Automated User-Generated Content Curation with Deep Learning Techniques

PhD Thesis Research Plan

Stalin Leonel Cruz

Supervisor: Dr. Ruben Tous
Dra. Beatriz Otero

Department of Computer Architecture
Universitat Politècnica de Catalunya

This Research Plan is submitted for the degree of
Doctor of Philosophy

Universitat Politècnica de Catalunya May 2017

Table of contents

Chapter 1

Introduction

1.1 Motivation

Nowadays, there is a growing interest in exploiting the photos that users share on social networks such as Instagram or Twitter [5], a part of the so-called user-generated content (UGC). On the one hand, users' photos can be analyzed to obtain knowledge about users behavior and opinions in general, or with respect to a certain products or brands. On the other hand, some users' photos can be of value themselves, as original and authentic content that can be used, upon users' permission, in different communication channels. However, discovering valuable images on social media streams is challenging. The amount of images to process is huge and, while they help, user defined tags are scarce and noisy. The most part of current solutions rely on costly manual curation tasks over random samples. This way many contents are not even processed, and many valuable photos go unnoticed.

Some works, such as [3], [4] or [1], propose scene-based and object-based image recognition techniques to enrich the metadata originally present in social media images in order to facilitate their processing. Automatically tagging the incoming images with tags that describe their semantics (e.g. "beach", "car", etc.) enables more expressive filters and search conditions, minimizing manual curation. This way the time between the image publication and its detection and usage is minimized, the quality of the resulting photos increases (as more photos are analyzed and only the best ones go through manual curation) and the cost is reduced.

Adoption of image recognition techniques in commercial UGC systems is currently very limited. In the best case they provide generic classifiers whose categories and original training data were not specific to UGC. Often these classifiers limit to the categories of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), but the vast majority of Instagram/Twitter photos are people-centric (selfies, food, clothes, etc.) while ILSVRC is

more generic (fauna, flora, etc.). An important particularity of UGC is the huge amount of *spam images*, i.e. images that, in the most usage scenarios, has no value neither as a knowledge carrier nor as a exploitable content. The incapacity of detecting the multiple types of spam images limits the usability and efficiency of existing solutions. Another difficulty of adopting image recognition techniques into UGC systems is the high computational cost of CNN-based image classifiers and object detectors. These systems need to process incoming streams of hundreds of images per second and a very volatile traffic. Any additional processing component need to be extremelly efficient and scalable.

For all these reasons, the development of UGC systems that overcome the constraints imposed by manual curation (cost, qualty and delay time) requires the development of efficient and scalable image recogition pipelines, tailored for the particularities of UGC.

[FEEL FREE TO CHANGE/REMOVE/ADD ANYTHING...]

[YO AÑADIRÍA UNA FIGURA CON PINTA "CIENTÍFICA", LO HABLAMOS]

1.2 Goals

In this thesis, our general goal is to improve some techniques currently applied for automatically annotating photos posted by social media users. The generated annotations facilitate searching, filtering and analyzing user-generated content (UGC). We will apply scene-based and object-based image recognition techniques to extend the metadata originally present in the images with tags that describe their visual content. In order to reach state-of-the-art precission values we will employ CNNs. Performing these tasks in the UGC scenario implies some problems that are currently constraining their applicability to commercial systems. The specific goals of the project, that aim solving or mitigating some of these problems, are the following:

- Development of an efficient and scalable solution to enable complex image recogition pipelines for automatic UGC curation, where multiple CNN-based classifiers and object detectors can be chained in a graph fashion. The resulting solution must provide the necessary throughtput under reasonable cost assumptions, and must be able to scale.

- Development of novel UGC-specific scene-based and object-based image recognition models, including but not limited to, a complete set of Instagram spam image detection models.

- Development of a solution for the scalable training of on-demand custom classifiers and object detectors (e.g. company logos). Because of the particularities of these

models, the resulting solution need to minimize training times and to be able to deal with small datasets.

- Examine the impact of different configurations and derive conclusions aiming to pave the way towards systematic and optimized methodologies for automatic UGC curation.

