Deep learning to automate the assessment of cultural ecosystem services from social media data

Ana Sofia Cardoso[1]
up201804335@fc.up.pt
Ana Sofia Vaz[2]
sofia.linovaz@gmail.com
Francesco Renna[1]
frarenna@dcc.fc.up.pt

[1]Instituto de Telecomunicações, Faculdade de Ciências da Universidade do Porto

[2]Andalusian Inter-University Institute for Earth System Research (IISTA-CEAMA), University of Granada & Research Centre in Biodiversity and Genetic Resources (CIBIO-InBIO), University of Porto

Abstract

Cultural ecosystem services (CES) result from the interactions between humans and nature, contributing to people's physical and mental well-being. Most social media content analyses considered in the context of CES are based on the manual classification of photos or texts shared by social media users. Inevitably, the manual classification of big photographic data is too time consuming and costly, particularly when it comes to large study areas and audiences. In this work we studied automated image classification techniques using deep learning approaches to address CES.

1 Introduction

Nowadays, computer science and related fields have been highly invested in the use and combination of methods that incorporate social media analytics [1]. Social media platforms represent a very significant fraction of all the available digital data, constituting an efficient method to collect big data that provide information on people's interactions with each other and with their environment [2]. Fast improvements in computational power and data storage capacity during the last years have motivated the emergent fields of Digital Conservation, iEcology and conservation culturomics [3]. These disciplinary fields refer to the use of digital (big) data and technology to understand human-nature interactions and to provide evidence in favour of nature conservation and of the sustainable management of ecosystems [4]. Among these human-nature interactions are cultural ecosystem services (CES), which constitute the non-material benefits that people can experience from nature, such as recreation and ecotourism, as well as those pertaining to spiritual, religious, aesthetic or heritage values, among others [5].

An approach that combines different data from social media with advanced analytics, besides spatial analysis, remains underexplored in the context of CES assessment. Thus, the investment in methods that can identify features of ecosystems and nature through the content analysis of shared photos (or text), can constitute an asset to support the evaluation of CES, particularly, related to aesthetics and recreation or ecotourism [6]. Lee *et al*., for example, proposed a method for analysing large amounts of social media photographs, as well as to derive indicators of socio-cultural usage of landscapes, through cluster detection with Convolutional Neural Networks (CNNs) [7]. This project aims to develop an automated classification of social media photographs that can be useful for CES evaluation and for providing innovative solutions to the scientific community. Specifically, this study aims to answer the following questions: (1) can deep learning algorithms be developed to support an automated classification of social media photographs in the context of CES? and (2) how can those algorithms and models be improved so as to promote statistically reliable image classifications? To achieve this, deep learning algorithms are developed and tested, more specifically CNNs and transfer learning strategies are applied to the classification of digital photographs of the "Peneda-Gerês" protected area (Northern Portugal) obtained from the social media platforms Flickr and Wikiloc.

2 Methods

2.1 Image classification methodology

We performed a classification of the content of photographs from the protected area "Peneda-Gerês" (Northern Portugal), that were withdrawn from the Flickr and Wikiloc social media platforms, specifying a time window of 2003-2017 (1778 images in total). This classification was based on "Nature" and "Human" labels (Figure 1). To achieve that, two different CNNs architectures were implemented, the VGG16 and the ResNet152, in order to verify the most appropriate and suitable for our study.

The proposed image classification methods were evaluated over the dataset using a 5-fold-cross validation method, following the literature and taking into account the computational resources and the running time. The considered performance metrics (accuracy, sensitivity, specificity, and F1-score) were computed as the mean of the performance metrics obtained over the 5 different folds. During training, in each of the 5 folds, 10% of the training data was retained to perform model validation, in order to determine the training parameters that guaranteed the highest accuracy over the validation set.

Since we are coping with a small dataset, in order to improve the generalization of the model and avoid the overfitting, transfer learning and data augmentation schemes were considered.

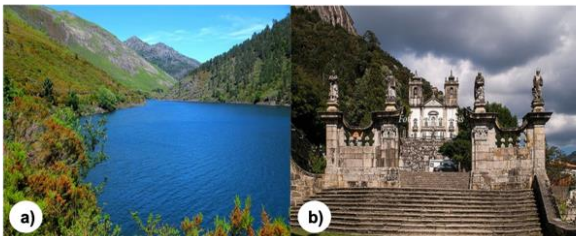

Figure 1: Examples of images belonging to the Nature and Human labels. a) Nature, b) Human.

