Cartography M.Sc.

Machine Learning Image Segmentation to Improve Object Recognition in Mixed Reality

Guillermo Fernando Esquivel Tabares

Outline

- Introduction
  - Research Objectives
  - Research questions
- Conceptual framework
  - Machine Learning
  - Image segmentation
  - Mixed Reality
  - Image enhancement
- Methodology
  - Object recognition model calibration
  - Image enhancement application
  - Implementation
- User tests
  - User test I: Identification of striking objects
  - User test II: Identification of striking objects applying image enhancement
- Results
  - User test findings
  - Discussion
  - Development of research questions
  - Limitations
- Conclusions
- References

INTRODUCTION

Introduction

- Models to assist navigation
- Mixed Reality (MR) as a mechanism to provide additional information of the environment
- This thesis aims at facilitating the navigation experience in MR by recognizing more efficiently those objects that provide relevant information to users to navigate
  - Detecting object instances
  - Modifying their depiction properties

Research objectives:
General

- Facilitate the recognition of objects in a Mixed Reality context
- Through image enhancement techniques to highlight and downgrade objects
- Accentuate elements that assist navigation

Research objectives:
Specific

1) Calibrate an object recognition model
2) Implement image enhancement techniques to modify depiction properties
3) Integrate the object recognition model and the image enhancement techniques

Research questions

- **What** is the difference between instant segmentation and object detection?
- **Which** existing model fits adequately to generate the desired segmentation process?
- **How** to integrate an object recognition model with image enhancement techniques to highlight and downgrade elements in a video?
- **Can** a preprocessed image segmentation tool together with image enhancement methods facilitate object recognition in Mixed Reality?

CONCEPTUAL FRAMEWORK

Mixed Reality

MR will be treated as a set of technologies that allows combining elements from both the actual and digital worlds (Costanza et al., 2009).

