Semantic Topic Analysis of Traffic Camera Images

Jeffrey Liu  
Civil and Environmental Engineering  
Massachusetts Institute of Technology  
Cambridge, MA 02139–4301  
jeffliu@mit.edu

Andrew Weinert  
Lincoln Laboratory  
Massachusetts Institute of Technology  
Lexington, MA 02421–642  
andrew.weinert@ll.mit.edu

Saurabh Amin  
Civil and Environmental Engineering  
Massachusetts Institute of Technology  
Cambridge, MA 02139–4301  
amins@mit.edu

Abstract—Traffic cameras are commonly deployed monitoring components in road infrastructure networks, providing operators visual information about conditions at critical points in the network. However, human observers are often limited in their ability to process simultaneous information sources. Recent advancements in computer vision, driven by deep learning methods, have enabled general object recognition, unlocking opportunities for camera-based sensing beyond the existing human observer paradigm. In this paper, we present a Natural Language Processing-inspired approach, entitled Bag-of-Label-Words (BoLW), for analyzing image data sets using exclusively textual labels. The BoLW model represents the data in a conventional matrix form, enabling data compression and decomposition techniques, while preserving semantic interpretability. We apply the Latent Dirichlet Allocation topic model to decompose the label data into a small number of semantic topics. To illustrate our approach, we use freeway camera images collected from the Boston area between December 2017–January 2018. We analyze the cameras' sensitivity to weather events; identify temporal traffic patterns; and analyze the impact of infrequent events, such as the winter holidays and the “bomb cyclone” winter storm. This study demonstrates the flexibility of our approach, which allows us to analyze weather events and freeway traffic using only traffic camera image labels.

I. INTRODUCTION

Monitoring the traffic state and driving conditions of a road infrastructure network is crucial for efficient operations. While embedded sensors such as loop detectors can give precise measurements of specific quantities, it is often impractical to install specific sensors for every quantity of interest. Thus, many transportation agencies have integrated cameras into their monitoring solutions. For example, one service provider, TrafficLand, operates over 18,000 cameras across over 200 cities in the United States [1]. Cameras offer several benefits: they are general purpose sensors that can detect multiple quantities in a region of space; they can be installed on existing physical and communications infrastructure; and human operators intuitively understand images. However, since humans are limited in their ability to process visual information from simultaneous information sources [2], it is unreliable to depend on human operators to constantly monitor all of the cameras.

While computers excel at processing large quantities of structured data, images are unstructured, and have been historically difficult for machines to parse. Prior attempts to explicitly parameterize object detection for traffic applications faced difficulties dealing with the separation of foreground and background elements, the occlusion of objects, variability in lighting conditions, and computational costs [3][4]. As a result, many of the successful applications of traffic cameras for automatic sensing have been limited to narrowly-defined problems, such as automatic license plate detection [4]. The use of cameras in these applications more closely resemble single-purpose sensors, as they only detect a specific quantity—e.g. license plate numbers. Thus, outside of these narrow applications and direct observation by operators, the potential of cameras for automatic, general-purpose, sensing in traffic applications has not yet been fully realized.

Recent advancements in computational power and deep learning methods, particularly deep convolutional neural networks (CNNs), have driven remarkable progress in recognizing objects in images [5]. These techniques infer rules for object detection based on large training data sets of labeled example images [6]. Once trained, predictions using CNNs are quick to perform and can be done even on mobile devices [7]. However, training neural networks is computationally expensive, and requires large quantities of labeled training data [8]. Thus, within the last few years, technology companies have begun to offer access to pre-trained image recognition algorithms as commercial services [9]. These services allow developers to quickly obtain labels in plain English for any image, without needing to build and train their own image recognition system.