2.2 CNN architectures and transfer learning

The VGG16 and ResNet152 were the chosen CNNs architectures. For both CNN architectures, three different sets of weights were considered: (1) weights obtained by training over the dataset "Places365", (2) weights obtained by training over the database "ImageNet" and (3) weights obtained by training the networks from scratch.

The Places365 dataset is the latest subset of the database Places, comprising around 1.8 million scene photographs of different places, labelled with 365 scene semantic categories, including photographs with similar elements to the ones under study. The ImageNet database constitutes a large-scale hierarchical image database, that has several applications in the broadest areas, comprising more than 14 million cleanly annotated images spread over around 21,000 categories. Both databases were selected due to their freely available online resources (weights and models).

Regarding the details of the transfer learning strategy implemented, all the convolutional layers were kept frozen when training over our dataset, while the remaining 3 (for VGG16) and 1 (for ResNet152) fully connected layers were trained with our dataset. Moreover, for both architectures, an additional dense layer with 128 units and a rectifier linear unit activation function was also included (to allow better fit of the model/network to the classification task) before the output layer, which was modified in order to have 2 units.

Regarding the training details, both networks were trained using the Adam optimizer. For VGG16, the best performance was verified when considering a learning rate of 0.000001 while, for ResNet152, it was 0.0001 the most accurate learning rate. Also, it was observed that, for VGG16, the model accuracy and loss had fully converged after 50 epochs, having been decided, because of that, to use only 50 epochs to build the VGG16 model, as well as the ResNet152 model, due to computing resource management.

2.3 Data augmentation

Regarding data augmentation, 5 transformations (including horizontal flip, width shift, height shift and zoom) were implemented individually for each of the images in the training set. The images in the validation set were not included in this process, in order to avoid biased results. The total number of transformations applied to each photograph (5 per image) was selected taking into account the overall running time of the algorithm, as well as the available computational memory.

3 Results

3.1 Nature vs. Human classification

When comparing the two transfer learning scenarios and the weights obtained by training only over our dataset (Figure 1), it was observed that, ImageNet had, overall, a higher accuracy for the two architectures under study (86.11 vs 87.18), followed by Places365 and weights trained only with our dataset, with the exception of Places365 in VGG16, that resulted in an equally high accuracy (87.01). Also, it was verified that, for Places365, VGG16 had a better performance when compared to ResNet152 (87.01 vs 86.00), while for the remaining scenarios, ResNet152 model was more accurate than the one for VGG16.

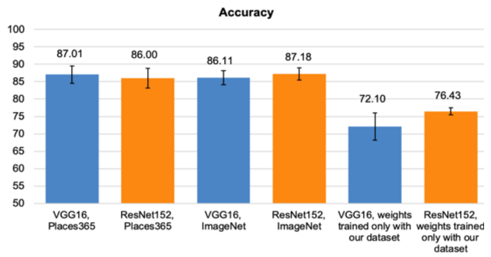

Figure 1: Accuracy of the VGG16 and ResNet152 model performance for the two transfer learning scenarios and the weights from scratch.

Considering sensitivity (Figure 2), it was verified that ImageNet had, overall, better results for the two architectures under study (86.71 and 86.78), followed by Places365 and weights trained only with our dataset, with the exception of Places365 in VGG16, that resulted in a higher sensitivity value (88.48). Likewise, it was observed that ResNet152 had slightly finer sensitivity results when compared to VGG16, except for Places365, where VGG16 showed the best result (88.48 vs 83.40).

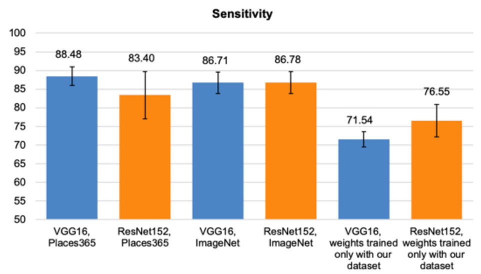

Figure 2: Sensitivity of the VGG16 and ResNet152 model performance for the two transfer learning scenarios and the weights from scratch.

