Perceiving Commerial Activeness Over Satellite Images

Zhiyuan He
Shanghai Key Laboratory of Intelligent Information Processing, School of Computer Science, Fudan University
Shanghai, China
16210240032@fudan.edu.cn

Su Yang*
Shanghai Key Laboratory of Intelligent Information Processing, School of Computer Science, Fudan University
Shanghai, China
suyang@fudan.edu.cn

Weishan Zhang
Department of Software Engineering, China University of Petroleum
Qingdao, China
zhangws@upc.edu.cn

Jiulong Zhang
School of Computer Science, Xi'an University of Technology
Xi'an, China
zjl@xaut.edu.cn

ABSTRACT

Different urban regions usually have different commercial hotness due to the different social contexts inside. As satellite imagery promises high-resolution, low-cost, real-time, and ubiquitous data acquisition, this study aims to solve commercial hotness prediction as well as the correlated social contexts mining problem via visual pattern analysis on satellite images. The goal is to reveal the underlying law correlating visual patterns of satellite images with commercial hotness so as to infer the commercial hotness map of a whole city for government regulation and business planning. We propose a novel deep learning-based model, which learns semantic information from raw satellite images to enable predicting regional commercial hotness. First, we collect satellite images from Google Map and label such images with POI categories according to the annotations from OpenStreetMap. Then, we train a model of deep convolutional networks that leverage raw images to infer the social attributes of the region of interest. Finally, we use three classical regression methods to predict regional commercial hotness from the corresponding social contexts reflected in satellite images in Shanghai, where the applied deep features are learned from the examples of Beijing to guarantee the generality. The result shows that the proposed model is robust enough to reach 82% precision at average. To the best of our knowledge, it is the first work focused on discovering relations between commercial hotness and satellite images. A web service is developed to demonstrate how business planning can be done in reference to the predicted commercial hotness of a given region.

CCS CONCEPTS

• **Applied computing** → *Sociology*;

KEYWORDS

Urban Perception, Deep Learning, Ubiquitous Computing

ACM Reference Format:
Zhiyuan He, Su Yang, Weishan Zhang, and Jiulong Zhang. 2018. Perceiving Commerial Activeness Over Satellite Images. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3184558.3186353

1 INTRODUCTION

The infrastructures of some modern cities change continuously fast. New commercial centers are arising quickly, while old centers might be decaying [18, 24]. In [19], some laws regarding the developing and decaying of regions are summarized from computer vision analysis over streetscapes. Except for such insight into the varying appearances of cities in a qualitative sense, we understand neither how city infrastructures affect economy nor we know the quantitative correlation between city infrastructures and economic indices. As some cities are developing vary fast, a critical issue is to learn lessons from the past to enable better city planning in the future. Due to the lack of knowledge regarding the impact of city infrastructures on commercial activeness, so far, city planning from the economic point of view remains a missing topic. Moreover, it is crucial for business owners to get recommendations on choosing suitable locations for their business so as to maximize the profits, for which an operational way is to establish a predictor based on revealing the relation between city infrastructures and commercial activeness.

So far, how commercial hotness is shaped by city infrastructures has never been fully understood. Luckily, the abundant resources of big data at this age provide new means to percept city infrastructures as well as the underlying laws governing the formation of commercial hotness. As city infrastructures change with regions, intuitively, the social functions underlying the infrastructures of a region should have impact on the formation of commercial hotness, for example, transportation networks, the number of houses and offices, and the scale of shopping centers if any. In general, different configuration of such resources for a region should lead to different commercial hotness. Since all city infrastructures are visible in satellite images, in this study, we propose to make use of satellite images to investigate the correlation between city infrastructures and commercial hotness. Satellite imagery has the

*Corresponding author. The author is also with School of Computer Science, Xi'an University of Technology.