Although there is significant attention currently focused on improving the performance of deep learning algorithms [5], our focus is on exploring new applications enabled by these technologies. To this end, we pose the motivating question: what operationally-relevant information about phenomena in a traffic network can be obtained using only the labels of traffic camera images? We illustrate the question with a thought experiment: imagine a blindfolded observer who is continuously told verbally whether the camera currently shows an item from a finite list of recognized objects; no additional information is given about the number, location, nor appearance of the objects in the frame, only simply whether or not it appears in the image. What can the blindfolded observer infer about the network state? In particular: can the blindfolded
To address these questions, we present in Sec. III an approach to analyzing sets of images using only the textual labels from an image recognition service. By treating sets of image labels as “documents” describing the images, we pose the problem as a Natural Language Processing (NLP) problem of analyzing a corpus of texts. We introduce the Bag-of-Label-Words (BoLW) model, inspired by the popular Bag-of-Words (BoW) model [10], which represents the image content labels in a conventional matrix structure, allowing for data compression and dimension reduction operations. We demonstrate the application of Latent Dirichlet Allocation (LDA) [11], a hierarchical Bayesian model of topics within text corpora, to decompose the label data into high-level semantic topics. We present a case study based on traffic camera images from the Boston area, described in Sec. II Sec. V describes the case study results regarding the detection of weather events, and the impact on weekly traffic patterns caused by the winter holidays and the “bomb cyclone” storm. Sec. V concludes and presents directions for future work.

II. TRAFFIC CAMERA DATA

A. Cameras and Images

We illustrate our approach using data from cameras in the Boston area. The data consists of images collected from seven freeway traffic cameras operated by the Massachusetts Department of Transportation (MassDOT). We regularly scraped the public Mass511 Traveler Information Service website [12] between December 17, 2017–January 31, 2018 to build a data set of 189,498 images, each with a resolution of 320 × 240 pixels, and an average sampling period of 3 minutes for each camera. Notable events during the collection period include the winter holidays and the “bomb cyclone” East Coast blizzard. Details for each camera, including their MassDOT-assigned identification number and name, locations, and sample images, are provided in Figure 1.

B. Labels

We obtained labels for the traffic camera images using the Google Cloud Vision (GCV) image recognition platform [1] in particular, we used the “label detection” service—referred to as Labeling Service 1 (LS1)—and the “web entity detection” service (resp. LS2). The label detection annotation service (LS1) provides annotations for “broad sets of categories within an image, ranging from modes of transportation to animals,” whereas the web detection service (LS2), provides more detailed information from the web, such as related websites and associated “web entities” [13]—elements of the Google Knowledge Graph ontology representing real-world entities and concepts [14]. We use the names of these related web entities as a second source of image labels. We define the term “vocabulary” of a service to represent the set of all labels that

Note that the approach in this paper is not specific to the GCV services, and any image labeling solution can be used in its place.

| ID   | Name                                           |
|------|------------------------------------------------|
| 1106-1 | I-93-ME-Boston-@ exit to HOV-E                 |
| 1137-1 | I-93-NB-Charlestown-@ Zakim South Tbr         |
| 1296-1 | Ramp 1-EB-TNL-ramp end 93N x20 c              |
| 1413-1 | Ramp K-NB-Boston-93N x20 b                     |
| 1500-1 | Ramp K-NB-Boston-93N x20 a                     |
| 1508-1 | Ramp CC-EB-Boston-90E x24C to 93S e           |
| 1600-1 | Road OHWY-SB-Boston-Leverett Circle            |

Fig. 1. Camera details. We selected a diverse set of cameras which depicted several different network locations and components, including a bridge (1137–1), underpass (1508–1), intersection (1600–1), HOV lane (1106–1), median (1296–1), and open freeway (1413–1, 1500–1) the particular service can return. In this paper, we consider a unified vocabulary, constructed from the disjoint union of the LS1 and LS2 vocabularies.

Labels and their respective scores for a sample image are given in Fig. 2. Note that the labels from LS1 are reported by the service in lowercase, and those from LS2 are reported in Title Case; when necessary, we may also prepend the originating service to the label to further distinguish labels from the respective service, e.g. “LS1: car” vs. “LS2: Car”. The score of each label corresponds to the confidence level reported by GCV. The label scores from LS1 are reported by GCV on a normalized scale, with a maximum value of 1 and truncated at 0.5, i.e. no labels are returned by the service whose scores is less than 0.5. The scores returned from LS2 are not normalized by GCV, and the documentation warns that the scores should not be compared between labels nor images [13]. As such, we discard all score information and binarize the data by setting all nonzero scores to 1. We demonstrate that even with such an aggressively data processing approach, we are still able to clearly identify topics and phenomena.

