

OBSERVATÓR!O2016

Exploring Rio-2016 image dataset through Deep Learning and visualization techniques



JULIA R GIANNELLA* VISGRAF LAB, IMPA (Rio de Janeiro, Brazil) | juliagiannella@gmail.com

LUIZ VELHO VISGRAF LAB, IMPA (Rio de Janeiro, Brazil) | lvelho@impa.br

INTRODUCTION

OBSERVATÓR!O2016 is a web-based platform for collecting, structuring and visualizing the online response to Rio-2016 from content shared on Twitter. Its website (<http://oo.impa.br>) provides an environment to explore comments and images about the Olympics through data visualizations. This poster, however, focuses on an unfolding project regarding an exploratory research to delve into large-scale visual data. We used Deep Learning and media visualization techniques to enhance discoverability and expression of subject features (Olympic torch and sports disciplines) within our image collection. This project was developed at the Vision and Computer Graphics Laboratory (VISGRAF) based at the Institute of Pure and Applied Mathematics (IMPA), from Rio de Janeiro, Brazil.

METHOD

OBSERVATÓR!O2016 collected around 180K unique images related to the Olympics from April 18th to August 25th, 2016. This large amount of data allied with the right tools for exploring it inaugurates new avenues for researchers to organize image collections and, potentially, discover visual trends in them. Aware of that, we conducted a first investigation attempt to enhance discoverability and expression of Rio-2016 dataset. We did so through the combination of two digital techniques: 1) Deep Learning to classify the images into subject features we've previously elected and 2) media visualization techniques to present the images from each category.

