

# Smart Citizen Sensing: A Proposed Computational System with Visual Sentiment Analysis and Big Data Architecture

Kaoutar Ben Ahmed  
FSTT, Abdelmalek  
Essaadi University  
Tangier, Morocco

Mohammed Bouhorma  
FSTT, Abdelmalek  
Essaadi University  
Tangier, Morocco

Mohamed Ben Ahmed  
FSTT, Abdelmalek  
Essaadi University  
Tangier, Morocco

## ABSTRACT

A city's "smartness" depends greatly on citizens' participation in smart city services. Furthermore, citizens are becoming technology-oriented in every aspect concerning their convenience, comfort and safety. Thus, they become sensing nodes—or citizen sensors—within smart-cities with both static information and a constantly emitting activity system. This paper presents a novel approach to perform visual sentiment analysis of big visual data shared on social networks (such as Facebook, Twitter, LinkedIn, and Pinterest) using transfer learning. The proposed approach aims at contributing to smart citizens sensing area of smart cities. This work explores deep features of photos shared by users in Twitter via convolutional neural networks and transfer learning to predict sentiments. Moreover, we propose big data architecture to extract, save and transform raw Twitter image posts into useful insights. We obtained an overall prediction accuracy of 83.35%, which indicates that neural networks are indeed capable of predicting sentiments. Therefore, revealing interesting research opportunities and applications in the domain of smart sensing.

## Keywords

Sentiment analysis; citizen sensing; opportunistic sensing; smart cities; big data; data warehousing

## 1. INTRODUCTION

The number of urban residents is growing by nearly 60 million every year. In addition, more than 60 percent of the world's population will be living in cities by 2050. As a result, people occupying just 2 percent of the world's land will consume about three-quarters of its resources. Moreover, more than 100 cities of 1 million people will be built in the next 10 years [1]. With this population growth, challenges including job creation, economic growth, environmental sustainability, and social resilience will arise. As such, "Digital urbanism" is rapidly becoming a central pillar for urban planners [2] and cities are witnessing a "smartness" evolution during an era where information and communication technologies (ICTs) have been exerting a growing and pervasive influence on the nature, structure and enactment of urban infrastructure, management, economic activity and everyday life [3].

Smart cities originally appeared in late 1990s ([4]; [5]) and have been evolving since then. Smart cities are currently the talk of the town due to their rapid increasing in discourses as well as in their real numbers. Smart city is a broad and variable concept, many studies and researches tried to define a smart city, namely the definition of Anthopoulos and Fitsilis [6] which presents smart city as "an ICT-based infrastructure and services environment that enhance a city's intelligence,

quality of life and other attributes (i.e., environment, entrepreneurship, education, culture, transportation etc.)" and according to Hall [7], a smart city is "a city that monitors and integrates conditions of all of its critical infrastructures, including roads, bridges, tunnels, rails, subways, airports, seaports, communications, water, power, event major buildings, can better optimize its resources, plan its preventive maintenance activities and monitor security aspects while maximizing services to its citizens".

While information and communication technologies (ICTs) have great potential for facilitating public participation, citizen engagement is a key element of most definitions of smart cities. Monitoring and managing data in order to feed back information into urban systems are practices that become constitutive of citizenship. Citizenship transforms into citizen sensing [3]. The concept of collecting data from sensors for monitoring purposes for which they were not originally designed is called citizen sensing or "opportunistic sensing" [8]. The idea is that mobile applications opportunistically exploit all the sensing technologies available in the environment (not requiring direct user interaction) [9]. Since 2006, when the term participatory sensing was first coined by Burke et al. [10], several studies have suggested and developed the concept of using citizens as sensors. The installation and maintenance of networks of physical sensors may be expensive and time consuming. Thus, citizen sensing provide a new, quick, cheap, and precise way to collect data about urban events. Citizen sensing can have higher impact on different sectors of society (e.g., health care and civil protection) than being solely specialized devices deployed for specific applications. This is due to their ability to exploit data collected throughout people's daily routines. The impact of citizen sensing can range from improvement of daily life routines and enhancing democratic debates [11] to the implementation of emergency/rescue applications.

Sentiment analysis is the examination of people's opinions, sentiments, evaluations, appraisals, attitudes, emotions, and personal preferences towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes [12]. Traditionally, sentiment analysis mines information from various text sources such as reviews, news, and blogs then classifies them on the basis of their polarity as positive, negative or neutral. An important preliminary task of sentiment analysis is to evaluate the subjective or objective nature of source texts. Subjectivity indicates that the text bears opinion content whereas objectivity indicates that the text is without opinion content. Recently, sentiment analysis aims to exploit audio, video, location, and other non-traditional data sources. Visual sentiment analysis is extremely challenging. Image sentiment analysis involves a much higher level of

abstraction and subjectivity in the human recognition process [13] and is more challenging than object recognition as the latter is usually well defined. Moreover, in order to use supervised learning, a large and diverse labeled training set, perhaps on the order of millions of images, is required. Additionally, the learning schemes need to have high generalizability to cover more different domains [14].