In general, the LS2 vocabulary includes more specific terms than the LS1 vocabulary; for example, we observed the labels “LS2: BMW,” “LS2: BMW 3 Series,” and “LS2: 2018 BMW 3 Series Sedan” in the vocabulary, whereas we found only the label “LS1: bmw” in the LS1 vocabulary. However, the LS2 was also prone to including more spurious labels; for

observer distinguish between a new instance of a phenomenon versus the persistence of an existing one, and can they discern differences in magnitude between phenomena?
III. METHODS

This section presents the methods used in our analysis. We formulate the Bag-of-Label-Words (BoLW) model in Sec. III-A, introduce the per-camera idf data weighting scheme in Sec. III-B, construct the image-label matrix representation in Sec. III-C, and describe the LDA topic model in Sec. III-D.

A. Bag-of-Label-Words

Consider a vector space, $\mathcal{L}$, where each dimension corresponds to an individual label in the label vocabulary. The dimension of $\mathcal{L}$—the number of terms in the label vocabulary—is denoted $M$. Any weighted list of labels can be represented as a vector in this label vector space, denoted $\ell \in \mathcal{L}$, where the nonzero components of $\ell$ correspond to the weight of the respective label in the list. Let $j$ index the vocabulary, and thus, let $\ell^j$ represent the component of the $\ell$ corresponding to term $j$. This vector representation is analogous to the BoW vector space model of documents in NLP, which represents documents as vectors where each component corresponds to the number of occurrences of a given word in the document $[10]$; hence, we refer to our model as BoLW.

In our model, an image, $I_i$, in a corpus, $\mathcal{I}$, is represented as the tuple $I_i = (c_i, \tau_i, \ell_i)$, where:
- $i \in \{1, \ldots, N\}$ indexes the image
- $c_i$ denotes its originating camera
- $\tau_i$ denotes its timestamp
- $\ell_i$ is its Bag-of-Label-Words label vector

The Bag-of-Label-Words vector model is defined as follows:
- A label word, $\lambda_j$, is defined as a single label in the label vocabulary, indexed by $j \in \{1, \ldots, M\}$. $\lambda_j$ is a unit-basis vector in $\mathcal{L}$ whose $j$th component equals one, and all other components equal zero.
- A bag of label words associated with image $i$ is a vector $\ell_i \in \mathcal{L}$. Any weighted list of labels describing an image $i$ can be represented as a bag of label words by setting the respective weights of $\ell_i$ equal to the weights of the corresponding list elements.
  - The total weight, $w_i$, of bag $\ell_i$ is defined as its $L^1$-norm:
    $$ w_i := \|\ell_i\|_1 = \sum_j |\ell_i^j| $$

While this paper focuses on the application of the BoLW model for traffic camera image analysis, our approach can be generically applied to any corpus of discretely labeled images.

B. Label Reweighting

We note that extremely common labels do not necessarily contribute much operationally useful information about the image contents. For example, labels such as “Road” and “Asphalt” appear extremely frequently in the Boston data set. While these labels are not incorrect—the images from freeway cameras do indeed contain roads made of asphalt—they are also not particularly informative, as we expect that most images from a traffic camera contain a road. Thus, we would like to attenuate the weight of labels which occur extremely frequently. We address this with the Term Frequency-Inverse

---

**Table:**

| LS1 label | Score |
|-----------|-------|
| snow      | 0.94  |
| infrastructure | 0.87  |
| mode of transport | 0.87  |
| lane      | 0.84  |
| winter storm | 0.84  |
| road      | 0.83  |
| transport | 0.82  |
| structure | 0.77  |
| phenomenon | 0.75  |
| blizzard | 0.7   |
| highway   | 0.71  |
| freezing  | 0.65  |
| automotive exterior | 0.57  |
| glass     | 0.55  |

| LS2 label | Score |
|-----------|-------|
| Blizzard | 1.21  |
| Lane      | 1.10  |
| Car       | 1.0   |
| Transport | 1.03  |
| Snow      | 0.75  |
| Highway   | 0.75  |
| Fog       | 0.72  |
| Glass     | 0.67  |
| Freezing  | 0.62  |
| Massachusetts Department of Transportation | 0.32 |

---

Fig. 2. An image taken during the “bomb cyclone” (a) and the labels and scores returned by the labeling services (b).
Document Frequency (tf-idf) weighting scheme, described below, to rescale each image’s label weights based on each label’s rarity for each camera.

