Where are our Providers?: Scene Recognition based on Locations of Brazilian Government Suppliers

Rodrigo P. Ferreira[1], Dr. Rommel N. Carvalho[2]

***Abstract*—The Observatory of Public Spending (or ODP, in Portuguese) is a special unit of Brazil's Ministry of Transparency, Monitoring and Office of the Comptroller-General (or CGU, in Portuguese) responsible for monitoring public spending and gathering managerial and audit information to support the work of CGU internal auditors. One of the most important themes monitored by this unit is Public Procurements and Government Suppliers which have won these procurement processes. Image analysis of many of these suppliers headquarters revealed suspicious landscapes, such as rural areas, isolated places or slums. These landscapes could be an indication of fake suppliers with poor capacity of delivering public goods and services. However, checking thousands of landscapes in order to find these fake suppliers would be a very expensive task. Our objective then is to discover what are the possible groups of scenes involving government suppliers, given that these images were not previously labeled, as automatically as possible. For that reason, we used Places CNN, a pretrained convolutional neural network for scene recognition presented by Zhou et al. [2], which was trained on 205 scene categories with 2.5 million images, for scene recognition on Brazilian Government Suppliers.**

1. INTRODUCTION

According to the Appendix I of the Decree 8910/2016 from the Presidency of the Federative Republic of Brazil [15], the Ministry of Transparency, Monitoring and Office of the Comptroller-General (CGU, in Portuguese) is a Federal Agency in the Executive Branch of the Brazilian Government which has the following functions: Public Assets Defense, Internal Control, Public Audit, Disciplinary Action, Prevention of Corruption, Ombudsman and Transparency.

In order to assist the agency in its decision making process, the good use of the available information is essential. To support this demand for information, in 2008, the Observatory of Public Spending (ODP, in Portuguese) was created. The ODP, according to De Barros and Camargo [11], is a special unit which works as a laboratory that monitors the government expenditure, making intensive use of Data Analysis. Many themes are monitored by this unit, such as government procurement, government corporate cards, outsourcing, and daily fees and flying tickets paid to public agents.

Public procurement, however, remains as one of the most important themes monitored by ODP, given the amount of money involved. More than 80 billion reais, or 24.7 billion dollars[1], were registered last year in *Sistema de Administração de Serviços Gerais* (SIASG), which is the main public purchase system in Brazil.

To monitor public procurement, ODP gathers mainly two types of information: Managerial and Audit. Managerial information offers a general vision about how the government makes its purchases, distributed by month or state, for instance. Therefore, this kind of information can help public managers to recognize trends in government spending, supporting their decision making process. Audit information is usually the result of applying data matching techniques between many different government databases to detect signals of fraud or misuse of public resources, offering an indicator of risk that can support the work of the federal auditors. Some examples of alerts generated by data matching techniques are: connections between companies and public agents, connections between companies who are competing in the same procurement process, agencies that split purchases in order to avoid the respective procurement procedure, companies that were recently created but already won procurement involving high values, and bidders who did not win the public auction despite giving the best bid.

In many of these alerts, it is possible to collect images of the companies headquarters that were involved in the procurement and became government suppliers. One way to do this is to search for the company address in Google Maps or Google Street View[2]. In real auditing procedures, many images found by these searching mechanisms revealed suspicious landscapes, such as rural areas, isolated places or slums.

[1]*Data Scientist at Observatory of Public Spending (ODP), Department of Research and Strategic Information (DIE), Brazil's Ministry of Transparency, Monitoring and Office of the Comptroller-General (CGU), SAS, Quadra 01, Bloco A, Edifcio Darcy Ribeiro, Braslia, DF, Brazil* odp@cgu.gov.br.

[2] *Dr. Rommel N. Carvalho is a Professor on the Applied Computer Science Masters at University of Brasilia (UnB) and Postdoctoral Fellow at George Mason University (GMU), USA, in the areas of artificial intelligence, data mining, uncertainty and knowledge discovery.*

1. According to the data available in CGU's Sistema de Administração de Serviços Gerais Data warehouse (DW SIASG) in August 30, 2016. Values equal or above 1 billion reais per item were excluded from the query due to common errors in data quality.