WWW '18 Companion, April 23–27, 2018, Lyon, France

ACM ISBN 978-1-4503-5640-4/18/04.
https://doi.org/10.1145/3184558.3186353

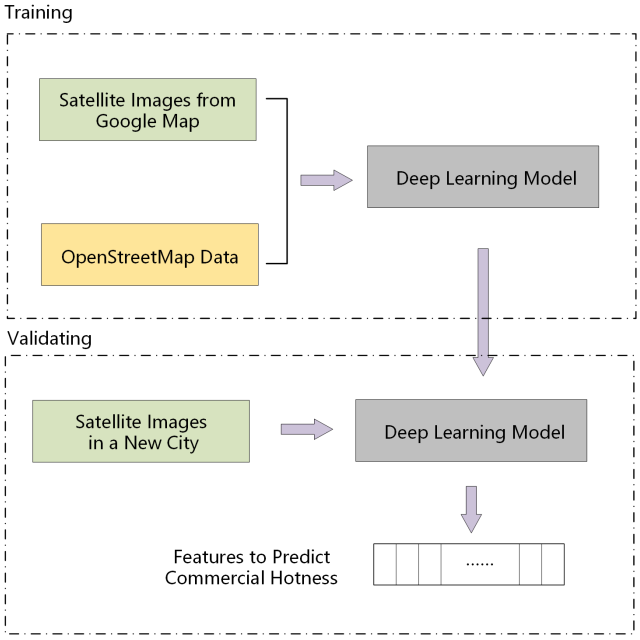

Figure 1: The main framework: First, train a deep learning-based model using satellite images and OpenStreetMap labels. Then, apply it to a new city, where regional features are generated as the basis to compute commercial hotness.

following natures: (1) Ubiquitous: Images of almost all cities on earth are available; (2) Low-cost: We can download free satellite images from websites such as Google Map while there are also paid services to obtain satellite images, which are not very expensive; (3) Nearly real-time: It is possible to obtain high-resolution satellite images per hour, and as a result, tracing the development of regions in a city becomes easy.

In fact, satellite imagery has become a major means for urban perception. The previous studies are mainly focused on recognizing ground objects such as roads, buildings, and natural plants, or detecting changes of ground objects through image registration. Recently, the trend of the state-of-the-art researches has been shifted to knowledge discovery on association between visual patterns of satellite image and socioeconomic indices so as to establish a predictive model to percept location-aware social contexts or economic issues over satellite images. In [15], Jean etc. propose to use convolutional neural networks along with transferring learning to learn features for predicting the poverty across a country from nighttime light intensities over satellite images. In [21], road safety is predicted from satellite images by learning the correlation between the historical data of accidents and the satellite images in the framework of deep learning. In [2], convolutional neural networks are applied to tackle the problem of land use classification. Nevertheless, learning the mapping from satellite images to broad-spectrum social-economic issues is an open problem yet, and commercial hotness prediction from satellite images is so far a missing topic in terms of urban perception.

The goal of this study is to discover the relation between the visual patterns of satellite images and commercial activeness. Once a predictor from satellite images to commercial hotness can be established by mining the correlation between them, city planning as well as business planning can be conducted in a rational way on the basis of referring to the map of commercial hotness over satellite images, which is a byproduct turned out from the predictor by applying the image of the region of interest as input. By producing the map of commercial hotness region by region, it leads to an integral profile in terms of commercial activeness over city infrastructures throughout the city, which visualizes the city from the economic perspective for city planners. Moreover, for newly developed regions with short history and few data, the predicted commercial hotness over satellite images becomes an important clue to foresee the future in terms of business planning. To approach high prediction precision, the key is to reveal the correlation between satellite images and commercial hotness from historical data, and develop an inference mechanism to approximate such correlation as close as possible. However, it is not an easy task due to the heterogeneous data. Although deep learning has been successfully applied to urban perception from satellite imagery [15, 30], directly predicting commercial hotness from satellite images is not practical. The main reason lies in that the satellite images are in a highly unstructured form with seemingly irregular visual patterns and deep learning models can hardly find explicit clues of commercial activities on raw satellite images. To overcome this problem, we propose to establish a semantic layer between satellite image and commercial activeness. The semantic layer captures social functions and city infrastructures from raw image data. Then, prediction of commercial activeness is performed based on identification of city infrastructures at the semantic layer. The motivation is as follows: On one hand, it has been revealed that the commercial hotness of a given region is subject to the social contexts in and around the region of interest [29] and Point of Interest (POI) plays an important role, namely, the class label of the city function for a given location such as shopping mall or transportation facility. On the other hand, it has been demonstrated that coarse-gained POI can be recognized from satellite images [2]. This gives rise to an interesting problem: Is it possible to predict commercial hotness from satellite images through POI? Therefore, we make use of the annotations from *OpenStreetMap*, which is a crowdsourcing project to create a free editable map of the world, to label each image with a class label of POI. Then, a deep convolutional model can be trained to classify these labeled images so as to identify the city infrastructures indicated by the corresponding POI categories. We divide the city of interest into a grid of multiple regions. In each region, we use the deep learning model to produce a mean feature vector representing the statistical distribution of POIs. Then, some regression methods, including Linear Regression, Support Vector Regression (SVR), and Gradient Boosting Decision Tree (GBDT), are used to predict commercial hotness given the regional features of POIs. To evaluate the generality of the proposed method, we apply the model trained from one city to predict the commercial hotness in another city. The highest prediction accuracy is 82%, which indicates that there is a close relation between satellite images and commercial activities. The main framework of the proposed method is shown in Fig. 1. Additionally, to demonstrate our idea of commercial hotness

Figure 2: An online demo to demonstrate inferring commercial hotness from satellite imagery. First, choose the region of interest on the satellite map of the city. The corresponding longitude and latitude of this region will appear on the top right corner. Then, click the "submit" button. The predicted value of the commercial hotness of this region will be calculated along with the statistical distribution of the ground objects corresponding with the 6 classes of POIs of interest.

prediction from satellite imagery, we develop a web service portal, which is shown in Fig. 2.