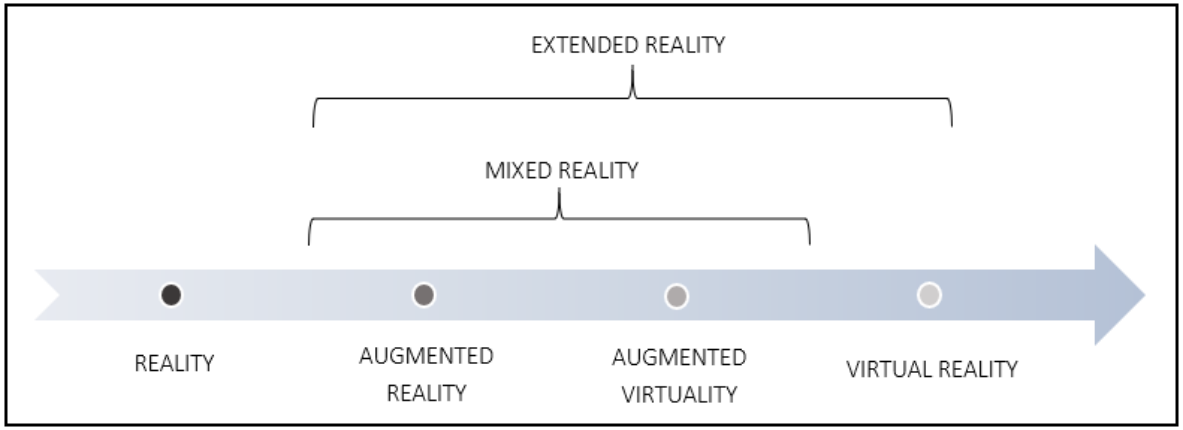
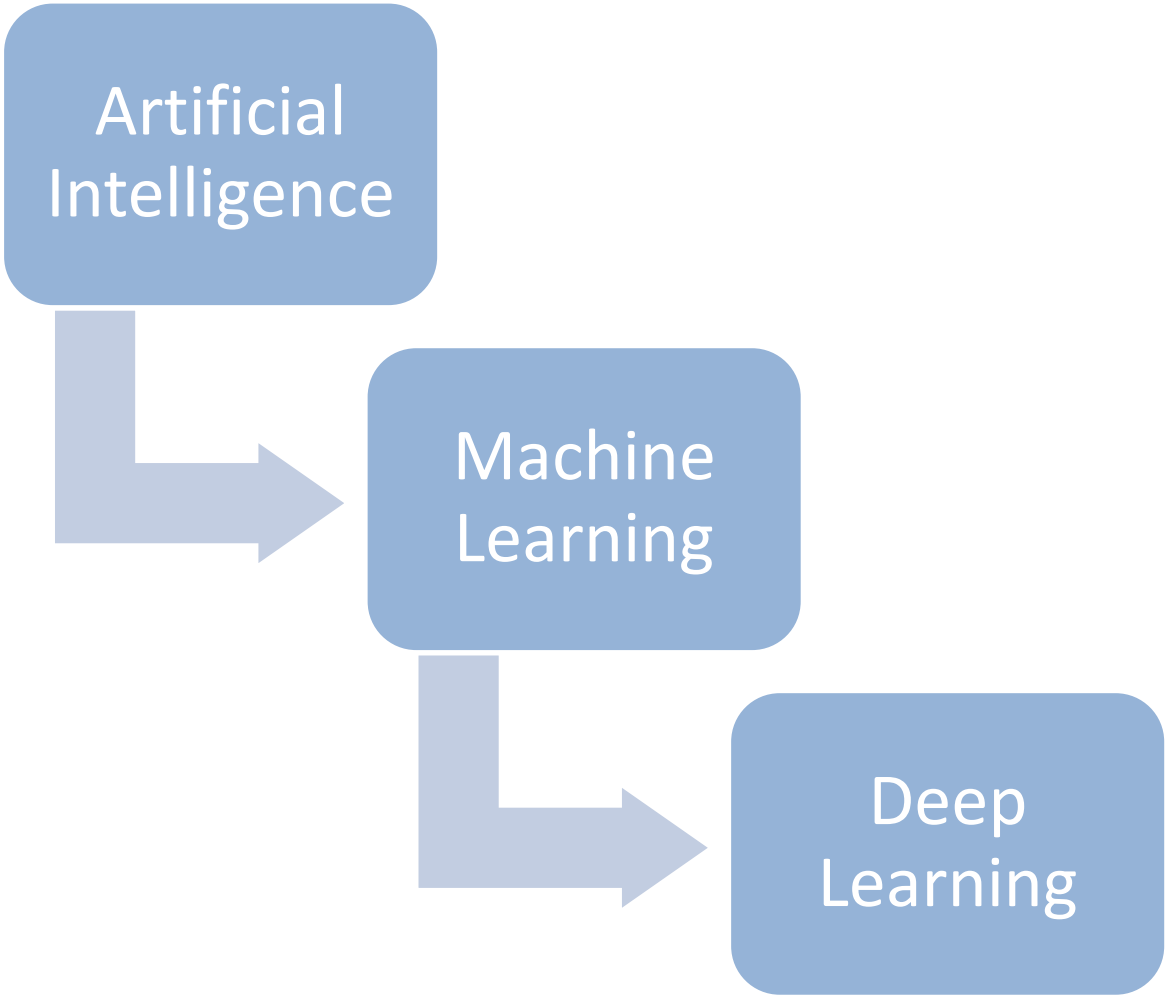


Figure 1. Dimensions covered by Mixed Reality and Extended Reality.
Source: Author design based on Çöltekin et al. (2020), Fast-Berglund et al. (2018) and Mann et al. (2018).