2. Google Maps is a free search and visualization service on the Web for maps and satellite images from Earth supported and developed by Google (maps.google.com). Google Street View is a feature in Google Maps that provides panoramic views of locations and streets on the ground level (www.google.com/streetview/)

These suspicious landscapes are a relevant information for the auditors. They might indicate the existence of fake suppliers, which will be unable to deliver public goods and services with the expected quality previously established in contract.

However, given the government budget restrictions and lack of human resources, viewing thousands of images from all government suppliers would be a very expensive effort. Thus, it is essential to find a method to automatically analyze the landscapes of those images, applying machine learning for scene recognition, as proposed by the works of Oliva and Torralba [3], in all government suppliers images and reducing this set of images to a much smaller and suspicious one, which then could be more easily and efficiently analyzed by CGU auditors.

Nowadays, there are no labels for all the possible landscapes that could be applied to images of Brazilian Government suppliers. Furthermore, creating such labels would be an expensive and time consuming task. Therefore, we will use PlacesCNN, a pretrained convolutional neural network (CNN) for scene recognition presented by Zhou et al. [2], to classify these places, a model which considerably improved the state-of-the-art technique for Image Classification, as explained by LeCun, Bengio and Hinton [13].

This article contains the following structure: an overview of the theoretical framework which will give us support to the development of this project (Section 2), related works on scene recognition (Section 3), steps taken according to the Cross Industry Standard Process for Data Mining (CRISP-DM), proposed by Wirth and Hipp [1] (Section 4), analysis of the obtained results (Section 5), and a brief conclusion with suggestions for further work (Section 6).

2. THEORETICAL FRAMEWORK

In this section, we present an overview of the main subjects we are going to work with to support the development of this project.

2.1. The CRISP-DM methodology

This work is going to follow the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology. According to Wirth and Hipp [1], this methodology is composed of four levels of abstraction: phases, generic tasks, specialized tasks, and process instances. The phases are in the most abstract level and resume the Data Mining Process life cycle: *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment*.

2.2. Convolutional Neural Networks (CNN)

According to LeCun et al. [12] (Section II), traditional feed-forward neural networks have difficulties in dealing with the variability of images, such as translations and distortions, unless they have a big number of layers applying almost the same weight to many different parts of the image. The convolutional neural networks (CNN) were created to solve this kind of problem. The main difference between CNN and traditional neural networks is that the former applies two operations in the image features before performing the classification task: *convolution* and *sub-sampling*.

LeCun et al. proposed in [12] a convolution operation that divides and scans parts of the image with a local group of units called receptive fields. These fields must apply the same weights in all parts of the image. The main objective is to obtain and combine similar features that are spread in different places of the same image.

The other operation proposed by LeCun et al. [12] was a sub-sampling operation, also known as pooling, which tries to reduce the image dimensions, given that a great part of an image contains irrelevant information for the classification task and also causes an unnecessary amount of layers that might affect the model, making it more sensible to noise.

The convolution and sub-sampling operations can be repeated for many layers in the network until the output matrix achieves a reasonable degree of invariance to transformations, making it a more suitable input to the classification layers (known as fully-connected layers). As stated by LeCun et al. [12]: "A large degree of invariance to geometric transformations of the input can be achieved with this progressive reduction of spatial resolution compensated by a progressive increase of the richness of the representation".

2.3. The Places-CNN

To classify the scenes of Brazilian Government Suppliers, we will use a pretrained convolutional neural network proposed by Zhou et al. in [2], the Places-CNN, which has almost the same architecture of object recognition networks (e.g. ImageNet-CNN [7], proposed by Krizhevsky, Sutskever and Hinton). The main difference between them is in the way these networks were trained: Places-CNN was trained with scene-centric data whereas object recognition networks were trained with object-centric data. According to Zhou et al. [2], this difference was enough to change the features that these networks learn.

In [2], the Places-CNN was trained with about 2 million images from 205 categories. Each category contains between 5,000 and 15,000 images.