In this paper, we propose a novel approach to use visual sentiment analysis combined with big data warehousing as an emerging citizen sensing framework. We study the case of photos shared on Twitter posts. The rest is organized as follows. We review some related work to sentiment analysis and smart cities (section 2). Then, we explain in details our computational and storage system for sentiment prediction of Twitter photos using pre-trained deep neural networks (section 3). We limit our considerations, unless otherwise required, to the smart city application domain and do not aim to discuss sentiment analysis in its full generality.

## **2. RELATED WORK**

There is a departure from existing thinking on sensor networks because people are no longer just consumers of sensed data about some natural phenomenon or ecological process. Rather data about people are now sensed and collected such that the sets of producers and consumers of sensed data now overlap; people are ‘in the loop’ and may participate in both roles [15]. This ‘human-in-the-loop sensing’ combined with Social Media and mobile computing leads to the creation of citizen sensors. In 1999, Neil Gross predicted the future, in Bloomberg Business Week he said: “In the next century, planet earth will don an electronic skin. It will use the Internet as a scaffold to support and transmit its sensations. These will probe and monitor cities and endangered species, the atmosphere, our ships, highways and fleets of trucks, our conversations, our bodies-even our dream.” [16] Furthermore, Cuff et al. describe mobile devices as “passive sensors that can silently collect, exchange, and process information all day long” [17]. Moreover, Burke et al. assert that “participatory sensing will task deployed mobile devices to form interactive, participatory sensor networks that enable public and professional users to gather, analyze and share local knowledge” [10]. Some studies such as [18] mention “smart citizen” as a key parameter of the proactivity that citizens should be granted. Miluzzo et al. [19] introduced CenceMe, one of the first applications for smartphones capable of “sensing” the environment. Since then, several works in the literature proposed generic-purpose citizen sensing platforms ([20]; [21]).

Sheth [20] envisages that microblogging platforms (in which users can share short textual posts, URLs, videos, and images) as low-effort publishing services are of high interest to citizen sensing due to the large number of users connected on platforms such as Twitter, which enables the user to post from mobile devices with minimum effort [20]. As microblogging lends itself to instantaneous updates, data related to international events is posted before it can be reported on national television or traditional media. Sheth cites the example of Twitter posts during the Mumbai terrorist attacks in November 2008, when Twitter update by citizens using mobile devices reported observations of these events in almost real time [22].

As the number of citizens that are active on a social network continues to grow rapidly and public opinion views about government policies are scattered across the Internet, social sentiment analysis, has become one of the hottest solution

areas for Big Data. Sentiment analysis covers a wide range of applications in smart cities. The ability of government agencies to continuously keep a tab on pulse of its citizens by means of on-going sentiment analysis will pave the way to better governance. Sentiment analysis can address questions of primary importance to government agencies [23]. For example, how do citizens feel about the agency’s new programs and policies? What are the most talked about program? Is it perceived as good or bad? Sentiment analysis of citizens’ tweets can help identify critical events, such as road accidents, fires, street violence, etc., thus allowing to government agencies to provide fast response [24]. Moreover, sentiment analysis can assist decision-making processes and urban planning by providing a form of implicit feedback from the citizens ([25]; [26]). Sentiment analysis is instrumental in identifying problems by listening to the public, rather than by actively asking for input by means of surveys, and in doing so it is capable of ensuring a more accurate reflection of reality [27]. Sentiment mapping adds location to sentiment analysis. By knowing the places that tweets or other sources of sentiment are uttered from, or the names of places they refer to, it is possible to build a map of the areas where anger is running high or where travelers are happier, thus improving transportation services [28].

Sensing citizens’ sentiments about a scientific or cultural event can provide, in real time, the event’s organizers with meaningful insights into the degree of success of their initiatives [29]. Crowd pulse detection [30] and gross national happiness calculation [31] are examples of emerging sentiment analysis applications that will provide insight into the dominant mood of the population in real time. So far, the computational analysis of sentiment mostly concentrates on the textual content. Limited efforts have been conducted to analyze sentiments from visual content such as images and videos. Users of social media often post one or several images in addition to text in their messages. Existing methods for sentiment analysis are divided into three main categories by Cambria [32]: knowledge-based techniques, statistical methods, and hybrid approaches. Recent prominent deep learning methods ([33]; [34]) are becoming popular for their ability to learn sentiment representations from a large corpus of labeled and unlabeled text.