To the best of our knowledge, this is the first study to utilize satellite imagery to predict commercial hotness. Contributions made in this paper are summarized as follows:

- A framework is proposed for commercial hotness prediction on the basis of leveraging a deep neural network to percept the social functions reflected in satellite images, followed by a decision tree-based inference mechanism to mine the association between social functions and commercial hotness. This can help city planners to learn lessons for further city planning. From the technical point of view, the system works due to the implementation as follows: Raw satellite images make no sense in terms of semantics and it is hard to correlate directly the raw data of satellite images to commercial activeness. Thus, we propose to add a semantic middle layer between the visual patterns and commercial activeness. That is, we make use of OpenStreetMap to produce additional annotations so as to learn and infer the statistics of social functions, say, POIs, from satellite images.
- We predict the city-scale commercial hotness for Shanghai with an accuracy of 82% using the deep features learned from Beijing, which proves the generality of the method.
- We develop a web service portal to demonstrate predicting commercial hotness in an interactive manner for any region in Shanghai, which provides easy-to-use decision support for business location planning.

2 RELATED WORKS

To the best of our knowledge, so far, there is no open report on commercial hotness prediction with satellite imagery. Yet, some studies have been conducted aiming to predict social contexts from raw satellite images. In [15], deep learning and transfer learning are used to predict poverty using satellite images. Due to the lack of poverty data, it is almost impossible to train a network to predict poverty directly and night lights are used as a proxy to solve this problem. Satellite images are also used to predict urban land use [2, 4], road safety [21], and urban air quality [10].

There are also many other works leveraging computer vision methods to achieve urban perception. Urban perception refers to using modern sensors (for example, satellite sensors) to collect urban big data for analyzing, planning, and predicting urban social attributes. In [20], traditional computer vision methods are used to predict safety scores from streetscapes. Deep learning is utilized to advance technically this task in [22]. Except for safety scores, more social attributes can be predicted such as violent crimes [3] and the distance to the closest commercial center [16]. Some other studies are focused on city identity recognition, such as [8] and [31].

This study is the first work devoted to discovering the correlation between visual features of satellite images and commercial hotness. The most related work is [16], where they predict the closet

distance to a commercial center, which is implicitly correlated to commercial hotness. However, [16] uses streetscapes rather than satellite imagery, which should provide a more direct lens to measure the distance between ground objects. Although streetscapes and satellite images are both free data with easy accessbility, we prefer satellite imagery due to the following reasons: In view of the rapidly changing nature of city infrastructures, satellite imagery promises ubiquitous and real-time data acquisition against city development.

3 DATA

The data collection steps are as follows:

- Both satellite images and labels are needed to train the deep learning-based model. Using Google Map API, we download satellite images of Beijing[1]. All the images are at zoom level 18. Then, we download the OpenStreetMap data of Beijing to produce the corresponding image labels. We randomly sample 48,000 images for training, and 12,000 images for evaluation. The details will be described in the following section.
- We use the trained deep models to generate attributes for regions, which can be regarded as the features applied to predicting commercial hotness. To conduct unbiased validation, we predict commercial hotness in Shanghai using the deep model learned from Beijing. Similarly, we also download the satellite images of Shanghai[2] from Google Map.
- For the sake of training the predictors for Shanghai, we need a quantitative number to reflect the commercial hotness of an area. We use the data from Dianping.com, the biggest website in China to share comments on commercial entities, the function of which is similar to that of Yelp in US. The data include geographical locations, user comments and user's rating. Here, we use the number of comments as the proxy of commercial hotness since the online comments are from the consumers who have truly experienced the services and goods provided by the commercial entities.

In [29], urban big data including POI, human mobility, the number and purchasing power of local residents, and online reviews are used to predict commercial hotness of urban commercial districts, and it is revealed that POI plays an important role in commercial activeness prediction. So, POI identification is essential in terms of predicting commercial hotness. As far as we know, annotation of POI labels over geographical data highly relies on labor-intensive manual works. As satellite images can provide coarse profiles of region functions and have been used to classify land uses in [2], this study aims to automate POI classification and commercial hotness prediction from raw satellite image data as a whole, where deep learning is applied to achieve optimization for POI classification, which acts as the basis for commercial activeness prediction.

