, timestamps and vision content. It's inspired by an earlier work [32].

Timestamp is harder to be accurate due to the same reason of geo-tags and the time-zone problem. Thomee et.al. [33] show a detailed analysis in disagreement of camera time and GPS time. They also estimate a more accurate timestamp when users taking multiple images in a short timespan. [34] consider textual metadata to correct geo tags. They also found for users active on both Flickr and Twitter, the Twitter post at around the same time the images are taken can be a reliable reference to estimate the approximate location.

The specific application we consider here is inferring information about the state of the natural world from social media. Existing work has analyzed textual content, including text tags and Twitter feeds, in order to do this. Hyvarinen and Saltikoff [9] use tag search on Flickr to validate metereological satellite observations, although the analysis is done by hand. Zhang *et al* [10] take a large collection of geo-tagged and time-stamped Flickr photos and search for snow-related tags to produce estimates of geo-temporal snowfall distributions, and evaluate them against satellite snow maps. Singh *et al* [30] visualize geospatial distributions of photos tagged "snow" as an example of their Social Pixels framework, but they study the database theory needed to perform this analysis and do not consider the prediction problem.

Few papers have used actual image content analysis as we do here. Leung and Newsam [35] use scene analysis in geo-tagged photos to infer land cover and land use types. Murdock *et al* [36] analyze geo-referenced stationary webcam feeds to estimate cloud cover on a day-by-day basis, and then use these estimates to recreate satellite cloud cover maps. Webcams offer a complimentary data source to the social media images we consider here: on one hand, analyzing webcam data is made easier by the fact that the camera is stationary and offers dense temporal resolution; on the other hand, their observations are restricted to where public webcams exist, whereas photos on social media sites offer a potentially much denser spatial sampling of the world.

We note that these applications are related to citizen science projects where volunteers across a wide geographic area send in observations [1], [2], [3]. These projects often use social media, but require observations to be made explicitly, whereas in our work we "passively" analyze social media feeds generated by untrained and unwitting individuals.

Detecting snow in images We know of only a handful of papers that have explicitly considered snow detection in images. Perhaps the most relevant is the 2003 work of Singhal *et al* [37], [15] which studies this in the context of detecting "materials" like water, grass, sky, etc. They calculate local color and texture features at each pixel, and then compute a probability distribution over the materials at each pixel using a neural network. They partition the image into segments by thresholding these belief values, and assign a label to each segment with a probabilistic framework that considers both the beliefs and simple contextual information like relative location. They find that sky and grass are relatively easy to classify, while snow and water are most difficult. Follow-up work [16], [38] applied more modern techniques like support vector machines. Barnum *et al* [39] detect falling snow and rain, a complementary problem to the one we study here of detecting fallen snow.

Papers in the scene recognition literature have considered snowy scenes amongst their scene categories; for instance, Li *et al* [14], [13] mention snow as one possible component of their scene parsing framework, but do not present experimental results. The SUN database of Xiao *et al* [12] includes several snow-related classes like "snowfield," "ski slope," "ice shelf," and "mountain snowy," but other categories like "residential neighborhood" sometimes have snow and sometimes do not, such that detecting these scenes alone is not sufficient for our purposes.

Vegetation classification.

[40]identifies plant species by leaf images. They focus on accurate leaf segmentation according to color difference of leaf and background, curvature distribution over scale, and nearest neighbor matching.

[41]introduces multiple Gist models in scene classification. There happen to be a test set of vegetation shows Gist feature works great on vegetation classification.

[42]This one may be a little too old It's using man-made mask and neural networks. (could be compared with deep learning)

This paper [43] is the closest one to our purpose. They consider color, texture features in images and get good result. But they only test on a very limited dataset where the positive images are either with one tree in the center or full of trees or meadow. This is inadequate when we are working with very large number of public shared images.

Citizen science. While some volunteer-based biology efforts like the Lost Ladybug Project [44] and the Great Sunflower Project [1] use social networking sites to organize and recruit volunteer observers, we are not aware of any work that has attempted to passively mine ecological data from social media sites. The visual data in online social networking sites provide a unique resource for tracking biological phenomena: because they are images, this data can be verified in ways that simple text cannot. In addition, the rapidly expanding quantity of online images with geo-spatial and temporal metadata creates a fine-scale record of what is happening across the globe. However, to unlock the latent information in these vast photo collections, we need mining and recognition tools that can efficiently process large numbers of images, and robust statistical models that can handle incomplete and incorrect observations.

3 METHOD

3.1 DataSet

We use a sample of nearly 150 million geo-tagged, timestamped Flickr photos as our source of user-contributed observational data about the world. We collected this data using the public Flickr API, by repeatedly searching for photos within random time periods and geo-spatial regions, until the entire globe and all days between January 1, 2007 and December 31, 2010 had been covered. We applied filters to remove blatantly inaccurate metadata, in particular removing photos with geotag precision less than about city-scale (as reported by Flickr), and photos whose upload timestamp is the same as the EXIF camera timestamp (which usually means that the camera timestamp was missing).

For ground truth we use large-scale data originating from two independent sources: ground-based weather stations, and aerial observations from satellites. For the ground-based observations, we use publicly-available daily snowfall and snow depth observations from the U.S. National Oceanic and Atmospheric Administration (NOAA) Global Climate Observing System Surface Network (GSN) [45]. This data provides highly accurate daily data, but only at sites that have surface observing stations. For denser, more global coverage, we also use data from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument aboard NASA's Terra satellite. The satellite is in a polar orbit so that it scans the entire surface of the earth every day. The MODIS instrument measures spectral emissions at various wavelengths, and then post-processing uses these measurements to estimate ground cover. In this paper we use two datasets: the daily snow cover maps [11] and the two-week vegetation averages [21]. Both of these sets of data including an estimate of the percentage of snow or vegetation ground cover at each point on earth, along with a quality score indicating the confidence in the estimate. Low confidence is caused primarily by cloud cover (which changes the spectral emissions and prevents accurate ground cover from being estimated), but also by technical problems with the satellite. As an example, Figure 2 shows raw satellite snow data from one particular day.

3.1.1 Snow DataSet

The distribution of geo-tagged Flickr photos is highly non-uniform, with high peaks in population centers and tourist locations. Sampling uniformly at random from Flickr photos produces a dataset that mirrors this highly non-uniform distribution, biasing it towards cities and away from rural areas. Since our eventual goal is to reproduce continental-scale satellite maps, rural areas are very important. An alternative is biased sampling that attempts to select more uniformly over the globe, but has the disadvantage that it no longer reflects the distribution of Flickr photos. Other important considerations include how to find a variety of snowy and non-snowy images, including relatively difficult images that may include wintry scenes with ice but not snow, and how to prevent highly-active Flickr users from disproportionately affecting the datasets.

