98 Triggered trail camera images and machine learning based computer vision as alternative to established visitor monitoring approaches?

Marius Mayer[1], **Jeroen Staab[2,3]**, **Erica Udas[4,5]**, **Hannes Taubenböck[2,6]**, [1]University of Innsbruck, Austria. [2]German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), Germany. [3]Humboldt-University Berlin, Geography Department, Germany. [4]International Center for Integrated Mountain Development (ICIMOD), Nepal. [5]University of Greifswald, Institute of Botany and Landscape Ecology, Germany. [6]University of Würzburg, Germany

Visitor monitoring is crucial for many management and valuation tasks in protected areas and other recreational landscapes. Its core data are visitor numbers which are costly to estimate in absence of entry fees. Camera-based approaches have the potential to be both, accurate and deliver comprehensive data about visitor numbers, types and activities. So far, camera-based visitor monitoring is, however, costly due to time consuming manual image evaluation (Miller et al. 2017). To overcome this limitation, we deployed a convolutional neural network (CNN) and compared its hourly counts against existing visitor counting methods such as manual in-situ counting, a pressure sensor, and manual camera image evaluations.

The study site is Eldena Forest Nature Reserve (EFNR), which is located at the southeastern rim of the city of Greifswald, Germany. The forest is owned by Greifswald University and is spread over 411 ha. Given its close vicinity to the city of Greifswald, EFNR is frequently visited; however, exact visitor numbers have never been estimated so far (Udas et al. 2018). There are seven major entry points into EFNR, and based on the communication with local foresters, the entries at A, B, C and D are the ones mostly used due to their proximity to residential areas.

The methodology is explained in detail in Staab et al. 2021. While conducting systematic visitor monitoring in EFNR, three different visitor counting methods were deployed at different entrances in 2015. At all seven entrances, manual in-situ visitor counting was carried out as a fundamental benchmark following a visitor counting method used in many protected areas (Mayer et al. 2010). To determine the annual cumulative number of visits in EFNR, the actual data from the sampled days were extrapolated. The extrapolation procedure accounted for seasonality, weekends/weekdays and the weather situation and is a standard procedure used by many studies (Mayer et al. 2009). In addition, at one of the most frequented entrances, entrance A, a pressure sensor was installed to capture the seasonal variation of visitation, and triggered trail cameras were installed at entrances B, C, D and E. The images were first evaluated manually using three semantics of image interpretation to estimate visitor numbers. The number of visitors entering the EFNR was documented on an hourly basis for each day in a spreadsheet. Regarding the automated image analyses, advanced computer vision technologies such as deep CNNs can detect pedestrians at very high accuracies. We used a pre-trained image analyzing framework developed by Redmon et al. (2016), respectively Redmon & Farhadi (2017, 2018) and, as its name indicates, You Only Look Once (YOLO) is very fast in grasping an image's content. As a result of the versatile training data, the pre-trained algorithm detects several object classes in an image, among which are persons, bicycles, backpacks and dogs – categories of special interest to characterize visitors in recreational landscapes. We directly compared the results of all counting approaches. For each entrance the raw, hourly results per counting approach were set against each other. The statistical deviations were measured using Pearson's correlation and a linear model without intercept.

At entrance A, where manual in-situ observations were conducted next to the pressure sensor on five days, we found a strong and highly significant correlation (0.783, $p < 0.001$). The respective linear model fitted through the 50 hours of simultaneous observations further revealed that the automated approach does account for 88.4% of the visits counted by the manual in-situ observer (adj. $R^2 = 0.799$, $p < 0.001$).

Regarding the other two entrances where manual in-situ counting was conducted along with ongoing camera observations, only the manual and YOLO camera evaluations at entrance B correlate significantly with the corresponding 44 hours of manual in-situ observations due to too low sample size at entrance C. However, the regression models comparing the two camera evaluation approaches against manual in-situ observations are significant at entrances B and C.

When comparing the automated and the manual image evaluations of entrances B-D against each other, both approaches strongly correlate at very high significance levels (Ør = 0.818, p<0.001).

Further, the results of the automated image evalutation also often correlate significantly but with low to medium strength to the in-situ personal counts and the pressure sensor at the other entrances. This shows that YOLO is able to reflect the visitation trends over the year even though observations did not always take place exactly at the same locations.

Thus, the results show that the CNN derived comparable visitor numbers to the other visitor counting approaches regarding visitation patterns and numbers of visits. Therefore, we conclude that it is a fast and reliable method that could be used in protected areas as well as in a much wider array of visitor counting settings in other recreational landscapes. The approach also allows for counting dogs and recreational equipment such as backpacks and bicycles in automatic manner. While the accuracy of these categories has not been assessed yet, we suppose this new monitoring approach shall help managing particular user groups and helps avoiding conflicts in such recreational areas. Nevertheless, camera installation takes time and effort, regular maintenance (batteries, storage cards) and the automated evaluation requires specific hardware and expertise as well. Along ethical and legal concerns, other practical issues are theft, vandalism and the short lifespan of batteries, especially in the winter season.

References

Mayer et al. 2009. Wirtschaftliche Bedeutung des naturnahen Tourismus im Simmental und Diemtigtal – Regionalökonomische Effekte und Erfolgsfaktoren. Birmensdorf/Rapperswil: WSL/HSR. Mayer et al. 2010. **https://doi:10.1016/j.landurbplan.2010.04.013**. Miller et al. 2017. **https://doi:10.1016/j.jort.2016.09.007**. Redmon et al. 2016. You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016, 779–788. Redmon J & Farhadi A. 2017. YOLO9000: Better, Faster, Stronger. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017, 7263-7271. Redmon J & Farhadi A. 2018. YOLOv3: An Incremental Improvement. arXiv:1804.02767. Staab et al. 2021. Comparing established visitor monitoring approaches with triggered trail camera images and machine learning based computer vision. Journal of Outdoor Recreation and Tourism, in review. Udas et al. 2018. **https://doi:10.1108/IJSHE-05-2016-0089**.