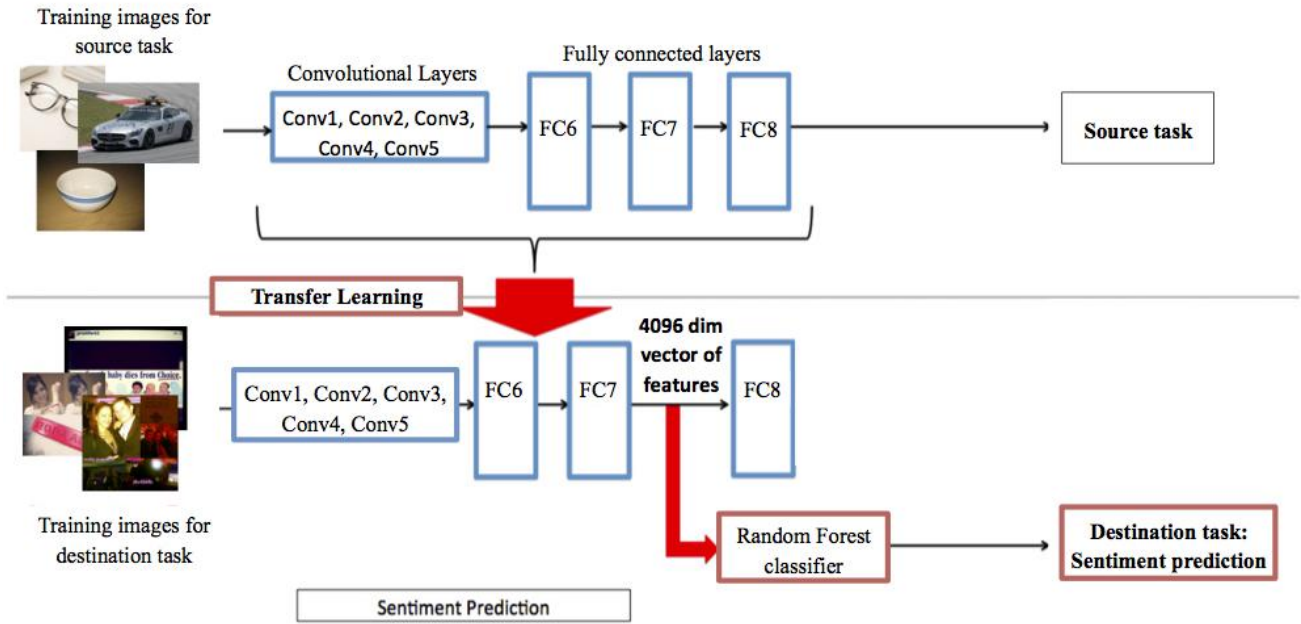
Kim and dos Santos et al. ([33]; [35]) used a convolutional neural network (CNN) to perform sentiment analysis of Twitter messages. Chen et al. [36] trained a deep CNN model called DeepSentiBank to classify visual sentiment concepts. Xu et al. [34] introduced a visual sentiment prediction framework that performs transfer learning from a pre-trained CNN with millions of parameters. You et al. [37] fine-tuned a CNN on a collection of images from Getty Image for visual sentiment analysis and trained a paragraph vector model for textual sentiment analysis. Yuhai et al. [38] also used deep learning to perform sentiment predictions by combining the textual and visual features.

## **3. DEEP CONVOLUTIONAL NEURAL NETWORKS**

Recently, The deep learning framework enables robust and accurate feature learning. Deep learning research has shown its state-of-the-art classification performance in the machine learning field ([39]; [40]) Especially on visual recognition tasks, convolutional neural networks (CNNs) has shown encouraging results outperforming hand-engineered image descriptors such as color histogram, HOG [41], SIFT [42] etc. Inspired by the recent successes of deep learning, we are

interested in solving the challenging visual sentiment analysis

task using deep learning algorithms.



**Fig 1: CNN-F architecture contains 5 convolutional layers (Conv 1–5) and three fully connected layers (FC 6-8)**

The pre-trained CNN we used in this study is the CNN-F architecture presented in Chatfield’s work [43]. As shown in Fig. 1, the CNN-F architecture has 5 convolutional layers followed by 3 fully connected layers, resulting in a total of 8 learnable concatenated layers. Fc6 and fc7 are regularized using dropout [39], while the last layer acts as a multi-way soft-max classifier. The activation function for all weight layers (except for fc8) is the Rectification Linear Unit (RELU) [39], defined in equation (1).

$$f(x) = \max(0, x) \quad (1)$$

Where x is the input to a neuron.

### 3.1 Transfer learning

Training an entire Convolutional Network from scratch with random initialization is a difficult task due to the difficulty of finding a dataset of sufficient size. Instead, it is common to use transfer learning which is to pre-train a CNN on a very large dataset and then transfer the knowledge learnt to a new task to improve learning. Recent works from Torresani et al. [44] and Li et al. [45] show the effectiveness of transfer unsupervised way on relatively small datasets like CIFAR and MNIST by using deep feature representations ([38]; [39]).

## 4. PROPOSED COMPUTATIONAL SYSTEM: VISUAL SENTIMENT PREDICTION USING TRANSFER LEARNING

While Researchers have largely relied on textual sentiment analysis, research on visual sentiment analysis is far behind. Nowadays, people share a lot of content on social media in the form of images. People with different backgrounds can easily understand the main content of an image or video, which makes a picture worth a thousand words. Thus, sentiment analysis of such large-scale visual content can help better understand user opinions toward events or topics. In this section, we focus on the problem of sentiment prediction purely based on the visual information within image tweets.

We propose to develop a suitable architecture for visual sentiment analysis as citizen sensing tool using convolutional neural networks and big data warehousing.

### 4.1 Dataset

Our experiment uses a real world dataset from Twitter (see Fig. 2 for samples). The benchmark [46] includes 603 tweets with photos. Ground truths of sentiment values were obtained by Amazon Mechanic Turk annotation, resulting in 470 positive and 133 negative labels.