Chapter 2

State of the art

2.1 Annotating social network images with image recognition techniques

The goals of this work are related to recent works attempting to facilitate the classification and search of images in social networks. Works such as [3], [4] or [1] apply image recognition techniques to automatically annotate the images.

[MORE ABOUT IMAGE RECOGNITION IN GENERAL. DON'T NEED TO DETAIL OLD APPROACHES BUT YOU CAN MENTION THEM...]

2.1.1 Deep Convolutional Neural Networks

All latest works rely on deep convolutional neural networks (CNNs) as an underlying technique.

[MORE ABOUT CNNs IN GENERAL... CITAR 2 o 3 ARTíCULOS MÁS RELEVANTES]

2.1.2 Transfer Learning for object recognition

When the amount of available data is not enough to properly train a CNN (because of the excessive overfitting) one possible solution is applying a Transfer Learning approach the same way Berkeley researches do in [2]. This approach could be relevant for our goal of finding a solution for the scalable training of on-demand custom classifiers (e.g. company logos). In addition to solving the problems related to small datasets, this approach has also advantages when one need to train many models and apply all of them in real-time.

[EXPLAIN TRANSFER LEARNING...]

2.1.3 Object detection

[OVERFEAT...]

2.2 Distributed CNNs

[LO DE TU TFM, DL4J, TENSORFLOW, etc....]

Chapter 3

Methodology

3.1 Project phases

We aim to address all the above mentioned goals by setting out following stages:

Phase 1: Development of novel UGC-specific image recognition models Select the proper CNN setup, acquire and preprocess the data and train a set of UGC-specific scene-based and object-based image recognition models, including a complete set of Instagram spam image detection models. Evaluate the performance and scalability of the combined execution of the resulting classifiers and detectors in real-time. Develop strategies that enable improving the performance and scalability of the solution. Evaluate the suitability of the solution when applied to custom models that need to be trained on-demand with small datasets.

Phase 2: Development of an arquitecture to enable complex image recogition pipelines for automatic UGC curation Taking as a starting point the setup resulting from Phase 1, it will be developed an efficient and scalable solution to enable chaining multiple classifiers and object detectors in a graph fashion. The resulting architecture will enable the definition of more complex image recogition pipelines for automatic UGC curation.

Phase 3: Evaluation of the performance and scalability of the obtained solution enabling complex image recogition pipelines in a real scenario Define a real scenario and evaluate the performance and scalability of the results of Phase 2 under stress conditions. Examine the impact of different configurations and derive conclusions aiming to pave the way towards systematic and optimized methodologies for automatic UGC curation.

3.2 Associated work plan

I plan to finish my Ph.D in a total of three years. A tentative work plan of my future activities can be found in Figure 3.1.