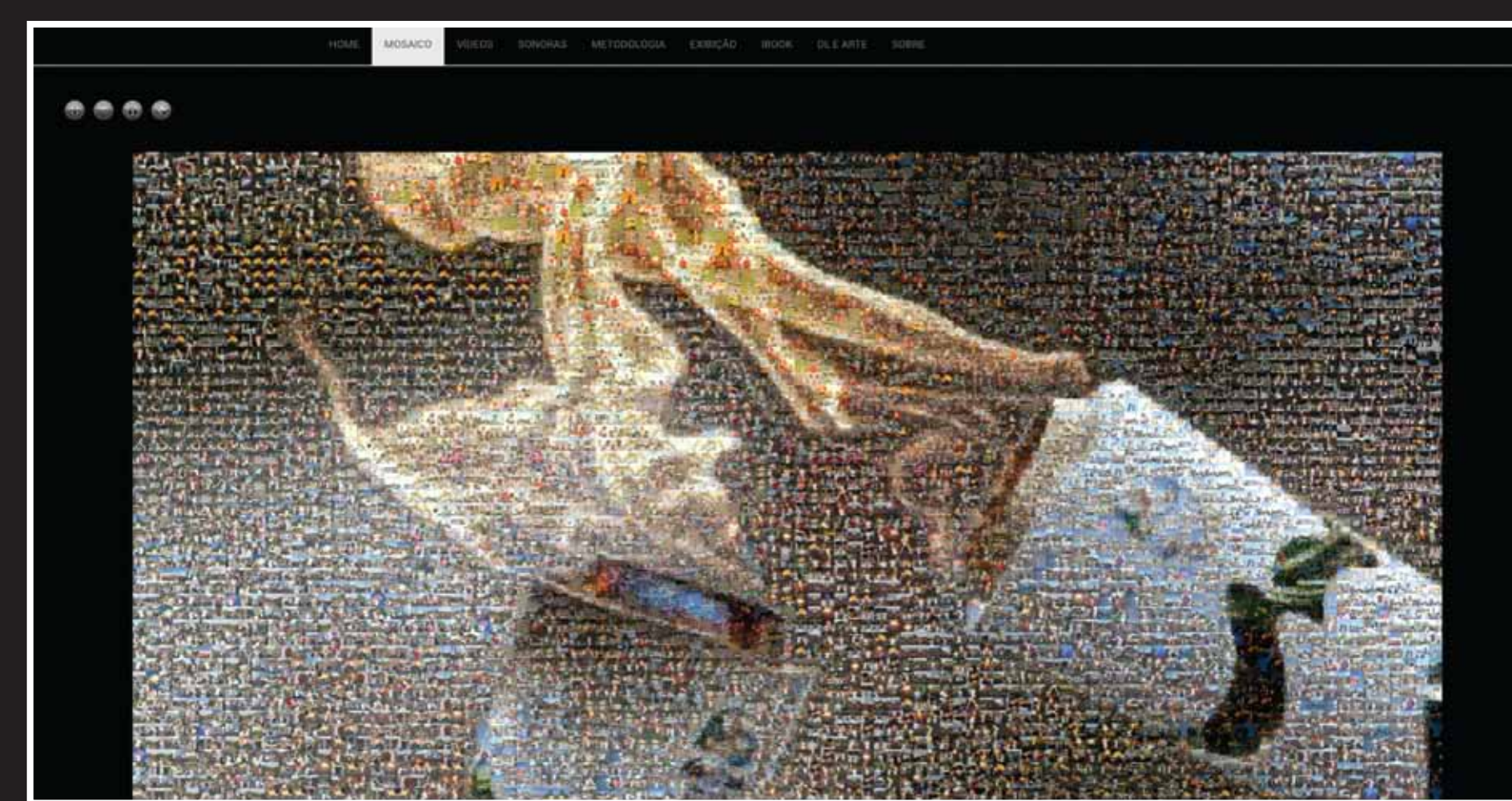
Deep Learning Technique

We referred to a Deep Learning technique to resolve an image recognition task: classify our image dataset into subject features. There are a few deep neural architectures, each one more suitable for certain problems domains. Computer vision tasks, for instance, have been recently improved by the efficient performance of Convolutional Neural Networks (CNNs) [1].

A CNN maps an input (e.g., the pixels of an image), to an output (e.g., the probabilities that an image is a member of a particular class), by automatically learning a rich hierarchy of internal image features given training data. That said, we started with a CNN pretrained on image classification (ImageNet ILSRVC2012) and, through a supervised learning approach, retrained the network on distinct image classification tasks (first the Olympic torch, then several sports disciplines). By creating a set of annotated training images, we updated the parameters of the CNN model. Finally, the retrained network ran over our dataset and returned images with their corresponding confidence score for the new categories.

Media visualization techniques

By June 25th, 1.5K images with over 83% confidence score for the Olympic torch category had been classified by our network. We used them to create a *mosaic visualization* that can be panned and zoomed (http://lvelho.impa.br/dl_rio2016/mosaico.html). The mosaic idea is that, given an image (*target images*), another image (*mosaic*) is automatically build up from several smaller images (*tile images*). To implement the mosaic we used a web-based viewer for high-resolution zoomable images called OpenSeadragon.



FROM LEFT TO RIGHT (CLOCKWISE): THE MOSAIC, THE TARGET IMAGE AND TILE IMAGES.

We also retrained the CNN to recognize some sports disciplines, such as tennis. However, we implemented a different visualization technique for the image of these categories. The *videosphere* (http://lvelho.impa.br/dl_rio2016/videos.php), which has zoom, swipe, and click-to-view enabled, is a 3D sphere containing thumbnails of dozens of videos. The videos are actually musical slideshows containing the still images [2]. Altogether, there are twelve distinct slideshows, each one representing a sport discipline automatically classified by the retrained CNN. To design the *videosphere*, we used an open code by Goo Engine built on HTML5, JavaScript, and WebGL.