**Fig 2: Examples of Twitter images that represent respectively from top to bottom positive sentiment and negative sentiment.**

### 4.2 Methods

We propose a novel framework that efficiently transfers Convolutional Neural Networks (CNN) learned on a large-scale dataset to the task of visual sentiment prediction. In our case, the learned parameters are transferred to the task of sentiment prediction, where the images are from a different domain and the labeled data is limited. The CNN-F was trained on ILSVRC-2012 with the ImageNet dataset [47]

which contains 1.2 million images with 1000 categories, using gradient descent with momentum.

As illustrated in Fig.1, in order to use the pre-trained CNN as a fixed feature extractor, we remove the last fully connected layer of the pre-trained CNN (the layer leading to the output units) and treat the activations of the last hidden layer as a deep feature representation for each 224x224 input image. As is demonstrated by J Donahue [48], the 7th layer (Fc7) output of pre-trained CNN generalizes well to object recognition and detection. In this experiment, we explore its potential capability to higher-level concept understanding, namely sentiment. The deep features are computed by forward propagating through 5 convolutional layers and two fully connected layers in the pre-trained CNN. The output is a 4096-dimensional deep feature vector. The pre-trained model was provided by a MATLAB toolbox called “MatConvNet” [49].

**Algorithm1:** Random Forests

1. For  $b=1$  to  $B$  (where  $B$  is the number of trees):
  - a. Draw a bootstrap sample of size  $N$  from the training data.
  - b. Grow a random-forest tree  $T_b$  to the bootstrapped data by repeating the following steps for each terminal node until the minimum node size  $n_{min}$  is reached.
    - i. Select randomly  $m$  variables from the  $p$  variables.
    - ii. Pick the best variable/split point among  $m$ .
    - iii. Split the node into two nodes.
2. Output the ensemble of trees  $\{T_b\}_1^B$ .
3. Predict new data by aggregating the predictions of the  $B$  trees (i.e., majority votes).

**4.2.1 Feature Selection**

When the number of features is very large as in our case (4096 features), irrelevant features will lead to greater computational cost and sometimes can even decrease accuracy by introducing noise [50]. As such, feature selection algorithms are used to select the right features for our system. The filter based approach for feature selection is usually chosen due to its computational efficiency. Therefore, in current work we have applied filter-based approach where symmetrical uncertainty (SU) [51] is used as a feature goodness measure based on information gain.

Let  $H(X)$  be the entropy of  $X$ ,  $H(X|Y)$  be the entropy of  $X$  after observing values of another variable  $Y$  and  $IG(X|Y)$  be the information gain about  $X$  provided by  $Y$ . A feature  $Y$  is regarded more correlated to feature  $X$  than to feature  $Z$ , if  $IG(X|Y) > IG(X|Z)$ .

$$H(X) = -\sum P(x_i) \log_2(P(x_i)) \tag{2}$$

$$H(X|Y) = -\sum_i P(y_j) \sum_i P(x_i|y_j) \log_2(P(x_i|y_j)) \tag{3}$$

$$IG(X|Y) = H(X) - H(X|Y) \tag{4}$$

Then the symmetrical uncertainty is:

$$SU(X, Y) = 2 \left[ \frac{IG(X|Y)}{H(X)+H(Y)} \right] \tag{5}$$

**4.2.2 Classification**

With one feature selected, we trained two decision tree classifiers: Random Forest and J48 from Weka [52], explained in the next sub-sections.

**4.2.2.1 Random Forest**

Decision trees are a popular decision support tools in machine learning. It uses graphs, which are similar to a tree where the nodes are decisions and their possible consequences. Random forests (RF) or random decision forests [53] is a tree-based ensemble learning used for classification, regression, among others. It has the advantage of being extremely fast, efficient on big data and capable of overcoming the over fitting to training set problem. RF is a voting based ensemble of  $L$  decision trees (DT). Each DT works as an independent classifier and predicts one activity from processing that particular tree. The final activity selected from the algorithm is the one selected by the majority of trees. The random forests algorithm for classification is as described in Algorithm1.

**4.2.2.2 J48 Decision Tree**

The J48 algorithm from Weka is used to implement Univariate Decision Tree approach. It is the implementation of the C4.5 algorithm used to generate a decision tree developed by Ross Quinlan. The general algorithm [54] for building decision trees is:

**Algorithm2:** J48

1. Check for the above base cases.
2. For each attribute  $a$ , find the normalized information gain ratio from splitting on  $a$ .
3. Let  $a\_best$  be the attribute with the highest normalized information gain.
4. Create a decision node that splits on  $a\_best$ .
5. Recur on the sublists obtained by splitting on  $a\_best$ , and add those nodes as children of node.

**4.2.3 Fine-Tuning**

Convolutional Neural Networks (CNNs) contain an enormous number of parameters so they often require large datasets to be trained from scratch. For tasks such as visual sentiment prediction, the size of the datasets is usually constrained due to the wide variability in visual content composing a positive or negative class and because of the difficulty and expense of acquire high-quality labels that depend so much on subjective reasoning [55]. In order to improve the prediction accuracy, we have fine-tuned the features of the pre-trained network and train another classifier on the features for the domain-specific task, which is sentiment prediction. Girshick et al. [40] showed how supervised pre-training followed by supervised fine-tuning of the network is a very effective approach for learning new tasks. Fine-tuning is done by initializing a network with weights optimized for ILSVRC-2012. Then the weights are updated using the target task training set. The learning rate used for fine-tuning is typically set to be less than the initial learning rate used for the full ImageNet dataset. This ensures that the features learnt from the larger dataset are not forgotten.