3. RELATED WORK

According to Oliva and Torralba [3], in the earlier days of computer vision, scene recognition was seen as a result of the integration of many different local objects (like shapes, angles, surfaces) that belonged to a certain image. The first model to propose an holistic representation of scenes as a computational model was the Space Envelope representation proposed by Oliva and Torralba in 2001 [3], which tries to capture the global properties of a scene (whether the scene is man-made or natural or how vast a scene is), rather than analyzing the image simply by its objects or regions.

In the end of the last decade, feature-based scene recognition became popular, as stated in the introduction of Zou et al. 2016 [10]. One of the main attempts in this direction was proposed by Fei-Fei and Perona (2005) [4], which decomposes a set of images into attributes called *codewords*. The objective of the training stage is to build a *codeword* distribution to each group of images. In the testing stage, a particular image is decomposed in *codewords* and then compared to the *codeword* distributions built in the training stage. The *codewords* are determined by the model, which means that the machine learning algorithm is unsupervised.

According to Zou et al. 2016, section 1 [10], the Bag-of-Visual-Words (BoVW) presented by Yang, Jiang and Hauptmann in [5] was another very popular technique for scene recognition in the last decade. This representation is similar to the ones used in text classification tasks, consisting of grouping local interest points in the image. These groups are treated like "visual words" in a bag of words representing the image. This bag of words is represented as a vector containing the frequency (histogram) of each visual word in the image document, which could them be processed like a feature vector. A disadvantage of this model, according to Lazebnik, Schmid and Ponce, 2006 [6], is that it cannot represent the spatial structure of the visual data, consequently there is considerable loss of information.

The Spatial Pyramid Matching Kernel (SPMK), proposed by Lazebnik, Schmid and Ponce in [6], was an improvement from the BoVW model, addressing the problem of lack of spatial representation. Like the BoVW model, it also represents the image as a bag of features, but it does it in many subdivisions of the image, so it can achieve an approximate geometric matching.

In 2012, there was a significant change in the state-of-the-art of Computer Vision when Convolutional Neural Networks were applied in image classification problems by Krizhevsky, Sutskever and Hinton [7]. They achieved a winning error rate of 15.3% in the ImageNet ILSVRC-2012 object recognition competition, in contrast to the 26.2% from the second-best achievement by Harada et al. [14].

Nowadays, the state-of-the-art algorithms for scene classification are divided between applications of CNN (e.g. supervised boosting with multiple CNN in [8], an unsupervised approach with CNN in [9] and the scene-centric Places CNN [2]) and improvements of older methods that involve global and local bag of features (e.g. [10], which merges the BoVW and SPMK models seen earlier with global descriptors of the image).

For the purpose of this work, we could use simpler methods like Bag-of-Visual-Words (BoVW) or Spatial Pyramid Matching Kernel (SPMK) to discover the clusters in suppliers landscapes, label the resulting groups and finally apply supervised learning to the scenes using CNN. However, given the availability of the scene-centric Places CNN, already trained with 205 scene categories and 2.5 million images [2], we find it reasonable, as a first step, to analyze the results of applying this pretrained model to Brazilian Government suppliers which had or have active contracts in 2016.

4. METHODOLOGY

In this section, we present the steps taken to obtain our results, following the phases specified by the Cross Industry Standard Process for Data Mining (CRISP-DM) [1]. The Business Understanding phase was covered by the introduction section (Section 1).

Therefore, the next sections will cover the following: Data Understanding, Data Preparation, Modeling and Evaluation (which will be covered in the Results, in section 5). We consider that, given the early stages of this work, the Deployment phase will be a subject for future works.

4.1. Data Understanding

To obtain the scenes in which government suppliers are located, one alternative would be to obtain their specific coordinates. Unfortunately, this information is currently unavailable in our databases. However, by matching data from suppliers ID (in Brazil, called "CPF" or "CNPJ") in our main public purchase system (in Portuguese, *Sistema de Administração de Serviços Gerais* - SIASG) with data about Brazilian companies in Brazilian Federal Revenue Office (in Portuguese, RFB) database, we can successfully obtain suppliers addresses, which, in most cases[3], will be enough to enable us to get the desired landscapes.