The tf-idf weighting scheme is a heuristic used in NLP to reweight terms in the BoW vector to account for the natural difference in term prevalence in a language [10]. Terms that are commonly used in a language will highly represented in any given document, simply due to their prevalence in the language, regardless of their relevance to the subject matter of the document. These extremely common terms can end up dominating the weight of a bag, and thus, it may be desirable to attenuate them. The tf-idf weight is computed as the product of its two titular components: the term frequency (tf) and the inverse document frequency (idf) [10]. The term frequency of a given document I and term j corresponds to the number of occurrences of the term within the document; in our case, our term frequency for image I and label j is given simply by the binary variable:

$$\text{tf}(i, j) = \begin{cases} 1 & \text{if image } i \text{ has label } j \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The inverse document frequency (idf) of a term j is typically computed as the negative logarithm the empirical document frequency: $\text{idf}(j) = -\log(f_j) = \log \left( \frac{N}{n_j} \right)$. We use a variant of idf, which we term the per-camera idf, computed as:

$$\text{idf}(j, c) = \log \left( \frac{N_c}{n_{j,c}} \right) \quad (3)$$

where $N_c$ is the total number of images for camera c. The per-camera idf considers the relative rarity of a label j within the context of the other images from that camera. This is motivated by the fact that the label distributions are different across cameras; for example, the presence of the label “Snow” is more unusual and notable for images from a camera in a tunnel than those from a camera out in the open.

C. Image-label matrix

We construct the $N \times M$ image-label matrix, denoted $\Lambda$, by vertically concatenating row vectors $\ell_i$ for all images $i \in \{1 \ldots N\}$, where the components of $\ell_i$ are the per-camera tf-idf values, given by:

$$\ell_i = \text{tf}(i, j) \times \text{idf}(j, c_i). \quad (4)$$

Each row of $\Lambda$ corresponds to an image, and each column correspond to a label. This image-label matrix is analogous to the document-term matrix in NLP, and in general, our usage of the terms “image” and “label” in this paper correspond to “document” and “term” respectively in the NLP literature.

The image-label matrix representation of the data set presents the data in a familiarly-structured form: a matrix of $N$ observations of an $M$-dimensional system. Thus, the image labeling process transforms the unstructured traffic camera data into a conventional, structured $M$-dimensional time series analysis problem. However, as a label vocabulary can span hundreds to thousands of words, the dimension of $\mathcal{L}$ may still be prohibitively large for human interpretation and certain computations. Since $\Lambda$ is simply a $N \times M$ matrix, many conventional matrix compression techniques can be applied to the data to reduce its dimensionality. However, depending on the technique, the compressed representation may not be semantically meaningful; this ends up forgoing much of the advantages in interpretability by representing the data as labels in the first place. Thus, we focus on topic models, which decompose the data into semantically distinct and interpretable topics; in particular, the Latent Dirichlet Allocation topic model.

D. Topic Identification via Latent Dirichlet Allocation

Algorithm 1: LDA BoLW generation procedure

```
input : Target weight, $\bar{w}_i$, of bag $\ell_i$
        image-topic prior hyperparameter $\alpha$
        topic-label prior hyperparameter $\beta$

output: bag of label words $\ell_i$

Initialize $\ell_i = 0$
Draw $\theta \sim \text{Dirichlet}(\alpha)$
Draw $\phi \sim \text{Dirichlet}(\beta)$
while $w_i < \bar{w}_i$ do
    Draw topic $z \sim \text{Multinomial}(\theta^z)$
    Draw a label word $\lambda \sim \text{Multinomial}(\phi^z)$
    Add $\lambda$ to the bag: $\ell_i = \ell_i + \lambda$
end
```