We replaced the last 1000-dimensional layer (previous fc8) with a 3-dimensional to suit our classification needs. Our 3 classes would be: Positive, Negative and Neutral. The pre-trained network contains several million parameters. Running a fine-tuning on a CPU is time consuming. Therefore, the

forward and backward passes of the networks ran on a GeForce GTX 960. The network was trained by stochastic gradient descent, with a learning rate of 10<sup>-4</sup> and momentum of 0.9. We lower the learning rate by an order of magnitude after every five iterations. This generally results in a slightly better performance for the first iterations.

### 4.3 Baselines

Low-level visual features: It has recently been shown that a set of low-level visual features can be useful for characterizing sentiment clues such as scenes, textures, faces as well as other abstract concepts [46].

Mid-level Feature-based: The second baseline is the SentiBank approach [56] that proposes new concept representation defined as an adjective-noun pair, e.g. happy guy, healthy food, etc. The ontology is constructed based on psychology studies and web mining.

### 4.4 Experimental Results and Discussions

In this section, we evaluate the performance of the transfer learning approach. The evaluation is made with the 5-folds cross validation evaluation strategy, which is standard for our problem scope. The 5-folds cross validation splits the dataset into 5 partitions and evaluates them 5 times. Each time, one of the 5 subsets is used as the test set and the other 4 subsets are grouped together to form a training set. Then, the average error across all 5 runs is computed for final result.

We compare in table 1 the performance of J48 and Random Forest. The results are presented in terms of precision, recall and F1-measure. Each metric is a function of the true positives (TP), false positives (FP), and false negatives (FN). The precision is the ratio of correctly classified positive instances to the total number of instances classified as positive. It is given by,

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

Recall is the ratio of correctly classified positive instances to the total number of positive instances, and is given by,

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

The F1-measure combines precision and recall into a single Value. It is given by,

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

**Table 1: The precision, recall, F1 and accuracy of J48 and RF classifiers**

Classifier	Accuracy	Precision	Recall	F1
J48	0.762	0.669	0.762	0.688
RF	<b>0.763</b>	<b>0.701</b>	<b>0.763</b>	<b>0.709</b>

It can be seen that RF slightly outperforms J48. Given the results we decided to use RF classifier after fine-tuning the network.

We fine-tune the pre-trained neural network in the following way. We randomly divide the Twitter dataset into 5 equal partitions. Every time, we use 4 out of the 5 partitions to fine-tune our pre-trained model and evaluate the new model on the remaining partition. The averaged evaluation results are reported in table 2. The algorithm is detailed below:

**Input:**  $X = \{x_1, x_2, \dots, x_n\}$  a set of images of size  $224 \times 224$   
 $Y = \{y_1, y_2, \dots, y_n\}$  sentiment labels of X  
 Pre-trained network N

1. Split X and Y into 5 equal groups  $\{(X_1, Y_1), \dots, (X_5, Y_5)\}$ .
2. Let  $i=1$ .
3. **do**
  4. Let  $(X_0, Y_0) = (X, Y) - (X_i, Y_i)$
  5. Fine-tune N with input  $(X_0, Y_0)$  to obtain new network  $N_i$
  6. Evaluate the performance of  $N_i$  on  $(X_i, Y_i)$
  7.  $i++$ .
8. **while** ( $i \leq 5$ )
9. **Output** The averaged accuracy of  $N_i$  on  $(X_i, Y_i)$  (where  $i$  in [1-5])

Similar to Borth et al. [46], we also employ 5-fold cross-validation and logistic regression classifier to evaluate the performance of all the baseline algorithms. It can be seen in table 2 that our proposed method of using a pre-trained network as feature extractor (CNN) outperforms both baseline approaches with a prediction accuracy of 77%. Additionally, experimental results showed that domain specific fine tuning is effective for improving the performance of pre-trained neural networks.