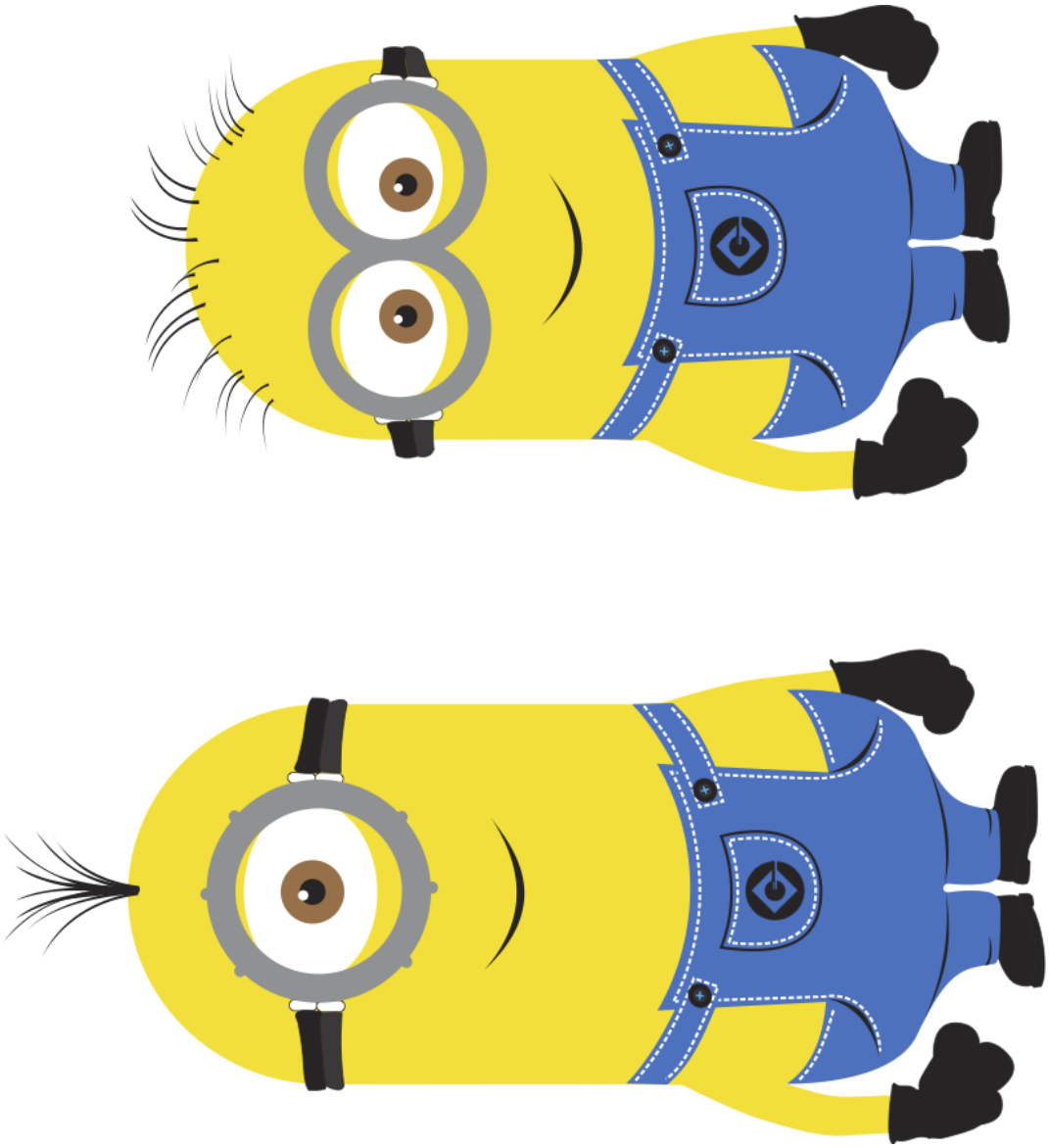

Fig. 3.1 Here you can place the GANTT diagram as an image

- J1: This publication will describe the ..TODO.
- C2: This publication will evaluate ...TODO...
- C3 : This publication will detail...TODO.
- J2: This publication will summarize previous contributions on a real scenario...

3.3 Completed activities

In this last year, the first six months have been principally spent studying the topic and having theoretical research. Then I implemented .....

[DESCRIBE HERE THE WORK WITH SPARK4MN + DL4J. CLARIFY WHEN THE TESTS FINISHED (WHEN YOUR ACCOUNT WAS CLOSED)]

3.4 Resources

3.4.1 Data

[TODO. Clarify that enven if the public API will change you still have the enough data in a save place]

3.4.2 Hardware

[TODO. Mention Adsmurai machines and also machines from the CAP research group.]

References

[1] Denton, E., Weston, J., Paluri, M., Bourdev, L., and Fergus, R. (2015). User conditional hashtag prediction for images. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, pages 1731–1740, New York, NY, USA. ACM.

[2] Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., and Darrell, T. (2014). Decaf: A deep convolutional activation feature for generic visual recognition. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, pages 647–655.

[3] Park, M., Li, H., and Kim, J. (2016). HARRISON: A benchmark on hashtag recommendation for real-world images in social networks. *CoRR*, abs/1605.05054.

[4] Tous, R., Torres, J., and Ayguadé, E. (2015). Multimedia big data computing for in-depth event analysis. In *BigMM*, pages 144–147. IEEE.

[5] Tous, R., Wust, O., Gomez, M., Poveda, J., Elena, M., Torres, J., Makni, M., and Ayguade, E. (2016). User-generated content curation with deep convolutional neural networks. IEEE.