As defined by the Decree 8910/2016 [15], CGU is a Federal Agency of the Executive branch of the Brazilian government. As a consequence, data from suppliers contracted by subnational entities (e.g. states and municipalities) or by Legislative or Judiciary branches of the government will be excluded from our analysis.

For the purpose of this work we will only consider government suppliers which had or have contracts in 2016. This means suppliers whose contracts started or finished in 2016 or that are still active in this year. We also have to consider that most contracts are subject to amendments, so if a contract initially expires in 2015, but due to an amendment this expiration date was postponed to 2016, then we should consider this contract in the scope of this work. To perform this task, we had to discover where in our SIASG database the expiration date is and, in case of an amendment, where is the final corrected expiration date.

Once found the described fields to apply these filters on our scope, we have obtained data about 29,299[4] government suppliers in 110,441 contracts, including suppliers IDs, names and addresses, contract numbers, initial and expiration dates, the values involved (in reais) and information about the government contractor.

3. In some cases, these addresses can be inaccurate or incomplete, which may lead to wrong or non-existent locations

4. Data obtained from CGU's SIASG and RFB Data warehouses in September 28, 2016.

4.2. Data Preparation

To the data preparation phase, we loaded the suppliers data in an R environment[5].

We first converted the encoding of all character types from the imported data to "latin1", to make sure that the specific characters of the Portuguese language will be interpreted correctly by our preparation script. We have also removed all the control characters[6].

After fixing the encoding issue, we searched for abnormalities (or outliers) in our data. These outliers may indicate that some error occurred (e.g. some records were not filled correctly in the system) and therefore are not very trustworthy. It is a reasonable strategy to exclude these kind of data from our work. One example found in our data was the occurrence of contracts with surprisingly long durations (about 100 or 200 years), which probably indicate some mistake during the input process.

The date columns are essential to our scope definition (as seen in the Data Understanding section 4.1, we will only consider suppliers which had or have contracts with the Federal Government in 2016), so we have to make sure that all the dates we use will be as reliable as possible. In order to do this, we had to do the following:

- Remove contracts without initial or final dates;
- Remove contracts where the final dates are lower than the initial dates;
- Remove contracts where the amendment final dates are lower than the original final dates;
- Remove contracts where the duration is too long (by anomaly detection using k-means clustering [16]).

For the last item on the list, we have built two k-means clustering models to identify the contracts with anomalous duration (see Figure 1). One is for the duration considering the time lapse between the initial date and original final date, resulting in 5 clusters where 2 of them, with the higher mean values of duration in number of days, were considered outliers. The second one considers the time lapse between the initial date and the final date given by the amendment, resulting in 4 clusters where 2 of them, with the higher mean values of duration in number of days, were considered as outliers.

After this preparation, we obtained 29,162 suppliers in 110,441 contracts with valid dates, all extracted from our original data obtained in CGU's SIASG and RFB Data warehouses.

Starting from the addresses of these 29,162, it was possible to download the scenes we needed from calling the

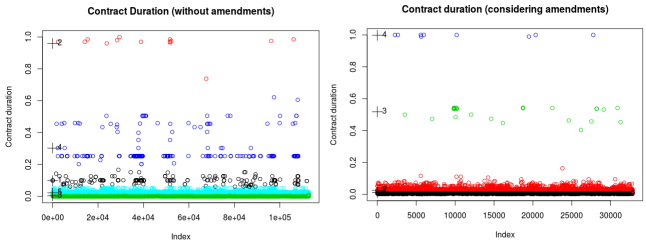

Figure 1. Results of k-means clustering on contract durations. The values were (0-1) normalized. In the first image (Contracts without amendments), clusters 2 (red) and 4 (blue) were removed as outliers. In the second image (Contracts with amendments), clusters 4 (blue) and 3 (green) were removed as outliers.

Google Street View API[7]. However, due to errors in the addresses, nonexistence of the referenced addresses in Google Street View or even possible errors during transmission, not all images were found. These problems generated some blank images or images with very low level of information. We had to exclude them from our work, removing all the images with size less than 10KB.