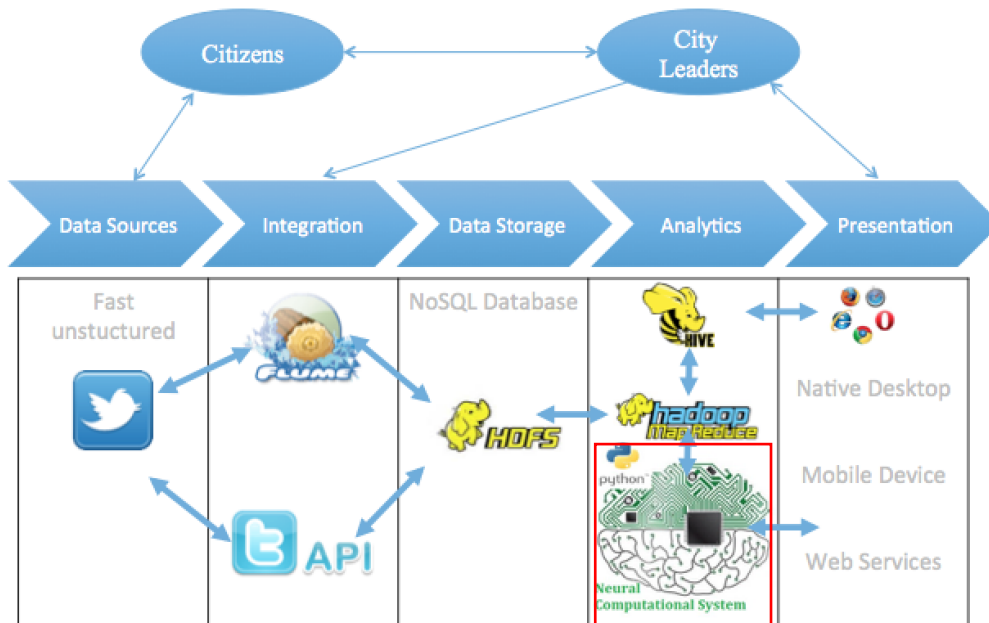
**Table 2: Comparison results of proposed method with baselines**

Method	Accuracy
Low-level Features	0.57
SentiBank	0.70
CNN	0.77
Fine-tuned CNN	<b>0.834</b>

This set of results demonstrates that the proposed system for visual sentiment prediction is capable of efficiently utilizing the power of the pre-trained Convolutional Neural Network. By transferring domain knowledge from the object detection domain to the sentiment prediction domain, and by domain specific fine-tuning we obtained an overall prediction accuracy of 83%. Moreover, the experimental results suggest that convolutional neural networks that are properly trained can outperform both classifiers that use low-level features or mid-level visual attributes for the challenging problem of visual sentiment analysis. We believe that the prediction accuracy of our system can be further improved with more training data, which is the goal of our future work.

## 5. PROPOSED ARCHITECTURE: VISUAL SENTIMENT ANALYSIS AS CITIZEN SENSING TOOL

The leaders of smart cities understand the power of social media sites. These leaders monitor the sentiment about how their city is being run that is expressed. Citizen sensing tool will provides city leaders with an easy to understand summary of the expressed sentiment about the themes of interest to them. City leaders must know the subjects that they want to track sentiment about.



**Fig. 3 The proposed architecture for a citizen sensing tool using our neural computational system**

Citizens interacting with social media websites such as Twitter generate enormous volume of big unstructured data, which refers to information that does not have a pre-defined model and/or cannot fit into relational databases.

Nearly no analytics tool can work directly on unstructured data. As such, there is a vital need to define the basic information/semantic models, architecture components and operational models that together comprise a so-called Big Data Ecosystem with the objective of transforming the unstructured raw Twitter posts into useful information about citizens' sentiments and opinions.

The overall architecture of our proposed system is explained in Fig. 3. Twitter posts can be downloaded and loaded into Hadoop [57] using purpose built tools like Apache Flume [58], which is a distributed system for collecting, aggregating, and moving large amounts of data from multiple sources into HDFS or another central data store. HDFS stores large files by dividing them into blocks (usually 64 or 128 MB) and replicating the blocks on three or more servers. Data can be gathered for free directly from Twitter public application interface. Once the data is loaded into Hadoop the next step is to transform it into a format that can be used for analysis. Data transformation in Hadoop is completed using a process called MapReduce. HDFS provides APIs for MapReduce applications to read and write data in parallel. MapReduce is the programming paradigm that allows to take unstructured data and transform (map) it to something meaningful, and then aggregate (reduce) for reporting. MapReduce jobs can be generated by tools such as Hive [59].

Once MapReduce jobs are done, the preprocessed data can now be loaded into our CNN-based computational system to perform sentiment analysis with Python programming language, which has recently gained popularity as a productive general purpose platform for statistical analysis as well as for web services and interface development. Finally, the result can be presented to city leaders using interactive and user-friendly interfaces.

## 6. CONCLUSIONS AND FUTURE WORK

We believe that sentiment analysis is a key factor in the development of all smart city domains. In this paper, we have designed a novel citizen sensing tool that leverages the potential of visual sentiment analysis. Our computational system uses convolutional neural networks and transfer learning for visual sentiment prediction. We have shown that deep architectures can learn useful features in recognizing visual sentiment in social images. We believe that our proposed technologies make an important contribution to the vision of citizen sensing.

In the future, we are planning to collect more data and study different network architectures to further validate our proposed system. Currently, we are looking at developing robust multimodality models that employ both the textual and visual content to increase the system's performance. Implementing the proposed big data architecture is part of our future work.

## 7. REFERENCES

- [1] "State of the World's Cities 2012/2013." [Online]. Available: <http://mirror.unhabitat.org/pmss/listItemDetails.aspx?publicationID=3387>. [Accessed: 08-Sep-2016].
- [2] T. Stonor, "Smart cities – why, what, how, how?," *The power of the network*, 06-Jun-2013. .
- [3] J. Gabrys, "Programming Environments: Environmentality and Citizen Sensing in the Smart City," *Environ. Plan. Soc. Space*, vol. 32, no. 1, pp. 30–48, Feb. 2014.
- [4] S. Graham and A. Aurigi, "Urbanising cyberspace?," *City*, vol. 2, no. 7, pp. 18–39, May 1997.
- [5] A. Mahizhnan, "Smart cities: The Singapore case," *Cities*, vol. 16, no. 1, pp. 13–18, Feb. 1999.
- [6] L. Anthopoulos and P. Fitsilis, "Exploring architectural and organizational features in smart cities," in *16th*



*International Conference on Advanced Communication Technology*, 2014, pp. 190–195.