We strike a compromise on these issues by combining together datasets sampled in different ways. We begin with a collection of about 100 million Flickr photos geo-tagged within North America and collected using the public API (by repeatedly querying at different times and geo-spatial areas, similar to [46]). From this set, we considered only photos taken before January 1, 2009 (so that we could use later years for creating a separate test set), and selected: (1) all photos tagged *snow, snowfall, snowstorm,* or *snowy* in English and 10 other common languages (about 500,000 images); (2) all photos tagged *winter* in English and about 10 other languages (about 500,000 images); (3) a random sample of 500,000 images. This yielded about 1.4 million images after removing duplicates. We further sampled from this set in two ways. First, we selected up to 20 random photos from each user, or all photos if a user had less than 20 photos, giving about 258,000 images. Second, we sampled up to 100 random photos from each $0.1^\circ \times 0.1^\circ$ latitude-longitude bin of the earth (roughly 10km $\times$ 10km at the mid latitudes), yielding about 300,000 images. The combination of these two datasets has about 425,000 images after removing duplicates, creating a diverse and realistic selection of images. We partitioned this dataset into test and training sets on a per-user basis, so that all of any given user's photos are in one set or the other (to reduce the potential for duplicate images appearing in both training and test).

We then presented a subset of these images to humans and collected annotations for each image. We asked people to label the images into one of four categories: (1) contains obvious snow near the camera; (2) contains a trace amount of snow near the camera; (3) contains obvious snow but far away from the camera (e.g. on a mountain peak); and (4) does not contain snow. For our application of reconstructing snowfall maps, we consider (1) and (2) to be positive classes and (3) and (4) to be negative, since snowfall in the distance does not give evidence of snow at the image's geo-tagged location. In total we labeled 10,000 images.

3.1.2 Vegetation DataSet

We build a data set with over 10000 images. They are taken before 2009, and are composed by images with "forest" and "summer" like tags and also random images without any tag preference. These images are labeled with categories *"Outdoor Greenery","Outdoor non-Greenery","Indoor","Other-modified"*,and *"Not available"*.

Finally, we build a positive set with images in category *"Outdoor Greenery"* and a negative set with images in categories *"Outdoor non-Greenery"* and *"Indoor"*. To learn a image classification model, we build a training set with 4000 images and a testing set with 1900 images. In training and testing set, there are equal number of positive and negative samples. To show the diversity of our Flickr image dataset, in figure 3 we present a random sample of images in our vegetation dataset labeled as positive and negative.

In continental-scale prediction, we only look at images on Flickr.com in year 2007 to 2010, with no tag limitation. We filter out the photos with inaccurate timestamps and geotags. We only use images with geotag precision no less than 13. (what's that mean?)

3.2 Extracting semantics using tags from individual images

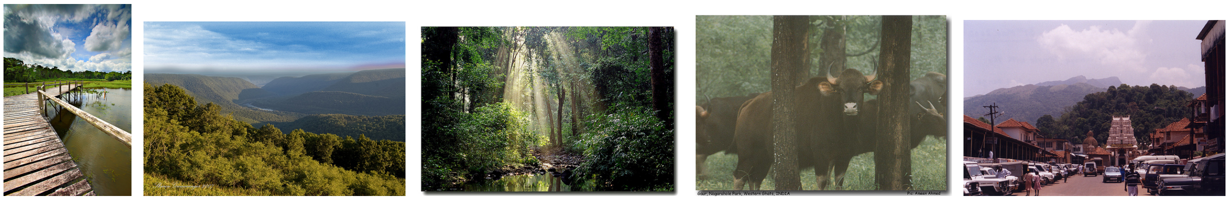
We consider two learning paradigms. The first is to produce a single exemplar for each bin in time and space consisting of the set of all tags used by all users. For each of these exemplars, the NASA and/or NOAA ground truth data gives a label (snow or non-snow). We then use standard machine learning algorithms like Support Vector Machines and decision trees to identify the most discriminative tags and tag combinations. In the second paradigm, our goal instead is to classify individual *photos* as containing snow or not, and then use these classifier outputs to compute the number of positive and non-positive photos in each bin (i.e., to compute m and n in the likelihood ratio described in the last section).

3.3 Extracting semantics using images from individual images

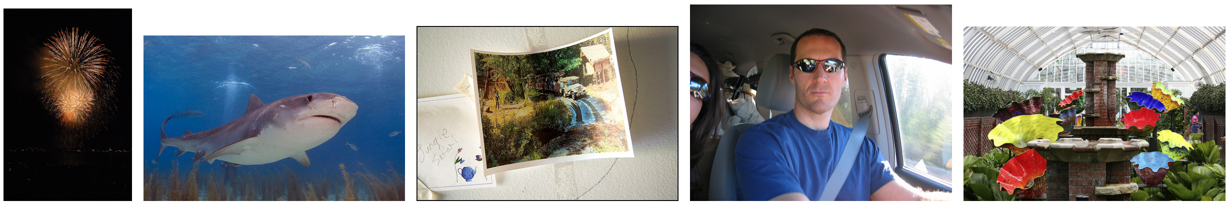
As noted above, we are aware of very little work that has considered the problem of detecting snow in images: the most relevant work [37] considers snow in the context of natural materials classification, but is over 10 years old, uses a small and biased dataset, and does not report classification results. Recent work on scene understanding [12] sometimes includes snow-related scenes, but none of this work applies directly to our problem because snow can appear across a range of different scene types. Snow is really an object, not a type of scene, but we are not aware of any work on recognizing snow in the object detection literature.

We thus begin by assembling a large-scale realistic image dataset as described above, and test a variety of modern classification techniques on the problem of snowy scene detection. We use a labeled subset of this dataset to train classifiers and to test their performance, and then apply these classifiers to the problem of generating satellite-like snowfall maps using image analysis on geo-tagged, time-stamped Flickr photos.

Snow is a somewhat unique visual phenomenon, and we claim that detecting it in images is a unique recognition task. In some



(a) Random positive images in vegetation dataset



(b) Random negative images in vegetation dataset

Fig. 3. Random images from our hand-labeled dataset. Public sharing images are various in quality, contents, illumination and view angle. Negative images like winter trees without leaves, or indoor images capturing a photo of forest are more confusing.

cases, snow can be detected by coarse scene recognition: ski slopes or snowy landscapes are distinctive scenes. But snow can appear in any kind of outdoor scene, and is thus like an object. However, unlike most objects that have some distinctive features, snow is simply a white, near-textureless material. (In fact, our informal observation is that humans detect snow not by recognizing its appearance, but by noticing that other expected features of a scene are occluded; in this sense, detecting snow is less about the features that are seen and more about the features that are *not* seen. We leave this as an observation to inspire future work.) We tested a variety of off-the-shelf visual features for classifying whether an image contains fallen snow. We used Support Vector Machines for classification, choosing kernels based on the feature type. Intuitively, color is a very important feature for detecting snow, and thus we focused on features that use color to some degree.

Similar to now , vegetation has the signature green color. The leaves of plants have distinctive visual texture. So we employ SIFT feature to analyze the local gradient distribution. And we also extract GIST feature to describe texture feature and global context.

