SRC: City-Scale Mapping of Pets Using Georeferenced Images

Aaron San Jose and Eduardo Hernandez (Mentor: Shawn Newsam)
University of California, Merced
5200 N. Lake Rd., Merced, CA 95343
[asanjose,ehernandez58]@ucmerced.edu

ABSTRACT

We investigate the mapping of pet activity using social media. Specifically, we perform cat and dog detection in a large collection of georeferenced images in San Francisco. We compare detection based on keyword search in user-supplied tags to detection based on image content using state-of-the-art deep-learning classification methods. The resulting city-scale spatial distribution of cat and dog activity makes sense based on our knowledge of the region. Our approach represents a general framework for mapping phenomena that are difficult to observe through traditional means.

CCS Concepts

•Information systems → **Geographic information systems; •Computing methodologies** → *Object detection;*

Keywords

Georeferenced images, geographic knowledge discovery

1. INTRODUCTION

There is a plethora of untapped data in Internet social media feeds that could be used to answer various interesting questions. For example, images uploaded on social media feeds are frequently of pets. Given a large number of such images, with their locations, one should be able to map where pets are. This is the focus of our project. We perform pet detection in a large number of georeferenced social media images. Mapping these detections allows us to analyze spatial trends of pet activity at a city scale.

The key technical challenge is automating the detection. We investigate two different approaches to this problem 1) applying text-based search algorithms to the user-submitted tag descriptors of the images, and 2) applying computer vision classification algorithms to the actual image content.

2. APPROACH

We seek to label each image as containing a dog or a cat. We then assign this detection to the location of the image in order to perform the spatial analysis.

2.1 Text-Based Detection

In order to classify the images through their text tags, we use keyword search algorithms. Each image has a varying number of text tags that have been provided by the user. If our search term, for example "dog", matches any of the tags, we mark the image as a detection. Text tags do not necessarily describe exactly what is in the image or completely prove that a pet is in the photo or not, though.

2.2 Image-Based Detection

Deep learning is recent, effective method of image classification that creates models based off of "learned features" of a visual class using convolutional neural networks (CNNs). CNNs are trained to recognize classes in a supervised fashion. A model is learned by feeding it labeled images. It can then be used to perform detection in unseen images. The training process involves tuning layers of neurons that perform simple tasks, like image convolution or subsampling, that culminate in a larger task, like image classification.

CNNs are useful as non-binary classifiers, or classifiers with multiple classes. During prediction, a CNN will produce a vector of size N, where N is the number of classes, of the probabilities that the image belongs to each class. A CNN usually will normalize the probabilities so they sum to one using a softmax function, and then return an encoding that gives the label for the class with the highest probability. The actual return value is another vector of size N that contains all zero values except for one index that holds the value one. This index indicates the predicted class.

We apply a CNN that has been trained to recognize a large number of visual classes including dogs and cats. This also includes specific breeds.

3. EXPERIMENTAL RESULTS

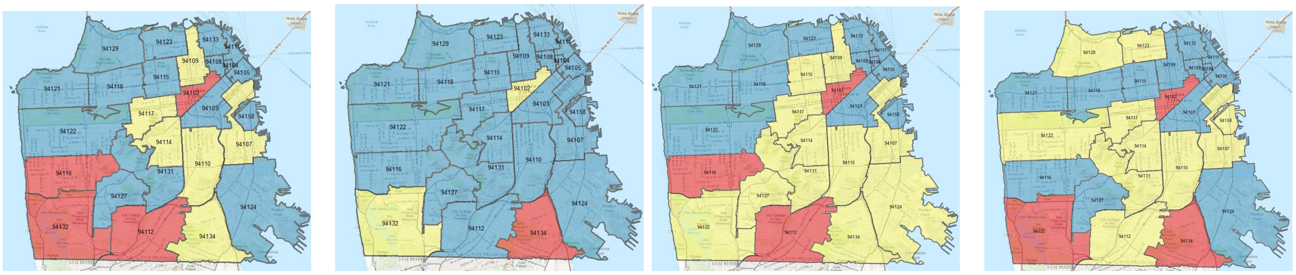
We applied the proposed approach to over one million Flickr images of San Francisco taken between 2008 and 2015.