Latent Dirichlet Allocation (LDA) is a hierarchical Bayesian topic model for document generation in NLP, first posed by [Blei et al.][11]. LDA represents documents as random mixtures of topics, denoted $\theta$, where each topic is, in turn, a probability distribution over words, denoted $\phi$. Griffiths and Steyvers present a variant that includes an additional Dirichlet prior over the topic-word distribution $\phi$ [15]. We adapt this variant of LDA to BoLWs below, and visualize it in plate notation in Figure 3. Each bag of label words $\ell_i$ in a corpus $\mathcal{L}$ is generated by the LDA model with the procedure described in Algorithm 1. The target weights $\bar{w}_i$ are set exogenously based on the empirical bag weights in the corpus; in addition, the target weights are rounded to the nearest integer, since the model requires an integer number of copies of the innermost plate in Fig. 3. The image-topic distribution is denoted $\theta$ and characterized by the $K$-dimensional hyperparameter $\alpha$, and $\theta^z = \theta(z|i)$. A topic is denoted $z \in \{1, \ldots, K\}$, where $K$ is the total number of topics, set exogenously. The topic-label distribution is denoted $\phi$ and characterized by the $M$-
The distribution of labels for a given topic is denoted \( \phi^\tau = \phi(\lambda|z) \).

Given the hyperparameters and a corpus of data \( I \), we would like to infer the most likely image-topic distribution, \( \theta \) and topic-word distribution, \( \phi \), which maximizes the posterior probability \( p(\theta, \phi|I, \alpha, \beta) \). We achieve this with the online variational Bayes algorithm presented in [16]. We assume symmetric priors on \( \theta \) and \( \phi \) with constant hyperparameter values \( \alpha = \frac{50}{K} \) and \( \beta = 0.1 \) based on [13]. For a detailed treatment of LDA and other probabilistic topic models, we recommend [17].

IV. RESULTS

A. Topics

In this section, we discuss the LDA topics, and present the top labels for a selection of representative topics. For the LDA decomposition in this example, we chose \( K \) qualitatively; we wanted the number of topics to be small enough to be easily visualized and interpreted, but large enough to distinguish between the following semantic categories in the data: i.) snow storms; ii.) the day/night cycle; iii.) traffic; iv.) physical infrastructure; and v.) error messages. For illustrative purposes, we found that \( K = 10 \) provides a manageable number of semantically distinct and relevant topics. In practice, the selection of the optimal number of topics, \( K \), depends on the target application; in [13], Griffiths and Steyvers propose selecting \( K \) which maximizes the log-likelihood of the data.

We present in Table 4 five representative topics—one for each of the aforementioned categories—along with their respective highest-probability labels. We have followed common practice and manually named topics—a posteriori—based on domain-specific knowledge of the data, for ease of reference. Note that Topic 10: “Error” is a special edge case; the topic appears only on the image depicted in Figure 46, which is returned when the web service is unable to provide the current feed for a given camera. This allows us to conveniently identify error images in the data; this is notable, as the LDA is not explicitly trained to identify error images. This suggests that the LDA decomposition may be a suitable transformation for image classification task, which we will explore in future work.

B. Topic time-series

The LDA image-topic distributions \( \theta^i \) can be interpreted as projections of the BoLW vector, \( \ell_i \), onto the LDA topic space. By plotting the probability of each topic for an image \( \theta^i(z) \) at its timestamp \( \tau_i \), we can construct a set of time-series for each topic and camera. Figures in this section were plotted with the data averaged into 15-minute bins. We examine the detection of weather events in Section IV-B1 and the effect of infrequent events on weekly traffic patterns in Section IV-B2.

1) Detecting weather events

We identified notable weather events during the data collection period via the NOAA Global Historical Climatology Network (NOAA-GHCN) data set [18]. We define a notable weather event as one with at least 1” of snow or rainfall. We identified six such events during the collection period: two with snow and rain on 2017-12-25 and 2018-01-04 (the “bomb cyclone”); two with only rain on 2018-01-13 and 2018-01-23; and two with only snow on 2018-01-17 and 2018-01-30.

We first consider the simple approach of using a single label: “LS1: snow” to detect weather events. Its time series—the column of \( \Lambda \) corresponding to “LS1: snow”—is plotted for each camera in Figure 5a. For most cameras, the “LS1: snow” time series is sensitive to the weather events that include snow. The exception is camera 1508–1; however, this is to be expected, as this camera is located in a underpass that is isolated from the elements (see Fig.1). We note two issues: first, the label is not sensitive to weather events that are exclusively rain (only two cameras, 1413–1 and 1600–1, have any signal during the 2018-01-23 rain event). This, however, can be overcome by manually including additional labels, such as the “rain” LS1 and LS2 labels. Second, there are significant daily recurrences of the “LS1: snow” label in many cameras following the weather events of 2017-12-25 and 2018-01-04. The recurrence corresponds to the detection of accumulated snow on the ground. The recurrence can make it difficult to differentiate new snowfall from recurring detection of existing snow (cf. cameras 1296–1, 1413–1, 1500–1). In addition, while the snowfall during the “bomb cyclone” event was much greater than the 2017-12-25 event, this is not directly apparent in the label time series.