We describe our visual feature

Color histograms We begin with perhaps the simplest of color features. We build joint histograms in CIELAB space, with 4 bins on the lightness dimension and 14 bins along each of the two color dimensions, for a total of 784 bins. We experimented with other quantizations and found that this arrangement worked best. We encode the histogram as a 784 dimensional feature and use an SVM with a chi-squared distance (as in [12]).

Tiny images We subsample images to 16×16 pixels, giving 256 pixels per RGB color plane and yielding a 768 dimensional feature vector. Drastically reducing the image dimensions yields a feature that is less sensitive to exact alignment and more computationally feasible [47].

Spatial Moments Tiny images capture coarse color and spatial scene layout information, but much information is discarded during subsampling. As an alternative approach, we convert the image to LUV color space, divide it into 49 blocks using a 7×7 grid, and then compute the mean and variance of each block in each color channel. Intuitively, this is a low-resolution image and a very simple texture feature, respectively. We also compute maximum, minimum, and median value within each cell, so that the final feature vector has 735 dimensions.

Color Local Binary Pattern (LBP) with pyramid pooling LBP represents each 9×9 pixel neighborhood as an 8-bit binary number by thresholding the 8 outer pixels by the value at the center. We build 256-bin histograms over these LBP values, both on the grayscale image and on each RGB color channel [48]. We compute these histograms in each cell of a three-level spatial pyramid, with 1 bin at the lowest level, 4 bins in a 2×2 grid at the second level, and 16 bins in a 4×4 grid at the third level. This yields a $(1 + 4 + 16) \times 4 \times 256$ = 21504 dimensional feature vector for each image.

GIST We also apply GIST features, which capture coarse texture and scene layout by applying a Gabor filter bank followed by down-sampling [49]. Our variant produces a 1536-dimensional vector and operates on color planes. Scaling images to have square aspect ratios before computing GIST improved classification results significantly [50].

Color SIFT histogram. We extract dense SIFT feature on each of the RGB color plane, and concatenate them to build color SIFT feature. The dense SIFT feature is extracted from every 2 pixels by 2 pixels bin, with a step size of 5 pixels. In this way, we achieve representative key points and reasonable computation complex.

3.3.1 Deep learning

Recently the Conventional Neural Network (CNN) [51] has gained a lot of attention in the vision community, as it outperformed all other techniques in the ImageNet challenge (the most famous object category detection competition) [52].

CNN is a special type of feed forward neural network inspired by the biological process [51] in cats' visual cortex. CNN enjoys additional features that distinguish it from the standard neural networks: shared wights and sparse connectivity. A layer in CNN may consist of three different stages: convolution, non-linear activation, and pooling. In the convolution stage, a set of convolution filters is applied in parallel. The output of a convolution filter is then passed to non-linear activation functions (e.g., rectified linear activation function, sigmoid activation function). The final stage is pooling, where the net output is manipulated based on its neighbors (e.g., max pooling , L_2 norm, and weighted average). Pooling makes the network invariant to the translation of the input

The key idea behind this approach is that instead of first designing low-level features by hand and then running a machine learning algorithm, a single unified algorithm should learn both the low-level features and the high-classifier simultaneously.

We apply CNN to detect snow and vegetation on image level. We followed Oquab [53] et al. and started with a model pre-trained on the huge ImageNet dataset then we train our models using hand-labeled data sets.

3.4 Combining evidence together across users

Our goal is to estimate the presence or absence of a given ecological phenomenon (like a species of plant or flower, or a meteorological feature like snow) on a given day and at a given place, using only the geo-tagged, time-stamped photos from Flickr. One way of viewing this problem is that every time a user takes a photo of a phenomenon of interest, they are casting a "vote" that the phenomenon actually occurred in a given geospatial region. We could simply look for tags indicating the presence of a feature – i.e. count the number of photos with the tag "snow" – but sources of noise and bias make this task challenging, including:

- *Sparse sampling:* The geospatial distribution of photos is highly non-uniform. A lack of photos of a phenomenon in a region does not necessarily mean that it was not there.
- *Observer bias:* Social media users are younger and wealthier than average, and most live in North America and Europe.
- *Incorrect, incomplete and misleading tags:* Photographers may use incorrect or ambiguous tags — e.g. the tag "snow" may refer to a snowy owl or interference on a TV screen.
- *Measurement errors:* Geo-tags and timestamps are often incorrect (e.g. because people forget to set their camera clocks).

A statistical test. We introduce a simple probabilistic model and use it to derive a statistical test that can deal with some such sources of noise and bias. The test could be used for estimating the presence of any phenomenon of interest; without loss of generality we use the particular case of snow here, for ease of explanation. Any given photo either contains evidence of snow (event s) or does not contain evidence of snow (event $\bar{s}$). We assume that a given photo taken at a time and place with snow has a fixed probability $P(s|snow)$ of containing evidence of snow; this probability is less than 1.0 because many photos are taken indoors, and outdoor photos might be composed in such a way that no snow is visible. We also assume that photos taken at a time and place without snow have some non-zero probability $P(s|\overline{snow})$ of containing evidence of snow; this incorporates various scenarios including incorrect timestamps or geo-tags and misleading visual evidence (e.g. man-made snow).

Let m be the number of snow photos (event s), and n be the number of non-snow photos (event $\bar{s}$) taken at a place and time of interest. Assuming that each photo is captured independently, we can use Bayes' Law to derive the probability that a given place has snow given its number of snow and non-snow photos,

$$\begin{aligned} P(snow|s^m,\bar{s}^n) &= \frac{P(s^m,\bar{s}^n|snow)P(snow)}{P(s^m,\bar{s}^n)} \\ &= \frac{\binom{m+n}{m}p^m(1-p)^nP(snow)}{P(s^m,\bar{s}^n)}, \end{aligned}$$

where we write $s^m, \bar{s}^n$ to denote m occurrences of event s and n occurrences of event $\bar{s}$, and where $p = P(s|snow)$ and $P(snow)$ is the prior probability of snow. A similar derivation gives the posterior probability that the bin does not contain snow,

$$P(\overline{snow}|s^m,\bar{s}^n) = \frac{\binom{m+n}{m}q^m(1-q)^nP(\overline{snow})}{P(s^m,\bar{s}^n)},$$

where $q = P(s|\overline{snow})$. Taking the ratio between these two posterior probabilities yields a likelihood ratio,

$$\frac{P(snow|s^m,\bar{s}^n)}{P(\overline{snow}|s^m,\bar{s}^n)} = \frac{P(snow)}{P(\overline{snow})}\left(\frac{p}{q}\right)^m\left(\frac{1-p}{1-q}\right)^n. \quad (1)$$

This ratio can be thought of as a measure of the confidence that a given time and place actually had snow, given photos from Flickr.