- [7] R. E. Hall, B. Bowerman, J. Braverman, J. Taylor, H. Todosow, and U. Von Wimmersperg, “The vision of a smart city,” Brookhaven National Lab., Upton, NY (US), 2000.
- [8] D. Havlik, S. Schade, Z. A. Sabeur, P. Mazzetti, K. Watson, A. J. Berre, and J. L. Mon, “From Sensor to Observation Web with Environmental Enablers in the Future Internet,” *Sensors*, vol. 11, no. 4, pp. 3874–3907, Mar. 2011.
- [9] M. Conti, M. Kumar, and others, “Opportunities in opportunistic computing,” *Computer*, vol. 43, no. 1, pp. 42–50, 2010.
- [10] J. A. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, “Participatory sensing,” *Cent. Embed. Netw. Sens.*, 2006.
- [11] A. Kavanaugh, A. Ahuja, M. Pérez-Quñones, J. Tedesco, and K. Madondo, “Encouraging Civic Participation Through Local News Aggregation,” in *Proceedings of the 14th Annual International Conference on Digital Government Research*, New York, NY, USA, 2013, pp. 172–179.
- [12] B. Liu, “Sentiment analysis and opinion mining,” *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.
- [13] D. Joshi, R. Datta, E. Fedorovskaya, Q.-T. Luong, J. Z. Wang, J. Li, and J. Luo, “Aesthetics and emotions in images,” *IEEE Signal Process. Mag.*, vol. 28, no. 5, pp. 94–115, 2011.
- [14] Q. You, J. Luo, H. Jin, and J. Yang, “Robust image sentiment analysis using progressively trained and domain transferred deep networks,” *ArXiv Prepr. ArXiv150906041*, 2015.
- [15] A. T. Campbell, S. B. Eisenman, N. D. Lane, E. Miluzzo, and R. A. Peterson, “People-centric urban sensing,” in *Proceedings of the 2nd annual international workshop on Wireless internet*, 2006, p. 18.
- [16] N. Gross, “14: The Earth Will Don An Electronic Skin,” *Bloomberg.com*, 30-Aug-1999.
- [17] D. Cuff, M. Hansen, and J. Kang, “Urban Sensing: Out of the Woods,” *Commun ACM*, vol. 51, no. 3, pp. 24–33, Mar. 2008.
- [18] Frost & Sullivan, “Strategic Opportunity Analysis of the Global Smart City Market Report Brochure,” 2013. [Online]. Available: <http://www.frost.com/prod/servlet/report-brochure.pag?id=M920-01-00-00-00>. [Accessed: 10-Sep-2016].
- [19] E. Miluzzo, N. D. Lane, S. B. Eisenman, and A. T. Campbell, “CenceMe—injecting sensing presence into social networking applications,” in *European Conference on Smart Sensing and Context*, 2007, pp. 1–28.
- [20] A. Sheth, “Citizen sensing, social signals, and enriching human experience,” *IEEE Internet Comput.*, vol. 13, no. 4, p. 87, 2009.
- [21] X. Hu, T. H. S. Chu, H. C. B. Chan, and V. C. M. Leung, “Vita: A Crowdsensing-Oriented Mobile Cyber-Physical System,” *IEEE Trans. Emerg. Top. Comput.*, vol. 1, no. 1, pp. 148–165, Jun. 2013.
- [22] B. C. Beaumont, “Mumbai attacks: Twitter and Flickr used to break news,” 27-Nov-2008.
- [23] R. Arunachalam and S. Sarkar, “The New Eye of Government: Citizen Sentiment Analysis in Social Media,” in *Sixth International Joint Conference on Natural Language Processing*, p. 23.
- [24] J. Villena-Román, “TweetAlert: Semantic Analytics in Social Networks for Citizen Opinion Mining in the City of the Future.”
- [25] B. Guthier, R. Alharthi, R. Abaalkhail, and A. El Saddik, “Detection and Visualization of Emotions in an Affect-Aware City,” in *Proceedings of the 1st International Workshop on Emerging Multimedia Applications and Services for Smart Cities*, New York, NY, USA, 2014, pp. 23–28.
- [26] A. Vakali, L. Anthopoulos, and S. Krco, “Smart Cities Data Streams Integration: Experimenting with Internet of Things and Social Data Flows,” in *Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS14)*, New York, NY, USA, 2014, pp. 60:1–60:5.
- [27] D. Osimo and F. Mureddu, “Research challenge on opinion mining and sentiment analysis,” *Univ. Paris-Sud Lab. LIMSI-CNRS Bâtim.*, vol. 508, 2012.
- [28] S. B. Davis and M. Saunders, “How social media can help improve and redesign transport systems,” *The Guardian*, 17-Jun-2014.
- [29] F. Antonelli, M. Azzi, M. Balduini, P. Ciuccarelli, E. D. Valle, and R. Larcher, “City Sensing: Visualising Mobile and Social Data About a City Scale Event,” in *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*, New York, NY, USA, 2014, pp. 337–338.
- [30] Athena Vakali, Despoina Chatzakou, Vassiliki Koutsonikola, and Georgios Andreadis, “Social Data Sentiment Analysis in Smart Environments - Extending Dual Polarities for Crowd Pulse Capturing,” 2013, pp. 175–182.
- [31] A. D. I. Kramer, “An Unobtrusive Behavioral Model of ‘Gross National Happiness,’” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2010, pp. 287–290.
- [32] E. Cambria, “Affective computing and sentiment analysis,” *IEEE Intell. Syst.*, vol. 31, no. 2, pp. 102–107, 2016.
- [33] Y. Kim, “Convolutional neural networks for sentence classification,” *ArXiv Prepr. ArXiv14085882*, 2014.
- [34] C. Xu, S. Cetintas, K.-C. Lee, and L.-J. Li, “Visual Sentiment Prediction with Deep Convolutional Neural Networks,” *ArXiv14115731 Cs Stat*, Nov. 2014.
- [35] C. N. dos Santos and M. Gatti, “Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts.”
- [36] T. Chen, D. Borth, T. Darrell, and S.-F. Chang, “DeepSentiBank: Visual Sentiment Concept Classification with Deep Convolutional Neural Networks,” *ArXiv14108586 Cs*, Oct. 2014.