Machine Learning



Machine Learning: Convolutional Neural Networks

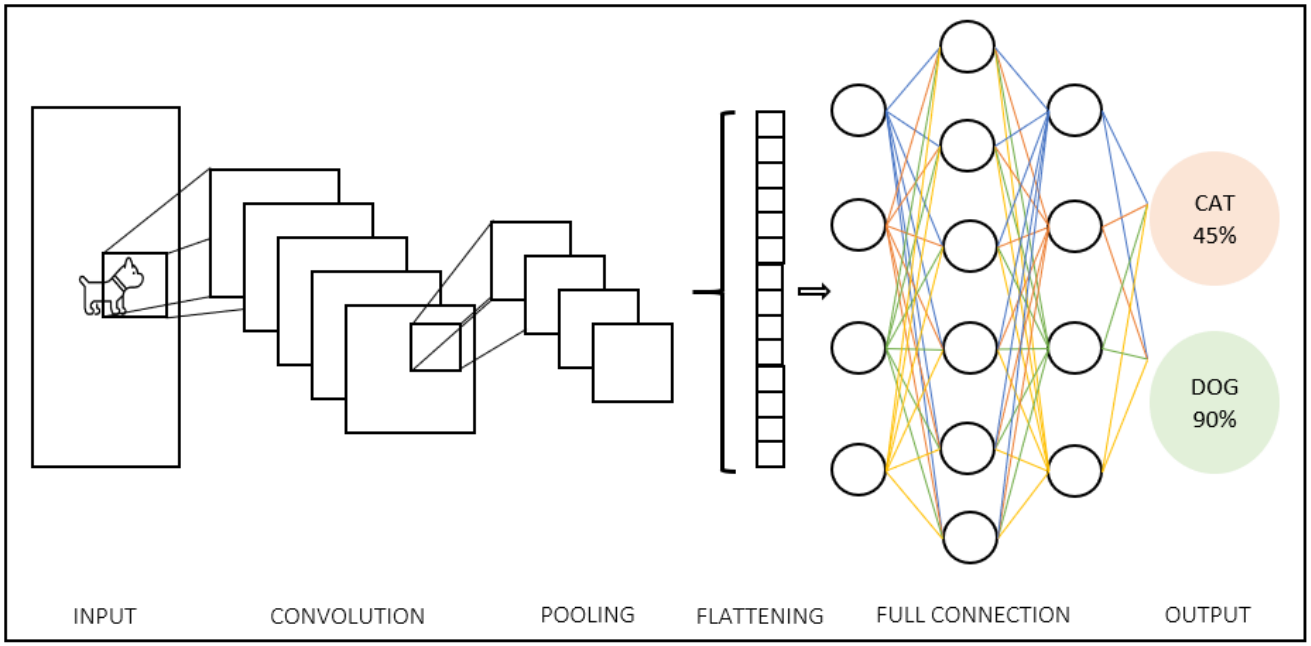


Figure 2. Structure of a simple Convolutional Neural Network. Source: Author design based on O'Shea and Nash (2015) and Zhu et al. (2020).

