Monitoring Visitor Numbers with Computer Vision

***Jeroen Staab**, German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Germany.*
***Hannes Taubenböck,** DFD, German Aerospace Center (DLR), Germany.*
***Hubert Job**, Institute of Geography and Geology, Julius-Maximilians-Universität Würzburg, Germany.*

Every day, a large diversity of visitors is encountered in protected areas - dog walking locals, hiking tourists or trekking bikers. Knowing their quantities and economic impact provides valuable arguments in favor of designating parks and, thus, helps to conserve our planet's biosphere. Furthermore, nature-based tourism and outdoor recreation help to develop regions of the rural periphery. To quantify visitation, a long list of methods and instruments is available to our domain (see CHESSFORD & MUHAR, 2003). Nevertheless, each tool comes with specific pros and cons, e.g. false triggers due to wildlife or lack of detailed information about the visitors. At the same time though, each visitor group has specific economic characteristics and demands, as well as corresponding ecological impacts.
Utilizing cameras to count visitors has proven to be accurate, traceable and rich in features (ARNBERGER et al., 2005). However, extracting data from the imagery manually consumes large resources, limiting the utilization of camera observations to short-term monitoring projects. In this work, we apply and test computer vision to characterize visitors at the *Biosphere Reserve Schorfheide-Chorin* in Germany in an automatic manner.

Convolutional Neural Networks

State of the art algorithms, such as *convolutional neural networks* (CNN), are not only well-known for speech recognition and mastering the game *Go*, but also for solving image classification problems. Therefore these machine-learning algorithms are utilized in autonomous cars, earth observation and other fields to detect objects (ZHANG et al., 2017). They empathize visual perception using layers of locally-sensitive, orientation-selective and connected neurons. At first, these neurons are randomly initialized. Then, the software architecture is trained on specialized *graphical processing units* (GPU) using a large number of samples. Hereby the weights organize autonomously. The resulting rule-set then can be exported and deployed by others. It is interesting to stress, that while training costs large computational resources, predicting with the model is less intense. As many CNNs are protected by intellectual property rights and developing a new one was outside our scope, we utilized a system available under public domain. *You Only Look Once* (YOLO) has been developed by REDMON et al. (2016). Their CNN is very efficient and handles variable image sizes. Most important though, it generalizes well into new domains, as it was trained with millions of images in hundreds of categories queried from the Internet (REDMON et al., 2016). Among these categories are, for example, *backpacks*, *bicycles* and *dogs*, which help to characterize visitors segments (see figure a+b).

Field Experiment

The experimental set-up was installed at the *Chorin Abbey* within the *Biosphere Reserve Schorfheide-Chorin*, located 60 km northeast of *Berlin*. To identify visitor characteristics and their economic impact, a combination of field observations and the collection of imagery were conducted in the field. Therefore, a low-cost setup consisting of a *RaspberryPi 3*, a corresponding *PiCam* and a Powerbank (20.000mAh) was utilized. The device was installed

diagonally five meters next to the census line at 1 meter height. Finally, the system was configured to routinely capture an image once every minute (time-laps).
By default, the CNN retrieved fair results. During eight hours of observation, 365 persons were determined by the visitor counter. Compared to a manual reference sample, hereby only 4.4% of all visitors were missed at 0.3 false positives per window. While the miss-rate is acceptable, counting one additional person every three frames erroneously is not accurate enough. A qualitative assessment was conducted to point out possible mistakes. The results show that (i) it is difficult to isolate individuals in crowded scenes (see figure c). (ii) Small people in the background were detected (see figure c), although they were not included in the reference data, as we did not expect YOLO to be that sensitive. Last but not least (iii), bright tree trunks with a significant contrast to their environment sometimes were classified incorrectly as *person* in otherwise blank images. The most interesting about the pre-trained CNN, however, was its variety of detected objects. 17 *backpacks* were detected (82.4% accuracy). These, together with solid shoes, for example are an indicator for hikers. Also 8 *bicycles* were found in the image archive, whereof 75% were classified correctly. It has to be mentioned though, that one false positive was actually a wheelchair. Unfortunately, no dogs were observed, but nevertheless, this class is available in YOLO without the costs of conducting any parameter optimizations. The following figure illustrates the CNN's capabilities. Even in crowded scenes, covered and cut off objects are detected well.

Figure: Bounding boxes around classified objects detected by YOLO

Discussion

The approach of applying computer vision proves capable for quantifying people, or as in our application, visitors of protected areas. However, it is mandatory to respect the legal framework when using cameras. In Germany for example, identifying individuals poses a threat to their personality rights. In our field experiment, visitors were informed by a sign about ongoing camera observations. Along that, legal issues can be circumvented by processing the images instantly, without backing them up. Otherwise the most crucial parameters are the camera lens, image resolution and distance to census line.
We found CNNs very promising for this domain, as they are capable of extracting specialized visitor characteristics. We also tested and compared two other technologies (*Change Detection*, *Histograms of Orientated Gradients*); however, CNNs resulted in the highest information gain (Staab, 2017). The introduced CNN by Redmon et al. (2016) provides user-friendly access to this latest computer vision technology. At the same time, we want to stress its possibilities to be retrained - e.g. to add a specific object class. Thus, opposed to commercial visitor counters, herewith an open-source project is presented, which may be

developed further according to its operational purpose. Consequently, it is also possible to apply the methods onto imagery captured with motion triggered cameras (MILLER, 2017). Last but not least, it may be interesting to measure the distance between objects (i.e. dog and person as leash indicator). The methods were wrapped into an *R* package named *wuepix*, which ensures a consistent interface. It is public accessible online and can be downloaded via https://github.com/georoen/wuepix.

References

Arnberger, A., Haider, W., Brandenburg, C., 2005. Evaluating Visitor-Monitoring Techniques: A Comparison of Counting and Video Observation Data. Environmental Management 36, 317–327.

Cessford, G., Muhar, A., 2003. Monitoring Options for Visitor Numbers in National Parks and Natural Areas. Journal for Nature Conservation 11, 240–250.

Miller, A.B., Leung, Y.-F., Kays, R., 2017. Coupling Visitor and Wildlife Monitoring in Protected Areas Using Camera Traps. Journal of Outdoor Recreation and Tourism 17, 44–53.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, Real-time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 779–788.

Staab, J., 2017. Applying Computer Vision for Monitoring Visitor Numbers - A Geographical Approach. Institute of Geography, Julius-Maximilians-Universität Würzburg, Germany, Master’s thesis.

Zhang, S., Benenson, R., Omran, M., Hosang, J., Schiele, B., 2017. Towards Reaching Human Performance in Pedestrian Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 973–986.