Combing Machine Learning Image Segmentation and Mixed Reality Design to improve Users´ Navigation Experience

Christian E. Murphy [a]*, Guillermo Fernando Esquivel Tabares [a]

[a] Technische Universität München, christian.murphy@tum.de, guillermo.esquivel@tum.de

Keywords: Mixed Reality, Machine Learning, Image Segmentation, Visual Highlighting

Abstract:

This work combines machine learning image segmentation and the design of mixed reality to create an improved navigation experience in mixed reality (MR). For this user´s experience, the visual appearance of reality is to manipulated in a way that crucial objects for the navigation task are to be highlighted (i.e. road signs and traffic lights), whereas non-significant context objects in view (i.e. building facades) are to be visually downgraded. This way, it is aimed to reduce the users' effort to focus on objects that need to be perceived. Two user tests with 50 participants reveal that the realistically highlighted objects were more striking and evident to the users.

Machine Learning make decisions and solve problems without human intervention (Holzinger et al., 2019). Furthermore, machine learning services are able to provide real-time analytics for big data (Audsley et al., 2016). This makes artificial intelligence a good solution for real-time image segmentation in a MR navigation application, as computer vision and machine learning can automatically analyse and annotate videos (Khan and Al-Habsi, 2020).

YOLACT is a fully-convolutional model for real-time instance segmentation (Bolya et al., 2019), The algorithm has been trained by 2.5 million labelled objects of about 328,000 images. The training is sourced by feeding YOLACT with the Microsoft Common Objects in Context database (Lin et al., 2014) featuring 91 different object classes. YOLACT was chosen for the real-time instance image segmentation in this work. Figure 1 shows such a result.
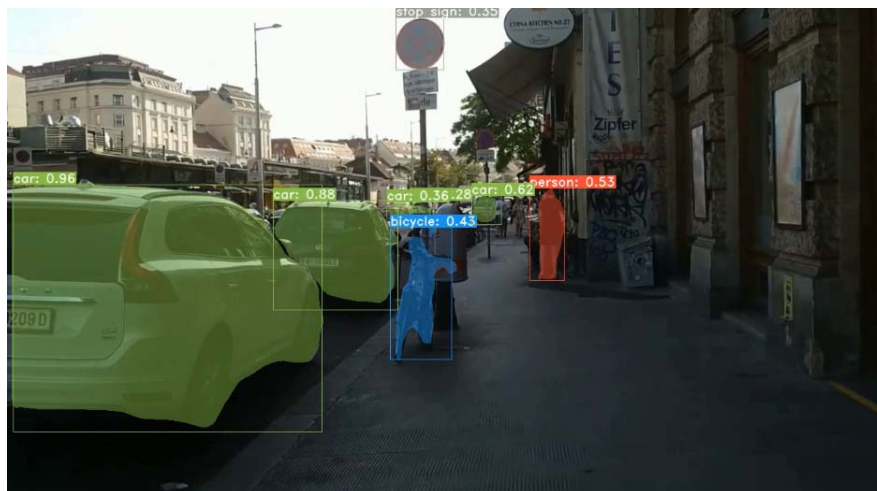

Figure 1. Object identification and segmentation results

Here, cars are correctly identified, segmented and marked for demonstration purposes. The cars are assigned a green transparent overlay, the bicycle´s is blue, a person´s is an orange overlay, and the traffic sign is in grey. The segmentation is in most cases geometrically exact to a degree where it segments individual objects visually precise.

MR design has been generally constrained to the cartographic map symbols. However, once discrete objects in the imagery are segmented, local, object-specific manipulations can be applied to the imagery. For this study, crucial objects for the navigation task are highlighted with increased brightness, contrast and saturation (see Figure 2), whereas non-significant context objects are visually downgraded by reducing contrast and blurring.

Figure 2. Visual object highlighting for a traffic sign

Two comparing user tests with 50 participants were conducted in order to reveal which highlighted objects were more striking and evident to the users. While the first user test was conducted based on a non-manipulated navigation video sequence, the second was a manipulated, designed navigation video sequence. In other words, the second user test featured highlighted objects as well as visually downgraded objects. After both tests the users had to identify the most striking object classes.

The user tests reveal that the realistically highlighted objects were more striking and evident to the users in comparison to the same object classes in the non-manipulated video (i.e. traffic signs). However, a significantly opposite effect did not occur when the characteristics of less relevant objects for the navigation task were de-emphasized (i.e. parking bicycles).

Further work has to be done in order to scientifically evaluate the intended improved user´s navigation experience. In particular, this work demonstrates the power and benefits of combining machine learning image segmentation and MR. The progress of this research project will lead one day to the question of how much of the user´s visual field should be manipulated, or in other words, how much unspoilt reality should be maintained for human navigation tasks.

References

AUDSLEY, N., BASANTA-VAL, P., WELLINGS, A., GRAY, I. & FERNANDEZ, N. 2016. Architecting Time-Critical Big-Data Systems. *IEEE Transactions on Big Data*.

BOLYA, D., ZHOU, C., XIAO, F. & LEE, Y. J. 2019. YOLACT: Real-Time Instance Segmentation. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9157-9166.

HOLZINGER, A., PLASS, M., KICKMEIER-RUST, M., HOLZINGER, K., CRIŞAN, G. C., PINTEA, C.-M. & PALADE, V. 2019. Interactive machine learning: experimental evidence for the human in the algorithmic loop. *Applied Intelligence,* 49**,** 2401-2414.

KHAN, A. I. & AL-HABSI, S. 2020. Machine Learning in Computer Vision. *Procedia Computer Science,* 167**,** 1444-1451.

LIN, T.-Y., MAIRE, M., BELONGIE, S., HAYS, J., PERONA, P., RAMANAN, D., DOLLÁR, P. & ZITNICK, C. L. Microsoft COCO: Common Objects in Context. *In:* FLEET, D., PAJDLA, T., SCHIELE, B. & TUYTELAARS, T., eds. Computer Vision – ECCV 2014, 2014// 2014 Cham. Springer International Publishing, 740-755.