A simple way of classifying a photo into a positive event s or a negative event $\bar{s}$ is to use text tags. We identify a set $\mathcal{S}$ of tags related to a phenomenon of interest. Any photo tagged with at least one tag in $\mathcal{S}$ is declared to be a positive event s, and otherwise it is considered a negative event $\bar{s}$. For the snow detection task, we use the set $\mathcal{S}$={snow, snowy, snowing, snowstorm}, which we selected by hand.

The above derivation assumes that photos are taken independently of one another, which is generally not true in reality. One particular source of dependency is that photos from the same user are highly correlated with one another. To mitigate this problem, instead of counting m and n as numbers of *photos*, we instead let m be the number of *photographers* having at least one photo with evidence of snow, while n is the numbers of photographers who did not upload any photos with evidence of snow.

The probability parameters in the likelihood ratio of equation (1) can be directly estimated from training data and ground truth. For example, for the snow cover results presented in Section **??**, the learned parameters are: $p = p(s|snow) = 17.12\%$, $q = p(s|\overline{snow}) = 0.14\%$. In other words, almost 1 of 5 people at a snowy place take a photo containing snow, whereas about 1 in 700 people take a photo containing evidence of snow at a non-snowy place.

Figure 2 shows a visualization of the likelihood ratio values for the U.S. on one particular day using this simple technique with $\mathcal{S}$={snow, snowy, snowing, snowstorm}. High likelihood ratio values are plotted in green, indicating a high confidence of snow in a geospatial bin, while low values are shown in blue and indicate high confidence of no snow. Black areas indicate a likelihood ratio near 1, showing little conference either way, and grey areas lack data entirely (having no Flickr photos in that bin on that day).

Voting. Voting is an interesting technique because of its simplicity. Voting simply counts the number of users who have annotated at least one photo in a given bin and day with a snow-related tag. Figure 4 plots precision versus the number of votes for snow retrieval. The shape of these curve illustrates why crowd-sourced observations of the world can be reliable, if enough people are involved: as the number of votes for snow increases, it becomes progressively less likely that these independent observations are coincidental, and more likely that they are caused by the presence or absence of an actual phenomenon. It is interesting to notice that when there are 7 or more snow voters, snow prediction precision becomes 100%, while the same is true for non-snow prediction when the number of non-snow voters reaches 33 if there are no snow voters in the bin.

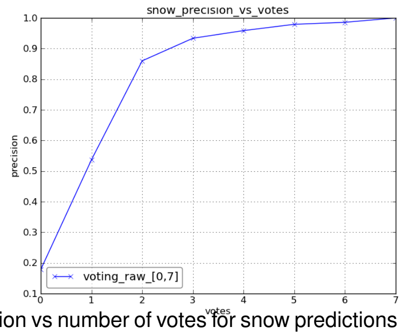


Fig. 4. Precision vs number of votes for snow predictions using the voting method.

Feature	Kernel	Accuracy
Random Baseline	—	50.0%
Gist	RBF	73.7%
Color	χ^2	74.1%
Tiny	RBF	74.3%
Spatial Color Moments	RBF	76.2%
Spatial pyramid LBP	RBF	**77.0%**
All features	linear	**80.5%**
CNN	-	**88.06%**

TABLE 1
Performance of different features for snow detection, all using SVMs for classification.

4 EXPERIMENTS AND RESULTS

4.1 Snow Case

4.1.1 Single Image classification

We used a variety of visual features for classifying whether an image contains fallen snow. We used Support Vector Machines for classification, choosing kernels based on the feature type.

We tested these approaches to detecting snow on our dataset of 10,000 hand-labeled images. We split this set into a training set of 8,000 images and a test set of 2,000 images, sampled to have an equal proportion of snow and non-snow images (so that the accuracy of a random baseline is 50%). Table 1 presents the results. We observe that all of the features perform significantly better than a random baseline. Gist, Color Histograms and Tiny Image all give very similar accuracies, within a half percentage point of 74%. Spatial Moments and LBP features perform slightly better at 76.2% and 77.0%. We also tested a combination of all features by learning a second-level linear SVM on the output of the five SVMs; this combination performed significantly better than any single feature, at 80.5%.

Figure 5 shows classification performance in terms of an ROC curve, as well as a precision-recall curve in which the task is to retrieval photos containing snow. The precision-recall curve shows that at about 20% recall, precision is very near to 100%, while even at 50% recall, precision is close to 90%. This is a nice feature because in many applications, it may not be necessarily to correct classify all images, but instead to find some images that most likely contain a subject of interest. To give a sense for the difficulty and failure modes of our dataset, we show a random sample of correct and incorrect classification results in Figure **??**.

The best performance we had using our traditional visual features using SVM is 80.5% accuracy. We also build CNN visual model for snow using Imagenet pre-trained model. We fine-tune our model using our training data. CNN achieves 88% accuracy which outperforms all other features by 7.44% . Therefore, we used CNN as our visual model for final predictions. Similar to visual model we also build SVM using only tags as features and our text classifier achieves 87% accuracy.

We now turn to presenting experimental results for estimating the geo-temporal distributions of snow.

4.1.2 Snow prediction on cities

We first test how well the Flickr data can predict snowfall at a local level, and in particular for cities in which high-quality surface-based snowfall observations exist and for which photo density is high.

We choose 4 U.S. metropolitan areas, New York City, Boston, Chicago and Philadelphia, and try to predict both daily snow presence as well as the quantity of snowfall. For each city, we define a corresponding geospatial bounding box and select the NOAA ground observation stations in that area. For example, Figure 6 shows the the stations and the bounding box for New York City. We calculate the ground truth daily snow quantity for a city as the average of the valid snowfall values from its stations.

We call any day with a non-zero snowfall or snowcover to be a snow day, and any other day to be a non-snow day.

Figure 6 also presents some basic statistics for these 4 cities. All of our experiments involve 4 years (1461 days) of data from January 2007 through December 2010; we reserve the first two years for training and validation, and the second two years for testing.

Daily snow classification for 4 cities. Figure 7(a) presents ROC curves for this daily snow versus non-snow classification task on New York City. The figure compares the likelihood ratio confidence score from equation (1) to the baseline approaches (voting and percentage), using the tag set $\mathcal{S}$={snow, snowy, snowing, snowstorm}. The area under the ROC curve (AUC) statistics are 0.929, 0.905, and 0.903 for confidence, percentage, and voting, respectively, and the improvement of the confidence method is statistically significant with $p = 0.0713$ according to the statistical test of [54]. The confidence method also outperforms other methods for the other three cities (not shown due to space constraints). ROC curves for all 4 cities using the likelihood scores are shown in Figure 7(b). Chicago has the best performance and Philadelphia has the worst; a possible explanation is that Chicago has the most active Flickr users per day (94.9) while Philadelphia has the least (43.7).