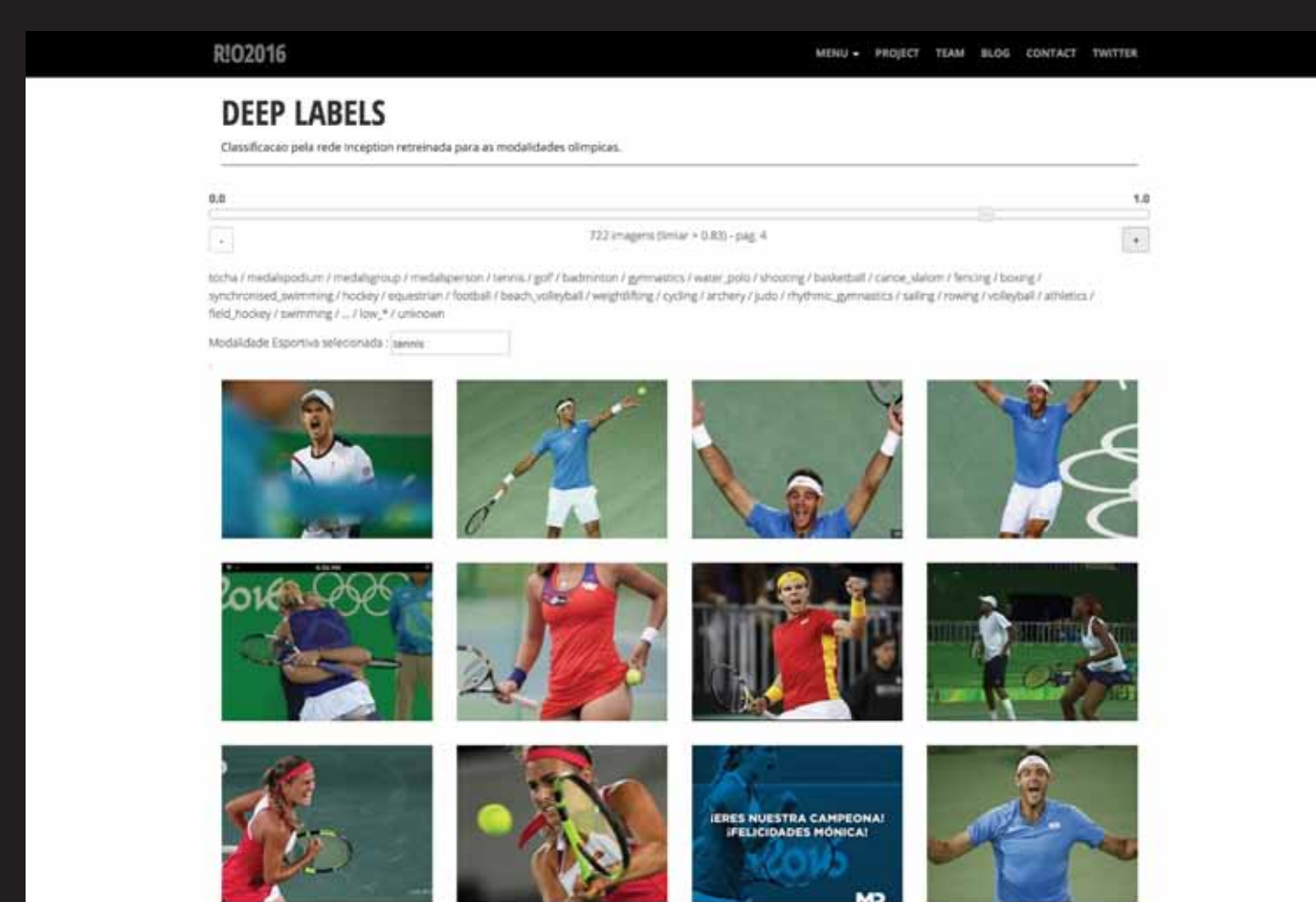
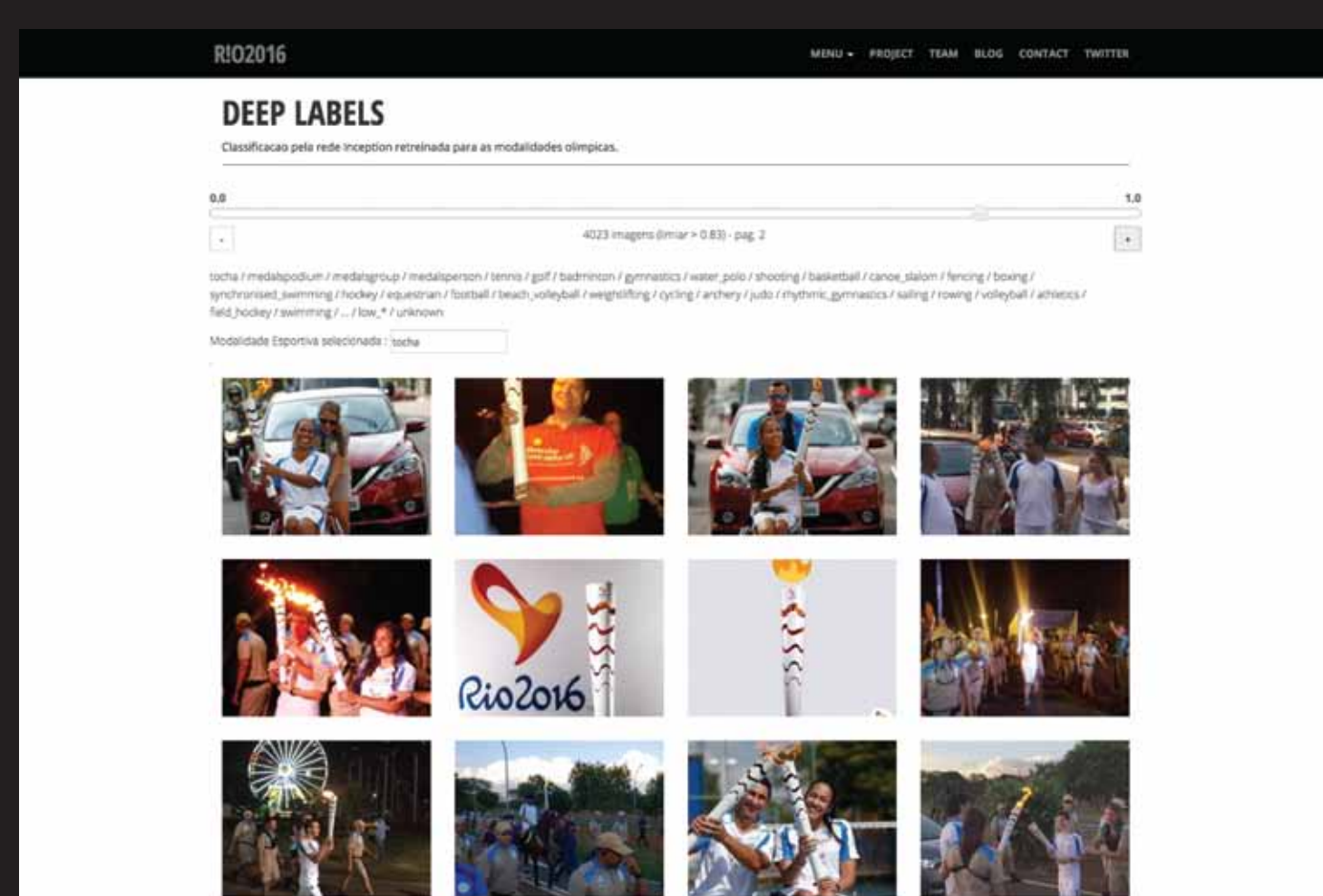
FROM LEFT TO RIGHT (CLOCKWISE): THE VIDEOSPHERE, A ZOOMED VIEW AND THE TENNIS SLIDESHOW

CONSIDERATIONS

This poster reflects an exploratory research towards a digital method to organize and visualize a large-scale imagery from Rio-2016. We expect our approach demonstrates the potentiality of CNNs allied to media visualization techniques for the area of visual discovery. For future work, though, we would like to address a more precise problem from our dataset. We could, for example, investigate Rio-2016 visual trends by measuring and visualising geo-spatio aspects of the image dataset.

REFERENCES

- [1] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks". *Advances in neural information processing systems*, pp. 1097-1105.
- [2] Giannella, J. & Velho, L. "OBSERVATÓR!O2016". Technical Report TR-08-2016, - IMPA, 2016.



A SUBSET OF CLASSIFIED IMAGES WITH CONFIDENCE SCORE OVER 83% FOR THE OLYMPIC TORCH AND TENNIS CATEGORIES, RESPECTIVELY