- [37] Q. You, J. Luo, H. Jin, and J. Yang, “Joint Visual-Textual Sentiment Analysis with Deep Neural Networks,” in *Proceedings of the 23rd ACM International Conference on Multimedia*, New York, NY, USA, 2015, pp. 1071–1074.
- [38] Y. Yu, H. Lin, J. Meng, and Z. Zhao, “Visual and Textual Sentiment Analysis of a Microblog Using Deep Convolutional Neural Networks,” *Algorithms*, vol. 9, no. 2, p. 41, Jun. 2016.
- [39] 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.
- [40] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.
- [41] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, 2005, vol. 1, pp. 886–893 vol. 1.
- [42] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [43] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, “Return of the devil in the details: Delving deep into convolutional nets,” *ArXiv Prepr. ArXiv14053531*, 2014.
- [44] L. Torresani, M. Szummer, and A. Fitzgibbon, “Efficient object category recognition using classemes,” in *Computer Vision—ECCV 2010*, Springer, 2010, pp. 776–789.
- [45] L.-J. Li, H. Su, L. Fei-Fei, and E. P. Xing, “Object bank: A high-level image representation for scene classification & semantic feature sparsification,” in *Advances in neural information processing systems*, 2010, pp. 1378–1386.
- [46] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang, “Large-scale visual sentiment ontology and detectors using adjective noun pairs,” in *Proceedings of the 21st ACM international conference on Multimedia*, 2013, pp. 223–232.
- [47] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 248–255.
- [48] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, “Decaf: A deep convolutional activation feature for generic visual recognition,” *ArXiv Prepr. ArXiv13101531*, 2013.
- [49] A. Vedaldi and K. Lenc, “MatConvNet: Convolutional neural networks for matlab,” in *Proceedings of the 23rd Annual ACM Conference on Multimedia Conference*, 2015, pp. 689–692.
- [50] H. Almuallim and T. G. Dietterich, “Learning With Many Irrelevant Features,” in *In Proceedings of the Ninth National Conference on Artificial Intelligence*, 1991, pp. 547–552.
- [51] L. Yu and H. Liu, “Feature selection for high-dimensional data: A fast correlation-based filter solution.”
- [52] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA Data Mining Software: An Update,” *SIGKDD Explor Newsl*, vol. 11, no. 1, pp. 10–18, Nov. 2009.
- [53] T. K. Ho, “Random decision forests,” in *Proceedings of the Third International Conference on Document Analysis and Recognition, 1995*, 1995, vol. 1, pp. 278–282 vol.1.
- [54] S. B. Kotsiantis, “Supervised Machine Learning: A Review of Classification Techniques,” in *Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*, Amsterdam, The Netherlands, The Netherlands, 2007, pp. 3–24.
- [55] V. Campos, B. Jou, and X. Giro-i-Nieto, “From Pixels to Sentiment: Fine-tuning CNNs for Visual Sentiment Prediction,” *ArXiv160403489 Cs*, Apr. 2016.
- [56] D. Borth, T. Chen, R. Ji, and S.-F. Chang, “Sentibank: large-scale ontology and classifiers for detecting sentiment and emotions in visual content,” in *Proceedings of the 21st ACM international conference on Multimedia*, 2013, pp. 459–460.
- [57] “Welcome to Apache™ Hadoop®!” [Online]. Available: <https://hadoop.apache.org/>. [Accessed: 30-May-2016].
- [58] “Welcome to Apache Flume — Apache Flume.” [Online]. Available: <https://flume.apache.org/>. [Accessed: 19-Sep-2016].
- [59] “Apache Hive™.” [Online]. Available: <https://hive.apache.org/>. [Accessed: 19-Sep-2016].