[1]Because predicting urban commercial hotness in non-urban area is meaningless, we only consider the main urban area of Beijing. To be specific, we use a rectangle area ranges from 40° 09' 24" N, 116° 09' 27" E to 39° 44' 01" N, 116° 40' 18" E.

[2]A rectangle area containing the main urban area ranges from 31° 25' 45" N, 121° 07' 57" E to 30° 49' 21" N, 121° 59' 22" E.

4 METHODOLOGY

Our goal is to predict commercial hotness from raw satellite images. However, commercial activities are subject to many factors, such as the total number of residents, the real purchasing power of the residents, and the density of shops. Directly mapping raw satellite images to commercial hotness is not practical, even with a deep learning predictor. The reason lies in that satellite imagery leads to a highly unstructured form of data such that it is hard to obtain explicit clues indicating reliably social attributes.

We use OpenStreetMap as a supplementary data source for data labeling with POI categories. OpenStreetMap is a project aiming to build map data over the world. All data on OpenStreetMap are free to download. Given this map data, we can assign a class label for each image. Thus, we can successfully train a deep learning classifier for raw satellite images. The trained model should be robust enough to be applicable to another city, where the model is used to generate a regional feature vector by classifying all the image blocks inside the region of interest. We view this feature vector as a *social context representation* since it is obtained under the supervision of POI from OpenStreetMap, based on which we can proceed to predicting commercial hotness. The main framework of the proposed method is shown in Fig. 1.

In the following, we describe the proposed method in detail. First, we explain how we assign the image labels using OpenStreetMap data. Then, we describe how we train the deep learning model for the labeled images and how we use the model to yield a regional feature vector. Finally, we describe how we use the feature vector to predict commercial hotness.

4.1 Label Images Using OpenStreetMap Data

OpenStreetMap offers detailed map information for almost all areas of the world. The basic data structure of OpenStreetMap is node, way, and relation. Node is the most basic concept. It stands for a single point on the map, which consists of a latitude, a longitude, and a node id. A way is an ordered list of nodes. It can be open or closed. In a closed way, the last node and the first node are identical but they are not the same for an open way. Roads, railways, and streams are represented with open ways, while closed ways are used to define meaningful areas like residential spots, business districts, industrial areas, and schools. Relation is a concept to describe relationship between nodes and ways.

In this paper, we use *closed way* to label raw satellite images. According to its definition, we can view a closed way as a polygonal area with tag information. Overall, there are over 2,000 tags on OpenStreetMap[3]. In practice, most tags are rarely used. We only use the tags that are widely used and related to commercial activities. One tag corresponds with one class.

How we assign class labels to satellite images is as follows: There is a contradiction here: A satellite image, which is a Google Map tile, is in a standard square shape, while a way is an irregular polygon. We calculate the overlap area in a satellite tile. If it exceeds a predefined threshold (In all experiments, it is 70%), we will assign the class label to the image tile according to the tag of the way, which is illustrated in Fig. 3.

[3]http://wiki.openstreetmap.org/wiki/Category:Tag_descriptions

Figure 3: Use OpenStreetMap data to generate labels for satellite tiles. An OpenStreetMap way is an irregular polygon. In one satellite tile, calculate the overlap area. If it exceeds a threshold, the tile will be assigned a class label according to the tag of the way.

4.2 Train a Deep Learning-based Classifier

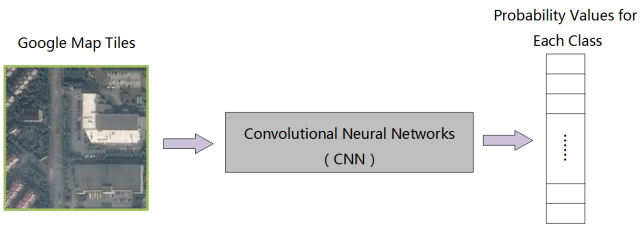

Figure 4: The CNN model to classify labeled satellite images: The model takes a single image S_i as an input; The output is a probability that the input belongs to each class (denoted as C_i).

Once labeled satellite images are available, we can train a deep learning-based classifier. In this paper, we employ a convolutional neural network (CNN) to classify raw satellite images. A CNN is a class of deep feed-forwarded networks. It has been proven successful when applied to computer vision tasks. Traditional machine learning methods for image data usually start with handcrafting features. This kind of work needs much domain knowledge and labor. A CNN model can learn to extract features automatically from raw images. Thus, using a CNN model can not only save time, but also improve the final performance dramatically.