These methods based on presence or absence of tags are simple and very fast, but they have a number of disadvantages, including that the tag set must be manually chosen and that negative correlations between tags and phenomena are not considered. We thus tried training a classifier to learn these relationships automatically. For each day in each city, we produce a single binary feature vector indicating whether or not a given tag was used on that day. Also we tried to build classifiers trained based on our likelihood ratio computed based on tags or our visual model predictions. Table 2 shows the results foe these classifiers. Best prefromance obtained when we combine the confidence scores of tags and visual model based on CNN.

4.1.3 Continental-scale snow prediction

Predicting snow for individual cities is of limited practical use because accurate meteorological data already exists for these highly populated areas. Here we ask whether phenomena can be monitored at a continental scale, a task for which existing data

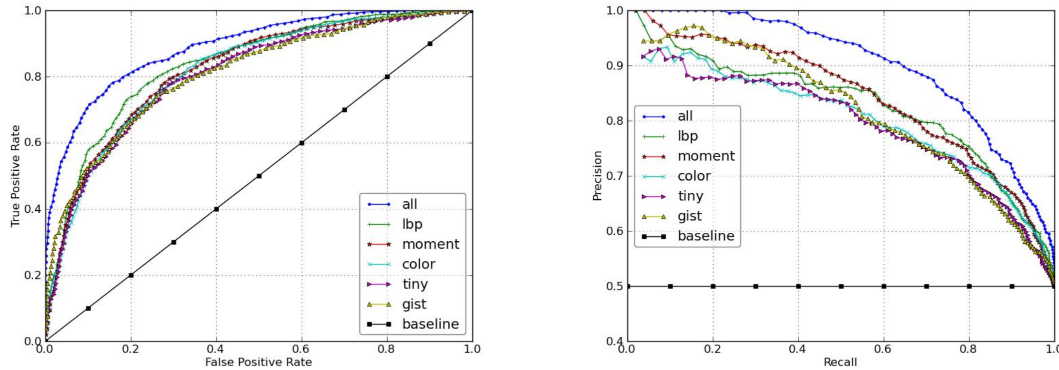
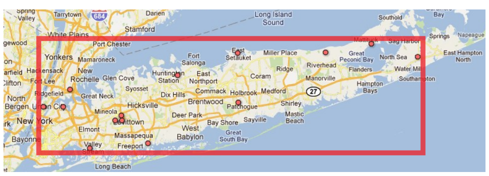


Fig. 5. Snow classification results for different features and combinations, in terms of *(left):* ROC curves for the task of classifying snow vs. non-snow images; and *(right):* Precision-Recall curves for the task of retrieving snow images.

TABLE 2
Results for Confidence score model using tags and visual classifiers for our 4 cities .

City	baseline	tags	tag confidence	vision conf	tags conf and vision conf
NYC	85%	85.75 %	90.4241 %	90.2873 %	92.3393 %
Chicago)	72.80%	93.5616 %	94.1176 %	93.1601 %	95.0752 %
Boston	75.60%	90.5479 %	89.1781 %	85.2055 %	91.2329 %
Philly	80.50%	85.34%	89.1929 %	85.0889 %	89.1929 %



	NYC	Chicago	Boston	Philadelphia
Mean active Flickr users / day	65.6	94.9	59.7	43.7
Approx. city area (km^2)	3,712	11,584	11,456	9,472
User density (avg users/unit area)	112.4	52.5	33.5	29.6
Mean daily snow (inches)	0.28	0.82	0.70	0.35
Snow days (snow¿0 inches)	185	418	373	280
Number of obs. stations	14	20	41	26

Fig. 6. *Top:* New York City geospatial bounding box used to select Flickr photos, and locations of NOAA observation stations. *Bottom:* Statistics about spatial area, photo density, and ground truth for each of the 4 cities.

sources are less complete and accurate. We use the photo data and ground truth described in Section **??**, although for the experiments presented in this paper we restrict our dataset to North America (which we defined to be a rectangular region spanning from 10 degrees north, -130 degrees west to 70 degrees north, -50 degrees west). (We did this because Flickr is a dominant photo-sharing site in North America, while other regions have other popular sites — e.g. Fotolog in Latin America and Renren in China.)

The spatial resolution of the NASA satellite ground truth datasets is 0.05 degrees latitude by 0.05 degrees longitude, or about $5 \times 5km^2$ at the equator. (Note that the surface area of these bins is non-uniform because lines of longitude get closer together near the poles.) However, because the number of photos uploaded to Flickr on any particular day and at any given spatial location is relatively low, and because of imprecision in Flickr geo-tags, we produce estimates at a coarser resolution of 1 degree square, or roughly $100 \times 100km^2$. To make the NASA maps comparable, we downsample them to this same resolution by averaging the high confidence observations within the coarser bin. We then threshold the confidence and snow cover percentages to annotate each bin with one of three ground truth labels:

— Snow bin, if confidence is above 90 and coverage above 80,
— Non-snow bin, if confidence is above 90 and coverage is 0,
— Unknown bin, otherwise.

Figure 8 shows the precision and recall curve of snow prediction in continental-scale. Here we limit our predictions for the bins which have photos taken at that time and location, we do this by keeping the bins have ground truth and photos at the same time. We computed our confidence scores based on tags and image-classification, then we trained simple decision tree to learn the correct thresholds to make final prediction. We achieve almost 0.5% over the baseline (cutting the error rate by more than 20%), the baseline in our case is the majority class which predicts now snow all the time.