For the text-based detection, we also used specific dog breeds (e.g., "terrier", "hound") and dog synonyms ("canine") as keywords to detect dogs. Similarly, we used specific cat breeds (e.g, "Siamese", "Egyptian") and cat synonyms ("feline") for detecting cats. Rows two and three of Table 1 show the number of cat and dog detections per year based on performing keyword search on the user-provided tags.

For the image-based detection, we used a CNN called Inception-v3 [2] that has been trained on the ImageNet Large

Table 1: **Rows two through five indicate the number of detections. The last row shows the total images.**

Approach	2008	2009	2010	2011	2012	2013	2014	2015	Total
Text Cat	217	272	78	235	128	177	217	177	1501
Text Dog	624	537	196	722	323	456	252	199	3309
Image Cat	258	313	79	426	240	464	351	396	2527
Image Dog	735	941	228	1554	891	1486	901	1069	7805
Total Photos	158720	184742	61942	202154	116826	195390	143824	72784	1136382



(a) Text Cat (b) Text Dog (c) Image Cat (d) Image Dog

Figure 1: Per-zip code activity mapping. Red, yellow, and blue indicate high, medium, and low activity. (a) and (b) are the results of text-based detection and (c) and (d) are of image-based.

Scale Visualization Dataset [1]. This dataset contains 1000 different classes, including various breeds of cats and dogs. The Inception-v3 CNN returns the five most likely classes for an image and we mark a detection if any of these five classes are related to cats or dogs. This network has been shown to be very effective, achieving an error rate of just 3.46% for the top five predictions [2] on the 1000 class ImageNet data set. Rows four and five of Table 1 show the number of cat and dog detections per year in our data set based on image content.

We aggregated the detections by zip code to perform our spatial analysis. We calculated a pet activity value for each zip code by normalizing the number of detections by the total number of Flickr images in that zip code. Fig. 1 shows the resulting maps where each of the 26 zip codes is labeled as having low (blue), medium (yellow), or high (red) activity.

4. DISCUSSION

We do not have a ground truth to evaluate our results. However, we make the following observations based on Table 1 and Fig. 1.

Our image-based method results in over twice as many detections as the text-based. This demonstrates the potential benefit of exploiting the image content through state-of-the-start image understanding.

Both methods, text- and image-based, result in more dog detections. This could indicate that there are more dogs in San Francisco than cats (or, really, that people take more pictures of dogs).

We observe the following spatial patterns in Fig. 1.

- The two methods result in very similar spatial distributions for each type of pet. Compare the similarities between the text- and image-based cat activity in Figs. 1(a) and 1(c) and text- and image-based dog activity in Figs. 1(b) and 1(d). While image content results in more overall detections, the spatial distributions of the two methods are very much in agreement.
- Dog activity is high where there are parks. Fig. 1(d) shows high dog activity in 94132 which includes the sizable Lake Merced Park and 94134 which includes the sizable John McLaren Park and medium dog activity in 94122 which includes Golden Gate Park and 94129 which include the Presidio. In contrast, as seen in Fig. 1(c), cat activity is lower in the zip codes with parks and higher in more residential zip codes such as 94116 which contains the Sunset District and 94112 which contains Ingleside, Excelsior, and the Outer Mission.
- Despite of there being more dog detections overall (Table 1), they are more concentrated. Compare the dog detections in Figs. 1(b) and 1(d) with the cat detections in Figs. 1(a) and 1(c).
- Fisherman's Wharf, North Beach, and the Embarcadero, tourist regions in 94133 and 94111, contain very little pet activity.
- We detect high pet activity in 94102 which is downtown and very urban. This is somewhat surprising and warrants further investigation.

5. CONCLUSION

We demonstrated a framework that uses georeferenced social media to measure phenomena that might not be observable through other means. Specifically, we explored two methods to detect pets in Flickr images and then mapped the results at the city-scale. The spatial distributions make sense based on our knowledge of the region.

6. ACKNOWLEDGMENTS

This work was funded in part by an NSF CAREER grant, #IIS-1150115, through an REU supplement. We would like to thank the UC Merced Spatial Analysis and Research Center (SpARC) for help with preparing the maps.

7. REFERENCES

[1] O. Russakovsky et al. ImageNet Large Scale Visual Recognition Challenge. *IJCV*, 115(3):211–252, 2015.

[2] C. Szegedy et al. Rethinking the inception architecture for computer vision. *CoRR*, abs/1512.00567, 2015.