The CNN model takes a raw satellite image (denoted as S_i) as an input and produces class-dependent probabilities as an output, where S_i is an RGB image with fixed width and height. Although raw satellite images are unstructured and only contain low-level features, the CNN model can learn to approach the semantically meaningful targets under the supervision of POI labels. This process is accomplished by stacked convolutional layers. One convolutional layer usually consists of 3 parts: The first part is a convolutional kernel, which is responsible to perform convolutional transformation upon the inputed 3-D tensor. The 3 dimensions of the tensor are width, height, and channel. This tensor is also known as the feature map. Then, the result is fed to an activation function to perform nonlinear transform, where Rectified Linear Unit (ReLU) is often used. The third part is a pooling layer, which makes the output to be more robust against minor disturbances. The stacked structure of convolutional layers makes it easier to capture the visual patterns of an RGB image hierarchically. The lower convolutional layers can learn to extract low-level features such as colors and edges, while the deeper layers function to extract semantic features like shapes and objects. Those deep features are followed by some fully-connected layers with a softmax function to produce the probability belonging to each class.

The CNN model is usually trained in a supervised manner using stochastic gradient descent methods.

We denote the output of the CNN model as C_i, which is a class-dependent probability with k possible classes in total. The flowchart of the method is shown in Fig. 4.

4.3 Compute Regional Feature Vectors

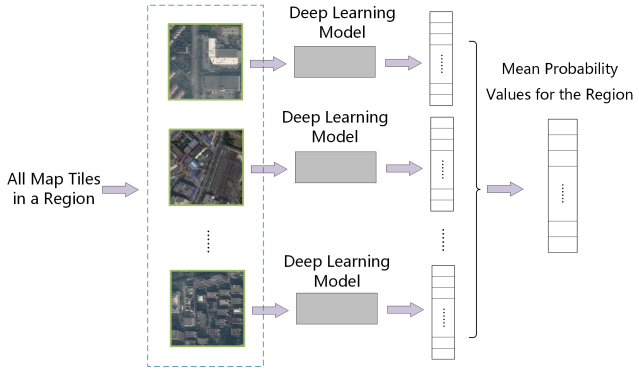

Figure 5: Feature extraction for each region: Using the trained deep model, we can obtain the probability values for every single image. The feature for a region is the average of all the feature vectors in association with the images in this region.

It is almost impossible to predict commercial hotness for a single satellite image tile. For example, one satellite image could only contain a part of a house at a high zoom level. This kind of information is too minor to make sense. However, if we switch to a lower zoom level, some details will be missing due to the limited resolution. Thus, we predict commercial hotness over regions instead. We divide the whole city into a grid of $n \times m$ regions (Denoted as $Region_j, j = 1..n \times m$). Each region contains a couple of image blocks. As shown in Fig. 5, we can obtain class-dependent probability values for each image block using the trained deep model. The features for the region is the average of the class-dependent probabilities of all the image blocks in this region:

$$Feature_j = \frac{\sum_{S_i \in Region_j} C_j}{|Region_j|}$$

We view $Feature_j$ as the vector to reflect the social contexts in the region. In fact, it figures out quantitatively the proportion of different classes, such as residential area, business area, industrial area, and school. Here, we compute the features as such is due to the following consideration: It has been experimentally verified that the statistical distribution of the POIs in the region of interest plays

an important role in predicting the regional commercial activeness [29]. $Feature_j$ can reflect the social environment and composition of the region. Compared with the original satellite image features, this vector contains high-level semantic information, which is considered to be a powerful feature to predict social attributes such as commercial hotness in the region.

4.4 Predict Regional Commercial Hotness

We use the total number of the web reviews on the commercial entities in this area as the proxy of commercial hotness since only the customs who have experienced the services and goods provided by the commercial entities wrote such comments on web, and the number of the comments reflect how much the commercial entities attract attentions, namely, commercial hotness. Here, the review data are from Dianping.com, which is the best-known web site for leaving online reviews in China, like Yelp in US. The regional feature vector $feature_j$ is used to predict the commercial hotness defined as such. We use three standard regression methods to perform prediction from $feature_j$: Linear Regression, Support Vector Regression (SVR), and Gradient Boosting Decision Tree (GBDT). The results are provided in the following section.

5 EXPERIMENT

We first train a CNN model with labeled satellite images in Beijing. Then, we validate the model by predicting the commercial hotness in another city, Shanghai. This proves the generality of the predictive model with deep learning.

5.1 CNN Training

Below we explain how we train the CNN classifier:

Datasets: We download all satellite images of Beijing from Google Map API. All images are at zoom level 18 and have a resolution of 256×256. Then, we use OpenStreetMap data to produce labeled images. The method to label every single image has been described previously. We use 6 tags, in other words, 6 classes, to train the CNN classifier: Water, wood, residential area, industrial area, commercial area, and farmland. 10,000 images are sampled for each class randomly. Then, we divide the images into training and validation set. For each class, there are 8,000 images used for training and 2,000 images for validation.