4.2 Vegetation case

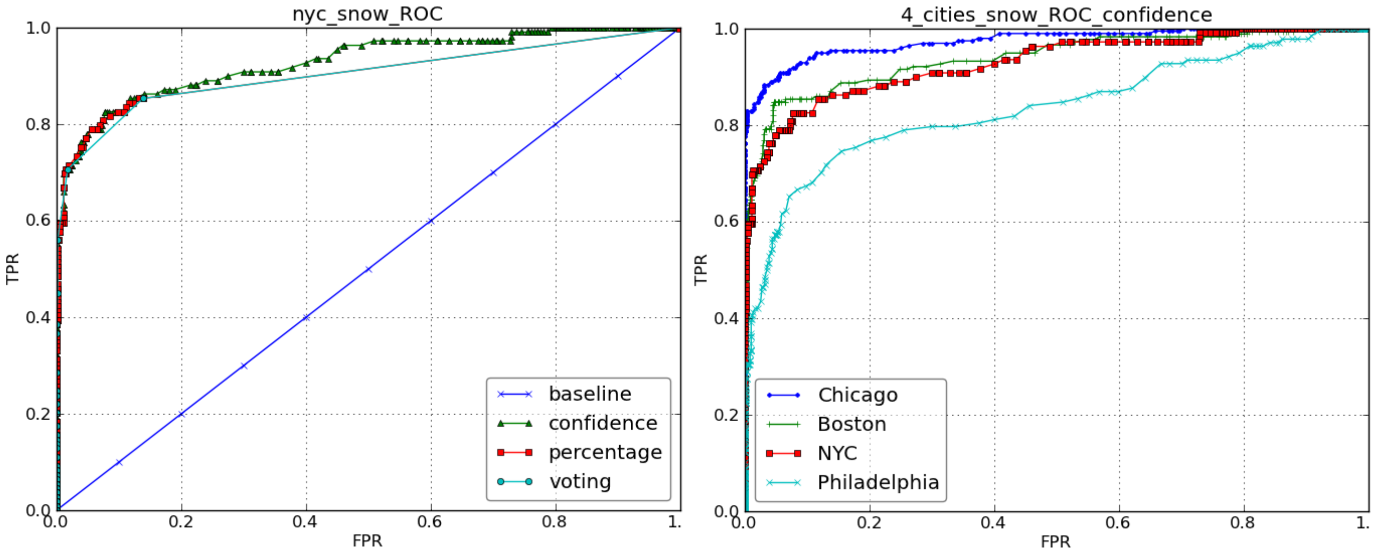
4.2.1 *Single image classification*

Using the method we describe in Section 3, we train and test the vision model on our hand-labeled data set. There are 4000 images in training set and 2000 images in testing set. In both training and testing set, the number of positive and negative images are the same. Here we present the results on image classification level.

4.2.2 *Vegetation coverage over time and space*

We consider north America area has more images uploaded to photo-sharing website, and is also where Ecologists in the US would be interested in the changing color of vegetation.

We combine all the evidence over space and time in North America from 2007 to 2010. We compute confidence score described in Section 3. The prior probability of a place being covered by vegetation at some time is 75.2%. For an image taken from a



(a)(b)

Fig. 7. ROC curves for binary snow predictions: (a) ROC curves for New York City, comparing likelihood ratio confidence score to voting and percentage approaches, (b) ROC curves for 4 cities using the likelihood scores

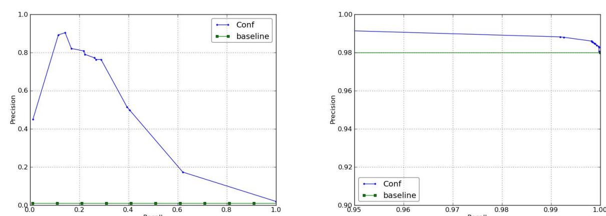


Fig. 8. Precision and recall curve of snow prediction (left) and nonsnow (right) in continental scale.

TABLE 3
Results for our visual models for vegetation.

Visual feature	Accuracy
Random Baseline	50.00%
Color SIFT	78.10%
Color GIST	82.58%
SIFT and GIST	85.9%
CNN%	88.0%

place covered by green vegetation at that time, the probability of this image being a green image is 27.18%. On the other hand, it's only 3.03% probability to see a green image in a place not covered by enough green vegetation at that time.

While the satellite has ground truth for 87594 bins in North America, our method predicts 61602 bins (70.3% in quantity). Moreover, about 20% of satellite ground truth locate in north Canada. On the other hand, our data is from users in social media. So our prediction focus on more populated locations or places people like to visit such as natrual scenery.

In North America, the overall accuracy of our method is 93.2% comparing to the 86.6% majority baseline. The precision of green bins is 98.8% and the precision of non-green bins is 68.2%. Recall of green bins is 93.3% and recall of non-green bins is 92.5%.

Figure 9 shows the precision and recall curve of snow prediction in continental-scale.

Generally, all the false negative error is due to the sparseness of data. While not enough images are collected at certain location during some time, there is either no green image found or green

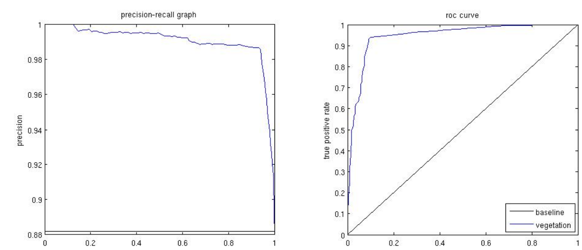


Fig. 9. Precision and recall curve of vegetation prediction in continental scale.

images are too few compare to the quantity of non-green images. On the other hand, false positive error is rare (less than 1%) and complex. We found most images in the false positive bins are actually green vegetation images. (here we need some more explanation) In figure 10, we show some examples of images in false positive bins.

4.2.3 *Performance at single place over time*

Figure 11 shows vegetation coverage of 6 places over 2009 and 2010. Prediction results on top usually have more data available than groud truth on the bottom. We use public sharing Flickr images. These images are more likely taken from more populated or more popular locations. The satellite ground truth

4.2.4 *Single time over place*

Sample maps are presented in figure 12. (I should print an outline of continent as background of these maps.)

5 DISCUSSION AND CONCLUSION

Big wrap-up, lots of ideas for future work.

In this paper, we show a real case of using public-sharing data to estimate ecology information. We used get this information only from some institution with the help of satellite.

In this paper, we propose using photo-sharing social media sites as a means of observing the state of the natural world, by automatically recognizing specific types of scenes and objects in large-scale social image collections. This work is an initial step towards a long-term goal of monitoring important ecological