In the end of the Data Preparation phase, considering the discard of the contracts and images explained earlier, we obtained a resulting set of 27,414 supplier's landscapes, all identified by the supplier's ID ("CPF" or "CNPJ").

4.3. Modeling

Our model objective is to classify the suppliers landscapes. Given that we have already obtained the images we need, we are now ready to test them with the Places CNN [2]. With this objective in mind, we loaded a Places CNN instance in an Ubuntu 14.04 server hosted from Amazon Web Services (AWS)[8].

The Places CNN is a convolutional neural network with an architecture similar to the one proposed by Krizhevsky, Sutskever and Hinton [7] in the ImageNet ILSVRC-2012 object recognition competition data set. It is composed of a sequence following layers (from input to output): "input, conv1, pool1, norm1, conv2, pool2, conv3, conv4, conv5, pool5, fc6, fc7, fc8, output", where the "conv" layers apply convolution operations, the "poll" layers apply subsampling, the "norm" layer applies a lateral response normalization (following the method described in Section 3.3 of Krizhevsky, Sutskever and Hinton [7]) and the final "fc" are the fully-connected layers that in fact are responsible for the classification task.

To apply this model to our scenes, we had to apply some transformations in our data, such as:

5. From https://www.r-project.org: "R is a free software environment for statistical computing and graphics."

6. See "Regular Expressions as used in R", from R Documentation: "Control characters. In ASCII, these characters have octal codes 000 through 037, and 177 (DEL). In another character set, these are the equivalent characters, if any."

7. Application Program Interface (API) is a set of subroutine and programming patterns to access a software or a service hosted on the web. The API for Google Street View is available in https://developers.google.com/maps/documentation/streetview. All the scene images were downloaded between September 28 and 29

8. Amazon Web Services (AWS) is a collection of cloud computing services offered by Amazon.com, a large e-commerce company based in Seattle, Washington, USA, available in https://aws.amazon.com

- Subtract the mean image from all the input images (image normalization[9]);
- Compute the transpose of the image matrix (this is expected by the network input);
- Change the image format from RGB standard to the model's standard (BGR image);
- Normalize the image to 0-255 scale; and
- Convert image size to 227x227 (also expected from the model).

After applying these preprocessing steps, we were ready to run the Places CNN with our images. The next section 5 will give an overview of the results found.

5. Results

At the moment, we have classified 14,997 landscapes with Places CNN from the 27,414 obtained through the previous steps. More images must be added in future works.

The most frequent landscapes classified by the network are in presented in Figure 2.

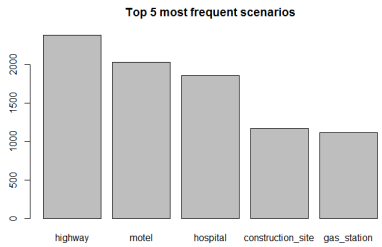

Figure 2. Most common landscapes classified by Places CNN

The amount of highway landscapes might be due to the fact that, in Google Street View, the photos usually show the streets by default. Perhaps changing the orientation of the images downloaded from the API would produce more interesting results (which should be done in future works).

One interesting result is that, from the classified images, we found 103 scenes classified as "slum". Many of these scenes would not be considered as slums in the Brazilian context (see Figure 3). These results suggest that we must create our own labels, considering only Brazilian landscapes for this purpose. A validation of these labels must also be conducted by CGU auditors or through opening a task in Amazon Mechanical Turk[10].

6. Conclusion and next steps

Even without designing and training a specific neural network, the classification of Brazilian Government supplier's scenes using a pretrained network like Places CNN [2] already gave us some interesting results. However, there is a lot to improve if we would like to transform this work into an effective tool for CGU's auditors.

Figure 3. Government Suppliers landscapes classifies by Places CNN. The first column (A) shows scenes correctly classified as "desert" and "forest". The second column (B) shows incorrect results, classified as "prison" and "slum". One should note that the misclassification may be due to the bars in the first image and perhaps due to the post and electrical cables in the second one - which is common in Brazilian streets.