For specificity (Figure 3), it was observed that Places365 had finer specificity results for the two architectures under study (85.54 and 88.46), followed by ImageNet and weights trained only with our dataset. Similarly, it was verified that ResNet152 had better specificity results when compared to VGG16, for all the scenarios under study.

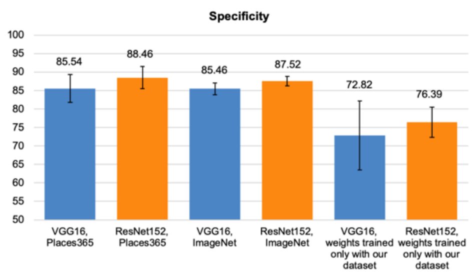

Figure 3: Specificity of the VGG16 and ResNet152 model performance for the two transfer learning scenarios and the weights from scratch.

Considering the F1-score (Figure 4), it was verified that ImageNet had slightly better F1-score results for the two architectures under study (86.53 and 87.44), followed by Places365 and weights trained only with our dataset. Also, it was observed that ResNet152 had finer F1-score results when compared to VGG16, except for Places365, where VGG16 showed the best result (87.53 vs 85.89).

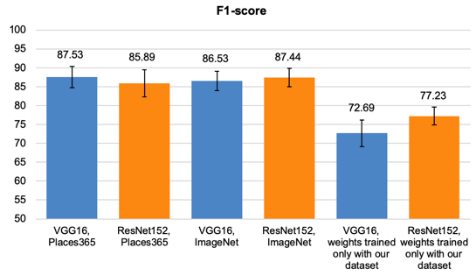

Figure 4: F1-score of the VGG16 and ResNet152 model performance for the two transfer learning scenarios and the weights from scratch.

4 Discussion and Conclusions

When comparing the two considered transfer learning scenarios and the weights obtained by training only over our dataset, it was expected that the model implemented with the Places365 weights would have a finer performance than the other two (with ImageNet weights and weights trained only with our dataset), since all the photographs contained in this dataset are exclusively related with landscapes and places in general, constituting the database that most resembles our dataset. Perhaps surprisingly, this was not the case for both VGG16 and ResNet152, as ImageNet was undoubtedly the database where the two transfer learning scenarios achieved better results. A possible explanation for this behavior can reside in the observation that deep learning models achieve more accurate results when trained in the presence of large datasets. In fact, ImageNet, by containing a larger number of photographs (more than 14 million) than Places365 (around 1.8 million), has led to a better performance of the model. Also, ImageNet contains a greater diversity of images that seems to contribute to a better generalization of the model.

The results showed that deep learning methods can offer significant contributions to assist in CES evaluation. Future work will focus on the improvement of the robustness of these models against scarcely labeled data via the use of semi-supervised approaches by leveraging autoencoder architectures and generative adversarial networks.

Acknowledgements

This work is funded by FCT/MCTES through national funds and when applicable co-funded EU funds under the project UIDB/50008/2020.

References

[1] Sherren Kate et al. Conservation culturomics should include images and a wider range of scholars. *Frontiers in Ecology and the Environment*, 2017, 15.6: 289-290. doi: 10.1002/fee.1507.

[2] Di Minin et al. Prospects and challenges for social media data in conservation science. *Frontiers in Environmental Science*, 2015, 3: 63. doi: 10.3389/fenvs.2015.00063.

[3] Jarić Ivan et al. iEcology: Harnessing Large Online Resources to Generate Ecological Insights. *Trends in Ecology & Evolution*, 2020. doi: 10.1016/j.tree.2020.03.003.

[4] Toivonen Tuuli et al. Social media data for conservation science: a methodological overview. *Biological Conservation*, 2019, 233: 298-315. doi: 10.1016/j.biocon.2019.01.023.

[5] Assessment Millennium Ecosystem et al. *Ecosystems and human well-being* (Vol. 5). United States of America: Island press, 2005. Doi:

[6] Richards Daniel R.; Tunçer, Bige. Using image recognition to automate assessment of cultural ecosystem services from social media photographs. *Ecosystem services*, 2018, 31: 318- 325. doi: 10.1016/j.ecoser.2017.09.004.

[7] Lee Heera et al. Mapping cultural ecosystem services 2.0–Potential and shortcomings from unlabeled crowd sourced images. *Ecological Indicators*, 2019, 96: 505-515. doi: 10.1016/j.ecolind.2018.08.035.