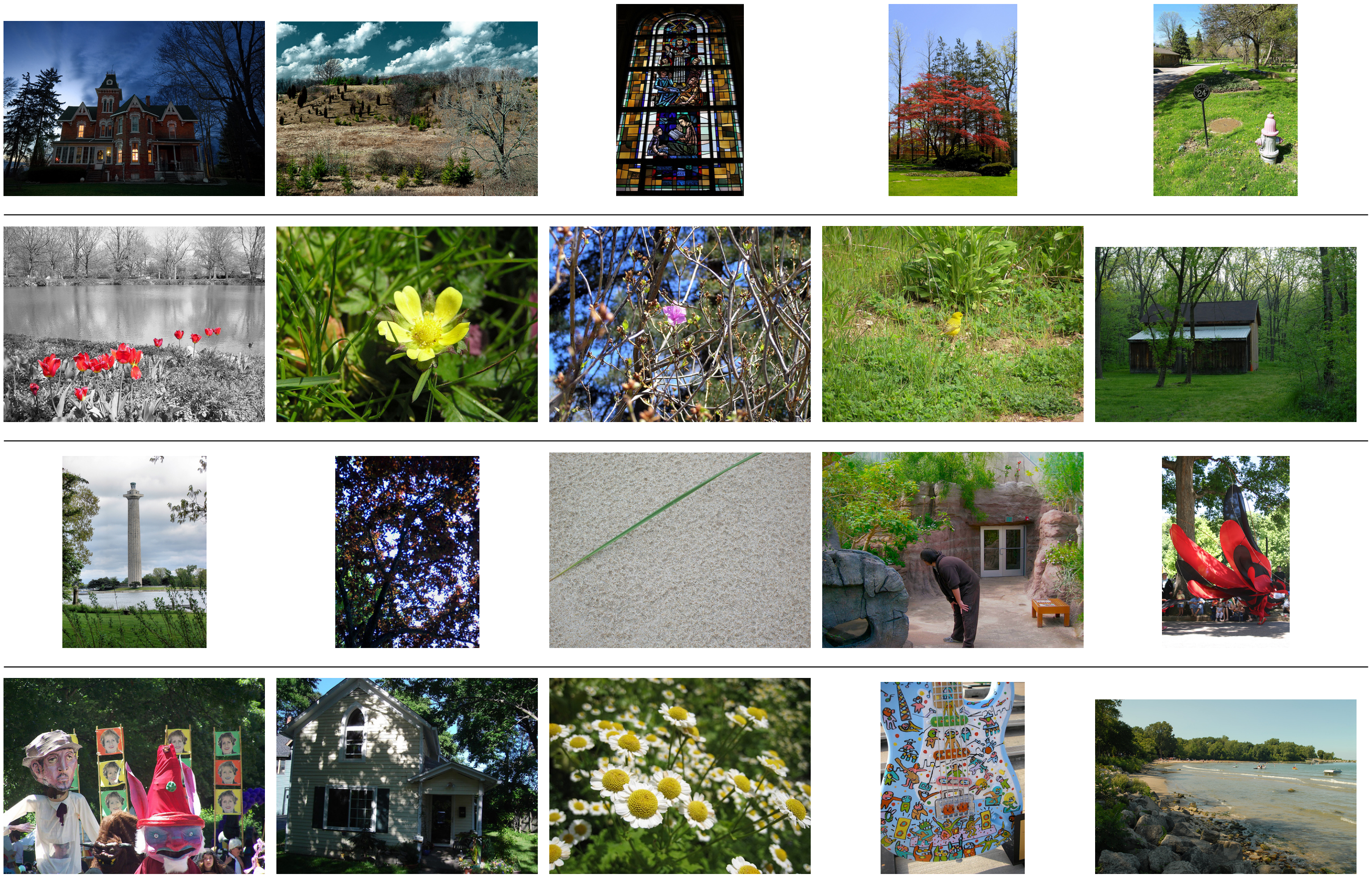


Fig. 10. In vegetation detection over North America in 2009 and 2010, among all false positive bins, there are ... images that are predicted as greenery. And these images are the reason these bins are predicted as green. Here are some random selected examples of the green images.

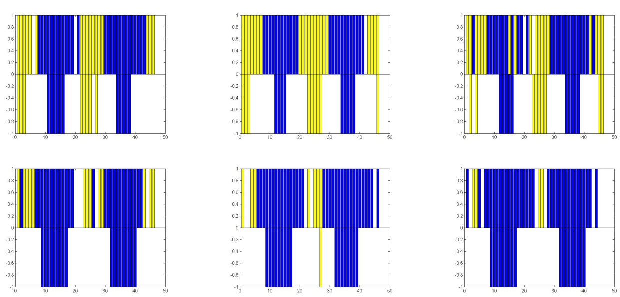


Fig. 11. Yellow bars show non-greenery at that time. Blue bars represent greenery. Prediction results on top shows 6 random places comparing to satellite ground truth. The ground truth on the bottom tends to disappear when leaves are turning yellow or green.

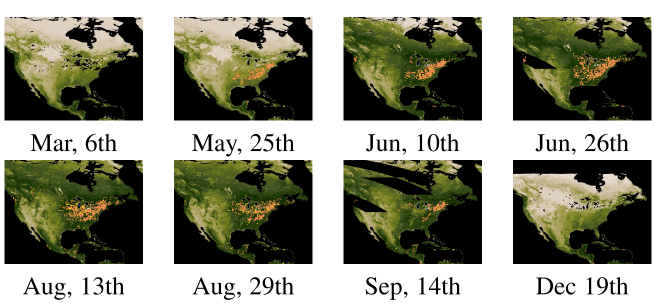


Fig. 12. We use prediction results to recreate vegetation coverage maps for each 16-days period. There are 8 maps picked in 2010. The dates under each map are the starting date of each 16-days period.

events and trends through online social media. Our study shows that snowy scene recognition is not nearly as easy a problem as one might expect, when applied to realistic consumer images; our best result using modern vision techniques gives 81% accuracy. Nevertheless, as a proof-of-concept we demonstrated that this recognition accuracy still yields a reasonable map that approximates observations from satellites. We also test recognition algorithms on their ability to recognize a particular species of flower, the California Poppy. In future work, we plan to combine evidence from tags and other metadata with visual features for more accurate estimates, and to develop novel techniques for these challenging recognition problems. More generally, we hope the idea of observing nature through photo-sharing websites will help spark renewed interest in recognizing natural and ecological phenomenon in consumer images.

In this paper, we propose using the massive collections of user-generated photos uploaded to social sharing websites as a source of observational evidence about the world, and in particular as a way of estimating the presence of ecological phenomena. As a first step towards this long-term goal, we used a collection of 150 million geo-tagged, timestamped photos from Flickr to estimate snow cover and greenery, and compared these estimates to fine-grained ground truth collected by earth-observing satellites and ground stations. We compared several techniques for performing the estimation from noisy, biased data, including simple voting mechanisms and a Bayesian likelihood ratio. We also tested several possible improvements to these basic methods, including using temporal smoothing and machine learning to improve the accuracy of estimates. We found that while the recall is relatively low due to the sparsity of photos on any given day, the precision can be quite high, suggesting that mining from photo sharing websites could be a reliable source of observational data for ecological and other scientific research. In future work, we plan to study additional features including using more sophisticated computer vision techniques to analyze visual content. Also we plan to study a variety of other ecological phenomena, including those for which high quality ground truth is not available, such as migration patterns of wildlife and the distributions of blooming flowers.

Appendix A
Proof of the First Zonklar Equation

Appendix one text goes here.

Appendix B

Appendix two text goes here.

Acknowledgments

The authors would like to thank...