As suggestions for further work to address these problems, we propose the following:

- Change the image orientation when downloading the scenes from Google Street View API;
- Change the image size downloaded from Street View API to get the maximum of information we can and change the standard input size from the neural network, to process larger images; and
- Apply image clustering to the government suppliers scenes through simpler methods (like Bag-of-Visual-Words or Spatial Pyramid Matching Kernel) or apply transfer learning[11] from Places CNN learned features, labeling the resulting groups considering the Brazilian context and using them to train the same Places CNN used in this work.

The given suggestions may help us to generate classes that represent these landscapes more accurately, thus giving the CGU's auditors a more reliable classification to work with.

References

[1] R. Wirth and J. Hipp, CRISP-DM: Towards a standard process model for data mining, in Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining, 2000, pp. 2939.

[2] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, Learning deep features for scene recognition using places database, in Advances in neural information processing systems, 2014, pp. 487495.

9. For details on how this normalization is implemented, see http://caffe.berkeleyvision.org/gathered/examples/imagenet.html

10. Amazon Mechanical Turk is a crowd-sourcing web tool where researchers can open a well defined Human Intelligence Task (HIT) which can be done by an on-line workforce around the globe: https://www.mturk.com/mturk/welcome.

11. According to the Broad Agency Announcement (BAA) 05-29 of Defense Advanced Research Projects Agency (DARPA)s Information Processing Technology Office (IPTO) [17], transfer learning is "the ability of a system to recognize and apply knowledge and skills learned in previous tasks to novel tasks".

[3] A. Oliva and A. Torralba, Modeling the shape of the scene: A holistic representation of the spatial envelope, International journal of computer vision, vol. 42, no. 3, pp. 145175, 2001.

[4] L. Fei-Fei and P. Perona, A bayesian hierarchical model for learning natural scene categories, in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05), 2005, vol. 2, pp. 524531.

[5] J. Yang, Y. Jiang, A. G. Hauptmann, C. Ngo, Evaluating bag-of-visual-words representations in scene classification, in Proceedings of the international workshop on Workshop on multimedia information retrieval, ACM, 2007, pp. 197-206.

[6] S. Lazebnik, C. Schmid, and J. Ponce, Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, in 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR06), 2006, vol. 2, pp. 21692178.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, Imagenet classification with deep convolutional neural networks, in Advances in neural information processing systems, 2012, pp. 10971105.

[8] F. Zhang, B. Du, and L. Zhang, Scene Classification via a Gradient Boosting Random Convolutional Network Framework, IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 3, pp. 17931802, Mar. 2016.

[9] Y. Li, C. Tao, Y. Tan, K. Shang, and J. Tian, Unsupervised Multilayer Feature Learning for Satellite Image Scene Classification, IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 2, pp. 157161, Feb. 2016.

[10] J. Zou, W. Li, C. Chen, and Q. Du, Scene classification using local and global features with collaborative representation fusion, Information Sciences, vol. 348, pp. 209226, Jun. 2016.

[11] A. De Barros, T. Camargo, Transparency and Control of Government Spending in Brazil: The Role of the Public Expenditure Observatory, Open Government and Targeted Transparency: Trends and Challenges for Latin America and the Caribbean, p. 87, Nov. 2012.

[12] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE, v. 86, n. 11, pp. 2278-2324, Nov. 1998.

[13] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature, vol. 521, no. 7553, pp. 436444, May. 2015.

[14] Harada, Tatsuya, and Y. Kuniyoshi. Graphical Gaussian vector for image categorization. Advances in Neural Information Processing Systems, pp. 1547-1555, 2012.

[15] Presidency of the Federative Republic of Brazil, Decree of the Presidency of the Federative Republic of Brazil No. 8910, Nov. 22, 2016.

[16] Hartigan, John A., and Manchek A. Wong. Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28.1, pp. 100-108, 1979.

[17] Pan, Sinno Jialin, and Qiang Yang. A survey on transfer learning. IEEE Transactions on knowledge and data engineering 22.10, pp.1345-1359, 2010.