Our model does not limit the total number of classes. In a future work, more classes can be considered to form a richer social context-represented vector.

Architecture: We use Inception V3 [27] as our basic CNN architecture. There are many successful CNN architectures, such as AlexNet [17], VGGNet [23], Inception networks [14, 25–27], and ResNet [13]. Inception V3 is one of these successful architectures. We use Inception V3 because there is a easy-to-use implementation and it also gives a strong baseline.

Training: Instead of training from scratch, we initialize the parameters based on a pretrained ImageNet [7] model. Overall, we have trained two CNN models: For one model, we only trained the last logit layer. Thus, we can view the pretrained CNN model as a traditional image feature extractor. Extracted image features are fed into a softmax classification model to give a result. The other model is trained on all layers of Inception V3.

	Top-1 Accuracy	Top-3 Accuracy
Only logit layer	0.685	0.942
All layers	**0.743**	**0.958**

Table 1: Classification accuracy of our CNN models. The training dataset consists of 48,000 images, 8,000 images for each class. The validation dataset consists of 12,000 images, 2,000 images for each class. We calculate top-1 and top-3 accuracy on the validation dataset. In sum, we have trained two models: One is only trained on the logit layer; The other is trained on all layers, which has better accuracy.

For these two models, we use 32 as the batch size and 0.00004 as the weight decay. The learning rates for the two models are different. We use a learning rate of 0.001 to train only logit layer and 0.0001 to train all layers. Both models are optimized by a RMSProp optimizer [28]. Finally, training is conducted with TensorFlow [1] framework running on a single Nvidia GeForce Titan Black GPU.

Evaluation: To evaluate the trained models, we calculate the classification accuracy on the testing dataset, which has 2,000 images for each class. We report top-1 and top-3 accuracy after 20,000 iterations. Top-k accuracy refers to the portion of the correct hits among all the predictions, where a correct hit means that one of the top-k candidates resulting from a prediction agrees with the ground truth. As shown in Table 1, both models promise a reasonable precision. The highest top-1 classification accuracy is 0.743.

5.2 Commercial Hotness Prediction

We use the trained deep learning model to predict the commercial hotness in another city. Instead of predicting commercial hotness from a single image, we perform predicting for every region in a city. The details are as follows:

Datasets: We divide Shanghai into $n \times m$ regions. In all the experiments, we let $n = 20$ and $m = 20$. For each region, the feature can be viewed as the mean class-dependent probability values of all the satellite images in this region to figure out the portion of different classes of POIs. It is generated by the trained deep learning-based model. The commercial hotness of each region is the total number of the online comments on the commercial entities in this region. The review data are gathered from Dianping.com. We perform a log transformation on the number of the online comments for each region and use this value as the ground truth to supervise the learning as well as for performance evaluation.

Linear Regression: We use linear regression as our baseline model. Given a dataset of p-dimension features $x_i = [x_{i1}, x_{i2}, ..., x_{ip}]$ and labels $y_i, i = 1, 2, ..., n$, a linear regression model assumes that there is linear relationship between x_i and y_i. There also exists a noise term ϵ_i. Thus, a linear regression model can be formulated as

$$y_i = \beta^T x_i + \epsilon_i$$

where β is a p-dimension parameter vector. Linear regression is one of the most basic regression models. We use it as a baseline model, which proves that commercial hotness can be predicted from satellite images successfully with the proposed method.

Support Vector Regression: In machine learning, support vector machine [6] is a supervised learning model used for classification.

	Linear Regression	SVR	GBDT
Only logit layer	0.743	0.748	0.799
All layers	0.804	0.813	**0.822**

Table 2: The R^2 values of the three regression models for commercial hotness prediction. The results are averaged over 5 folds. We validate the two deep learning-based models with three classical regression methods. It is shown that the proposed method can promise prediction with satisfactory precision.

A SVM model for regression, called support vector regression, is proposed in [9]. In the experiments, we also utilize SVR as one of the regression methods. Given the features x_i and labels $y_i, i = 1, 2, ..., n$, the original SVR model aims to minimize

$$\frac{1}{2}||w||^2$$

subject to

$$y_i - \epsilon \leq w^T \phi(x_i) + b \leq y_i + \epsilon$$

where $w^T \phi(x_i) + b$ is the model prediction for the sample x_i, and ϵ is a threshold parameter.