We now demonstrate how LDA ameliorates both of the above issues. Figure 5b illustrates the time series for Topic 1: “Wintry conditions” for each camera. These time series are still sensitive to the snowfall events (except camera 1508–1, as expected), but unlike the “LS1: snow” time series, they also show small signals in response to the rainfall events. In addition, we observe that the effect of the daily detection of accumulated snow is significantly attenuated compared to the signal of the “LS1: snow” time series. Cameras 1296–1 and 1600–1 still exhibit the daily recurrence trait—albeit to a lesser degree—particularly after the large “bomb cyclone” blizzard. These two cameras are angled closer to the road than the other outdoor cameras, and thus they see more of the ground. As such, we expect the effect of accumulated snow to be more pronounced for these two cameras. Additionally, whereas the “LS1: snow” time series did not clearly show any difference in magnitude in the label recurrence between the 2017-12-25 and 2018-01-04 events; the magnitude of the “Wintry conditions” topic recurrence following the “bomb cyclone” 2018-01-04 storm is larger, consistent with the storm’s greater snowfall.

2) Weekly traffic patterns and disturbances

In this section, we analyze the weekly daytime traffic patterns using the time series of Topic 4: “Car” from camera 1508–1. This camera is selected due to its location in an underpass, which isolates it from the direct effects of weather; in addition, we did not observe any camera angle changes during the observation period. Its isolation from weather and static camera angle ensures that any changes in the traffic pattern are due to changes in demand, and not a side effect of reduced visibility or change in camera angles.
We illustrate in Figure 6 the weekly time series of the “Car” topic for camera 1508–1. The light grey lines represent the data for all weeks, whereas the highlighted lines in Figure 6a and 6b highlight the weeks of the Christmas holiday, and the “bomb cyclone” storm respectively. We observe the weekday-weekend traffic pattern of greater demand during the weekdays than on the weekends. Additionally, we observe a larger peak during the evening rush hours on weekdays. This is consistent with the camera’s location on an on-ramp to I-93 South, leading out of Boston (Fig. 1). Indeed, though the labels do not explicitly encode the car counts, we demonstrate that some magnitude information can be inferred binarized label data.

In Figure 6, we observe that on Christmas, which occurred on a Monday, the “Car” topic was significantly lower than on other weeks. The rest of the week, however, was not significantly different than average. This is consistent with a reduction in traffic on Christmas day, as it is a national holiday, and many institutions and businesses are closed. We also note that while New Year’s day (Monday, Figure 6b) is also a national holiday, we observe more traffic on New Year’s than on Christmas. This is likely because in the US, many businesses are closed on Christmas but are open on New Year’s, albeit with limited hours.

We observe the effect of the bomb cyclone in Figure 6b. Whereas the snow did not start falling until the evening of January 4th, the observed counts are at near zero starting in the morning, and remain there until Friday evening. Even then, Friday evening and Saturday have lower readings than usual; only on Sunday do things return to normal. The low “Car” topic weights correspond with the City’s declaration of Snow Emergency and Parking Ban, which was in effect between 7 a.m., January 4th–5 p.m., January 5th.

V. CONCLUSION

In summary, we presented: the Bag-of-Label-Words vector model for representing images in a semantic label space; the application of the LDA topic model as a dimensionality reduction tool; and an analysis of freeway traffic cameras in the Boston metropolitan area using these techniques. We are able to distinguish between new snowfall and accumulated snow, and also to capture relative changes in magnitude in the example data set, despite using only binarized label data, which has potential application in data compression and privacy contexts.

We observe that disruptions, such as storms, manifest clearly in the topic time series. This offers a simple approach to performing change and anomaly detection on image data. Whereas most image change detection algorithms operate in the high-dimensional pixel space [20], our approach uses CNNs to transform the problem into the semantic space, where the BoLW representations and LDA dimension reductions allow for conventional univariate and multivariate algorithms to analyze the changes. Indeed, as traffic is an inherently dynamic phenomenon, pixels will always be changing due to passing vehicles. By representing the images in the semantic space, we can analyze phenomena that are difficult to parameterize in pixel space. We are currently working on automatic detection of anomalous traffic patterns using these techniques.