References

[1] "The great sunflower project," http://www.greatsunflower.org.
[2] D. Fink, T. Damoulas, and J. Dave, "Adaptive spatio-temporal exploratory models: Hemisphere-wide species distributions from massively crowdsourced eBird data," in *AAAI*, 2013.
[3] J. King, A. Cabrera, and R. Kelly, "The Snowtweets Project: Communicating snow depth measurements from specialists and non-specialists via mobile communication technologies and social networks," in *AGU Fall Meeting*, 2009.
[4] X. Jin, A. Gallagher, L. Cao, J. Luo, and J. Han, "The wisdom of social multimedia: Using Flickr for prediction and forecast," in *ACM Multimedia*, 2010.
[5] J. DiGrazia, K. McKelvey, J. Bollen, and F. Rojas, "More tweets, more votes: Social media as a quantitative indicator of political behavior," *PLOS One*, 2013.
[6] J. Bollen, H. Mao, and X.-J. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.
[7] J. Ginsberg, M. Mohebbi, R. Patel, L. Brammer, M. Smolinski, and L. Brilliant, "Detecting influenza epidemics using search engine query data," *Nature*, vol. 457, pp. 1012–1014, 2009.
[8] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: real-time event detection by social sensors," in *WWW*, 2010.
[9] O. Hyvarinen and E. Saltikoff, "Social media as a source of meteorological observations," *Monthly Weather Review*, vol. 138, no. 8, pp. 3175–3184, 2010.
[10] H. Zhang, M. Korayem, D. Crandall, and G. LeBuhn, "Mining Photo-sharing Websites to Study Ecological Phenomena," in *WWW*, 2012.
[11] 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.
[12] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, "SUN database: Large-scale scene recognition from abbey to zoo," in *CVPR*, 2010.
[13] L.-J. Li and L. Fei-Fei, "What, where and who? classifying events by scene and object recognition," in *ICCV*, 2007.
[14] L.-J. Li, R. Socher, and L. Fei-Fei, "Towards total scene understanding: classification, annotation and segmentation in an automatic framework." in *CVPR*, 2009.
[15] J. Luo, A. Singhal, and W. Zhu, "Natural object detection in outdoor scenes based on probabilistic spatial context models," in *ICME*, 2003.
[16] M. Boutell, A. Choudhury, J. Luo, and C. M. Brown, "Using semantic features for scene classification: How good do they need to be?" in *ICME*, 2006.
[17] K. Kremerskothen.
[18] B. De Longueville, R. S. Smith, and G. Luraschi, ""OMG, from here, i can see the flames!"," in *Proc. Intl. Workshop on Location Based Social Networks*, 2009, pp. 73–80.
[19] B. O'Connor, R. Balasubramanyan, B. Routedge, and N. Smith, "From tweets to polls: Linking text sentiment to public opinion time series," in *ICWSM*, 2010.
[20] M. L. Parry, O. F. Canziani, J. P. Palutikof, P. J. van der Linden, and C. E. Hanson, *IPCC, 2007: Climate Change 2007: Impacts, Adaptation, and Vulnerability*. Cambridge University Press, 2007.
[21] 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.
[22] D. Lazer et al, "Life in the network: the coming age of computational social science," *Science*, vol. 323, no. 5915, pp. 721–723, 2009.
[23] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri, "Feedback effects between similarity and social influence in online communities," in *KDD*, 2008.
[24] D. Liben-Nowell and J. Kleinberg, "Tracing information flow on a global scale using internet chain-letter data," *PNAS*.
[25] D. Brockmann, L. Hufnagel, and T. Geisel, "The scaling laws of human travel," *Nature*, vol. 439, pp. 462–465, 2006.
[26] A. Anagnostopoulos, R. Kumar, and M. Mahdian, "Influence and correlation in social networks," in *KDD*, 2008.
[27] C. Aggarwal and T. Abdelzaher, "Social Sensing," in *Managing and Mining Sensor Data*. Springer, 2013.
[28] S. A. Golder and M. Macy, "Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures," *Science*, vol. 333, no. 6051, pp. 1878–1881, Sep. 2011.
[29] A. Sadilek, H. Kautz, and V. Silenzio, "Predicting Disease Transmission from Geo-Tagged Micro-Blog Data," in *AAAI*, 2012.
[30] V. K. Singh, M. Gao, and R. Jain, "Social pixels: genesis and evaluation," in *ACM Multimedia*, 2010.
[31] V. Singh, S. Venkatesha, and A. K. Singh, "Geo-clustering of images with missing geotags," in *Granular Computing (GrC), 2010 IEEE International Conference on*. IEEE, 2010, pp. 420–425.
[32] D. J. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg, "Mapping the world's photos," in *Proceedings of the 18th international conference on World wide web*. ACM, 2009, pp. 761–770.
[33] B. Thomee, J. G. Moreno, and D. A. Shamma, "Who's time is it anyway?: Investigating the accuracy of camera timestamps," in *Proceedings of the 22nd ACM International Conference on Multimedia*, 2014.
[34] C. Hauff and G.-J. Houben, "Geo-location estimation of flickr images: social web based enrichment," in *Advances in Information Retrieval*. Springer, 2012, pp. 85–96.
[35] D. Leung and S. Newsam, "Proximate Sensing: Inferring What-Is-Where From Georeferenced Photo Collections," in *CVPR*, 2010.
[36] C. Murdock, N. Jacobs, and R. Pless, "Webcam2Satellite: Estimating cloud maps from webcam imagery," in *WACV*, 2013.
[37] A. Singhal, J. Luo, and W. Zhu, "Probabilistic spatial context models for scene content understanding," in *CVPR*, 2003.
[38] M. Boutell, "Exploiting context for semantic scene classification," Ph.D. dissertation, University of Rochester Department of Computer Science, 2005.
[39] P. Barnum, S. Narasimhan, and T. Kanade, "Analysis of rain and snow in frequency space," *IJCV*, vol. 86, no. 2-3, 2009.
[40] N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. C. Lopez, and J. V. Soares, "Leafsnap: A computer vision system for automatic plant species identification," in *Computer Vision–ECCV 2012*. Springer, 2012, pp. 502–516.
[41] C. Siagian and L. Itti, "Comparison of gist models in rapid scene categorization tasks," *Journal of Vision*, vol. 8, no. 6, pp. 734–734, 2008.
[42] P. G. Foschi and D. K. Smith, "Detecting subpixel woody vegetation in digital imagery using two artificial intelligence approaches," *Photogrammetric Engineering and Remote Sensing*, vol. 63, no. 5, pp. 493–499, 1997.
[43] P. Balamurugan and R. Rajesh, "Greenery image and non-greenery image classification using adaptive neuro-fuzzy inference system," in *Conference on Computational Intelligence and Multimedia Applications, 2007. International Conference on*, vol. 3. IEEE, 2007, pp. 431–435.
[44] "Lost ladybug project," http://www.lostladybug.org.
[45] http://www.ncdc.noaa.gov/oa/climate/ghcn-daily/.
[46] J. Hays and A. A. Efros, "IM2GPS: Estimating geographic information from a single image," in *CVPR*, 2008.
[47] A. Torralba, R. Fergus, and W. T. Freeman, "80 million tiny images: A large data set for nonparametric object and scene recognition," *PAMI*, vol. 30, no. 11, pp. 1958–1970, 2008.

[48] M. Korayem, A. Mohamed, D. Crandall, and R. Yampolskiy, "Solving avatar captchas automatically," in *AMLTA*, 2012.

[49] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *IJCV*, vol. 42, no. 3, pp. 145–175, 2001.

[50] M. Douze, H. Jegou, H. Sandhawalia, L. Amsaleg, and C. Schmid, "Evaluation of gist descriptors for web-scale image search," in *ICIVR*, 2009.

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

[52] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," 2014.

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

[54] I. Vergara, T. Norambuena, E. Ferrada, A. Slater, and F. Melo, "StAR: a simple tool for the statistical comparison of ROC curves," *BMC Bioinformatics*, vol. 9, no. 1, p. 265, 2008.