Gradient Boosting Decision Tree: Gradient Boosting [11, 12] is a machine learning technique, which combines weak predictors like decision trees to form a strong model. Gradient boosting decision tree (GBDT for short) is the third regression method used in the experiments. We also let x_i and y_i represent features and labels, respectively. The original GBDT can be viewed as a combination of N decision tree models $h_j, j = 1...N$:

$$F_N(x_i) = \sum_{j=1}^{N} h_j(x_i)$$

h_1 is directly learned by predicting $\{y_i\}$ from $\{x_i\}$ using decision trees. h_2 is learned by predicting $\{y_i - h_1(x_i)\}$ from $\{x_i\}$. h_3 is learned by predicting $\{y_i - h_1(x_i) - h_2(x_i)\}$ from $\{x_i\}$ and so on. In the experiments, we use a more complex implementation [5] of GBDT.

Training: For linear regression, no parameters need to be set. We use a 5-fold cross validation and grid search to find the best parameters for SVR and GBDT model. For SVR model, we utilize a radial basis function kernel. The penalty parameter C is chosen from $\{0.1, 1, 2, 10\}$, and the epsilon term ϵ from $\{0.05, 0.1, 0.15, 0.2, 0.5\}$. For GBDT model, the max tree depth is chosen from $\{2, 3, 4, 5\}$, the learning rate from $\{0.5, 0.1, 0.05, 0.01\}$, and the iteration number from $\{10, 50, 100, 150, 300\}$.

Evaluation: We utilize the coefficient of determination R^2 as the metric for evaluating the regression models. Given the observed values $\{y_1, y_2, ..., y_n\}$ and the predicted values $\{f_1, f_2, ..., f_n\}$, we first calculate the average of the observed values:

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Then, the R^2 value is given by:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - f_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

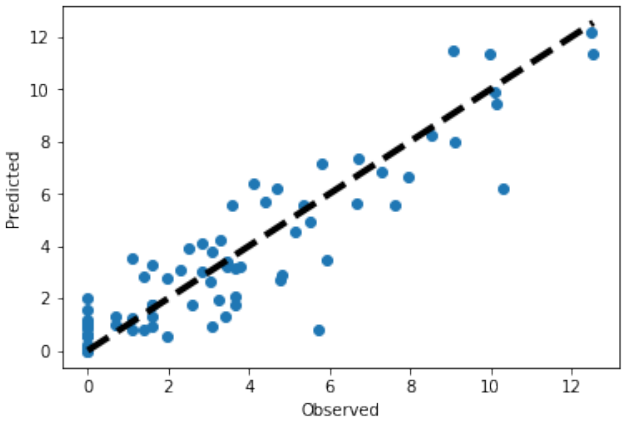

Figure 6: The observed commercial hotness (log of the comment number) and the predicted values resulting from the best model. The dotted line stands for function $y = x$.

The R^2 values are calculated over the 5 folds and are shown in Table 2. The best model is the one trained on all convolutional layers and followed by GBDT as the regression method. As shown in Fig. 6, this model can give a reasonable prediction of commercial hotness from satellite images.

6 SUMMARY & CONCLUSIONS

In this paper, we have investigated the use of deep learning to infer city-scale commercial hotness from raw satellite imagery. We propose a approach that takes advantage of the state-of-the-art CNN models and leverages open data to obtain a robust model for prediction.

To validate the proposed method, we first train a CNN model based on a large satellite image dataset, which contains 60,000 images of Beijing collected from Google Map API. The satellite images are labeled by using the OpenStreetMap data as a supplementary data source. Our best model achieves 74.3% classification accuracy on 6 classes in terms of identifying POI labels. Then, we apply the deep features resulting from the model to predict commercial hotness in Shanghai. It is done by predicting the number of comments on commercial entities region by region. We use the deep learning model to yield a social context feature vector for each region. On the other hand, the region is labeled using the log-form of online comments collected from Dianping.com. We employ three classical regression methods, Linear Regression (LR), Support Vector Regression (SVR), and gradient boosting decision tree (GBDT), to predict the commercial hotness value. The best result is achieved by the GBDT model with a R^2 value of 0.822, which proves that our model can promise commercial hotness prediction from satellite imagery with reasonable precision. Finally, we use the best model to build an online web service, which enables commercial hotness prediction of any region in Shanghai.

The experimental results confirm: (1) It is feasible to predict commercial hotness from satellite imagery. The visual features contained in satellite images can be a good indicator of commercial activities. (2) The state-of-the-art CNN model does help us to

achieve high prediction accuracy with good generality across cities. (3) We can use satellite imagery to predict commercial activities at city scale with relatively lower cost.