Finally, we highlight the general applicability of our approach; we use the same methodology to sense both weather events and traffic patterns. This is a step toward to realizing the potential of general purpose sensing with cameras. While our constraint of using only binarized label data was intentionally restrictive to study the information content of labels on their own, we are exploring the use of traffic camera label data in conjunction with other detectors, such as loop detectors, in sensor fusion applications.

REFERENCES

[1] TrafficLand, “Traffic Cameras, Traffic Video, Live Traffic Cams,” 2014. [Online]. Available: [https://perma.cc/6YDC-UM8J](https://perma.cc/6YDC-UM8J)

[2] R. Marois and J. Ivanoff, “Capacity limits of information processing in the brain,” Trends in Cognitive Sciences, vol. 9, no. 6, pp. 296–305, Jun 2005.

[3] V. Kastrinaki, M. Zervakis, and K. Kalaitzakis, “A survey of video processing techniques for traffic applications,” Image and Vision Computing, vol. 21, no. 4, pp. 359–381, 2003.

[4] N. Buch, S. A. Velastin, and J. Orwell, “A review of computer vision techniques for the analysis of urban traffic,” IEEE Transactions on Intelligent Transportation Systems, vol. 12, no. 3, pp. 920–939, 2011.

[5] Y. A. LeCun, Y. Bengio, and G. E. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, May 2015.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, 2012, pp. 1097–1105.

[7] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard
Fig. 5. Comparison of the sensitivity of the “LS1: snow” label (a) and Topic 1: “Wintry conditions” (b) to winter storms.

et al., “Tensorflow: A system for large-scale machine learning,” in *OSDI*, vol. 16, 2016, pp. 265–283.

Fig. 6. The “Car” topic of camera 1508–1 captures the weekday-weekend traffic demand cycle. The time series signal shows lower readings for the days corresponding to the Christmas holiday [6a] and the “Bomb Cyclone” winter storm [6b].

[8] R. Livni, S. Shalev-Shwartz, and O. Shamir, “On the Computational Efficiency of Training Neural Networks,” *NIPS*, pp. 1–15, 2014.

[9] Google, “Google Cloud Vision API enters Beta, open to all to try!” 2016. [Online]. Available: https://perma.cc/D6LF-NQ4X

[10] C. D. Manning, P. Raghavan, and H. Schutze, “An Introduction to Information Retrieval,” *Information Retrieval*, no. c, pp. 1–18, 2009.

[11] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet Allocation,” *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[12] Mass511, “Cameras — Traffic Cameras - Mass511.” [Online]. Available: https://mass511.com/list/cctv

[13] Google, “google-cloud vision 963db69 documentation.” [Online]. Available: https://perma.cc/3D26-2VXK

[14] A. Singhal, “Introducing the knowledge graph,” 2012. [Online]. Available: https://perma.cc/6BN7-WJL

[15] T. L. Griffiths and M. Steyvers, “Finding scientific topics,” *Proceedings of the National Academy of Sciences*, vol. 101, no. Supplement 1, pp. 5228–5235, 2004.

[16] M. Hoffman, D. Blei, and F. Bach, “Online learning for latent dirichlet allocation,” in *advances in neural information processing systems*, 2010, pp. 856–864.

[17] M. Steyvers and T. Griffiths, “Probabilistic Topic Models,” *Handbook of latent semantic analysis*, vol. 427, no. 7, pp. 424–440, 2007.

[18] M. J. Menne, I. Durre, B. Korzeniewski, S. McNeal, K. Thomas, X. Yin, and Others, “Global historical climatology network-daily (GHCN-Daily), Version 3,” pp. subset: December 2017–January 2018, 2012. [Online]. Available: https://perma.cc/3D26-2VXK

[19] T. Andersen and E. Sweeney, “Parking ban being lifted at 5 p.m. in Boston - The Boston Globe,” 2018. [Online]. Available: https://perma.cc/Q47F-ZWFT

[20] R. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, “Image change detection algorithms: a systematic survey,” *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 294–307, mar 2005.