Although our method is practical on account of the high accuracy, it suffers from some limitations: First, our model is not an end-to-end system, which directly maps commercial hotness from satellite imagery using deep learning method. The reason is that predicting commercial hotness from a single image is almost impossible. So, we do prediction region by region. Second, we treat the number of web comment as a proxy of commercial hotness. This is easy-to-obtain but might not be exactly precise. Third, we train our model in Beijing and validate it in Shanghai. It has been proved that the model has good generalization power within a country due to the similar visual appearance in satellite images. More experiments can be carried out to see if the model can be generalized over countries (e.g., training in China and then predicting in America). These limitations are to be solved in future works.

7 ACKNOWLEDGEMENTS

This work is supported by NSFC (grant NO. 61472087) and Shanghai Science and Technology Commission (grant No. 1751110420).

REFERENCES

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467* (2016).

[2] Adrian Albert, Jasleen Kaur, and Marta Gonzalez. 2017. Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale. *arXiv preprint arXiv:1704.02965* (2017).

[3] Sean M Arietta, Alexei A Efros, Ravi Ramamoorthi, and Maneesh Agrawala. 2014. City forensics: Using visual elements to predict non-visual city attributes. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 2624–2633.

[4] Michael J Barnsley and Stuart L Barr. 1996. Inferring urban land use from satellite sensor images using kernel-based spatial reclassification. *Photogrammetric engineering and remote sensing* 62, 8 (1996), 949–958.

[5] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 785–794.

[6] Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine learning* 20, 3 (1995), 273–297.

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 248–255.

[8] Carl Doersch, Saurabh Singh, Abhinav Gupta, Josef Sivic, and Alexei Efros. 2012. What makes Paris look like Paris? *ACM Transactions on Graphics* 31, 4 (2012).

[9] Harris Drucker, Christopher JC Burges, Linda Kaufman, Alex J Smola, and Vladimir Vapnik. 1997. Support vector regression machines. In *Advances in neural information processing systems*. 155–161.

[10] Jill A Engel-Cox, Christopher H Holloman, Basil W Coutant, and Raymond M Hoff. 2004. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmospheric Environment* 38, 16 (2004), 2495–2509.

[11] Jerome H Friedman. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* (2001), 1189–1232.

[12] Jerome H Friedman. 2002. Stochastic gradient boosting. *Computational Statistics & Data Analysis* 38, 4 (2002), 367–378.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[14] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*. 448–456.

[15] Neal Jean, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. 2016. Combining satellite imagery and machine learning to predict poverty. *Science* 353, 6301 (2016), 790–794.

[16] Aditya Khosla, Byoungkwon An An, Joseph J Lim, and Antonio Torralba. 2014. Looking beyond the visible scene. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3710–3717.

[17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.

[18] Yueliang Ma and Ruisong Xu. 2010. Remote sensing monitoring and driving force analysis of urban expansion in Guangzhou City, China. *Habitat International* 34, 2 (2010), 228–235.

[19] N Naik, S. D. Kominers, R Raskar, E. L. Glaeser, and C. A. Hidalgo. 2017. Computer vision uncovers predictors of physical urban change. *Proceedings of the National Academy of Sciences of the United States of America* (2017).

[20] Nikhil Naik, Jade Philipoom, Ramesh Raskar, and César Hidalgo. 2014. Streetscore-predicting the perceived safety of one million streetscapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 779–785.

[21] Alameen Najjar, Shun'ichi Kaneko, and Yoshikazu Miyanaga. 2017. Combining Satellite Imagery and Open Data to Map Road Safety.. In *AAAI*. 4524–4530.

[22] Lorenzo Porzi, Samuel Rota Bulò, Bruno Lepri, and Elisa Ricci. 2015. Predicting and understanding urban perception with convolutional neural networks. In *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, 139–148.

[23] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[24] William L Stefanov, Michael S Ramsey, and Philip R Christensen. 2001. Monitoring urban land cover change: An expert system approach to land cover classification of semiarid to arid urban centers. *Remote sensing of Environment* 77, 2 (2001), 173–185.

[25] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. 2017. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning.. In *AAAI*. 4278–4284.

[26] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–9.

[27] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2818–2826.

[28] Tijmen Tieleman and Geoffrey Hinton. 2012. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning* 4, 2 (2012), 26–31.

[29] Su Yang, Minjie Wang, Wenshan Wang, Yi Sun, Jun Gao, Weishan Zhang, and Jiulong Zhang. 2017. Predicting Commercial Activeness over Urban Big Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 119.

[30] Liangpei Zhang, Lefei Zhang, and Bo Du. 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine* 4, 2 (2016), 22–40.

[31] Bolei Zhou, Liu Liu, Aude Oliva, and Antonio Torralba. 2014. Recognizing city identity via attribute analysis of geo-tagged images. In *European conference on computer vision*. Springer, 519–534.