Convolution

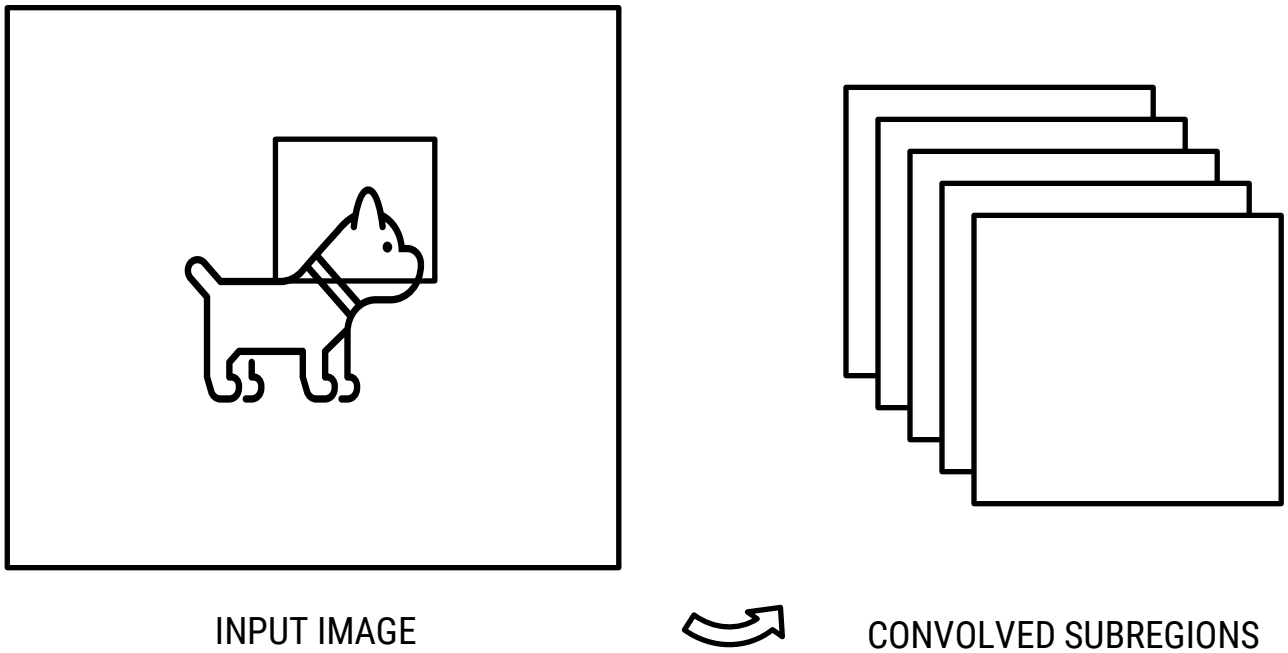
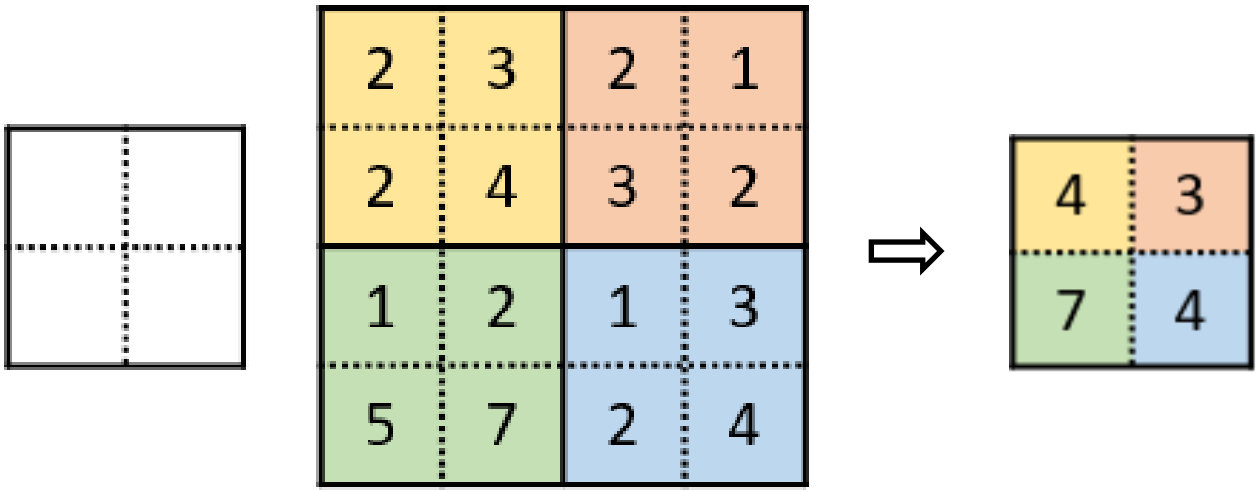


Figure 3. Convolution through every region of the image.

Pooling



2X2 MASK

PARTITIONED SUBLAYER

POOLING LAYER

Figure 4. Pooling execution using a 2x2 mask.

Flattening

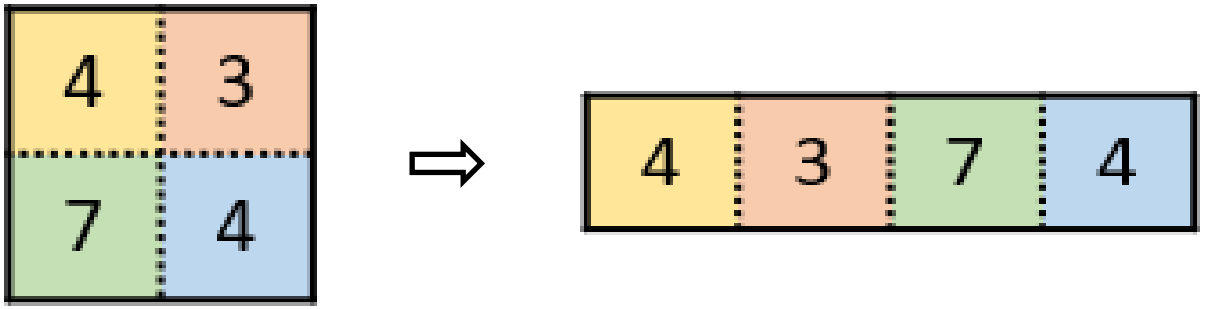


Figure 5. Flattening process.

Full connection

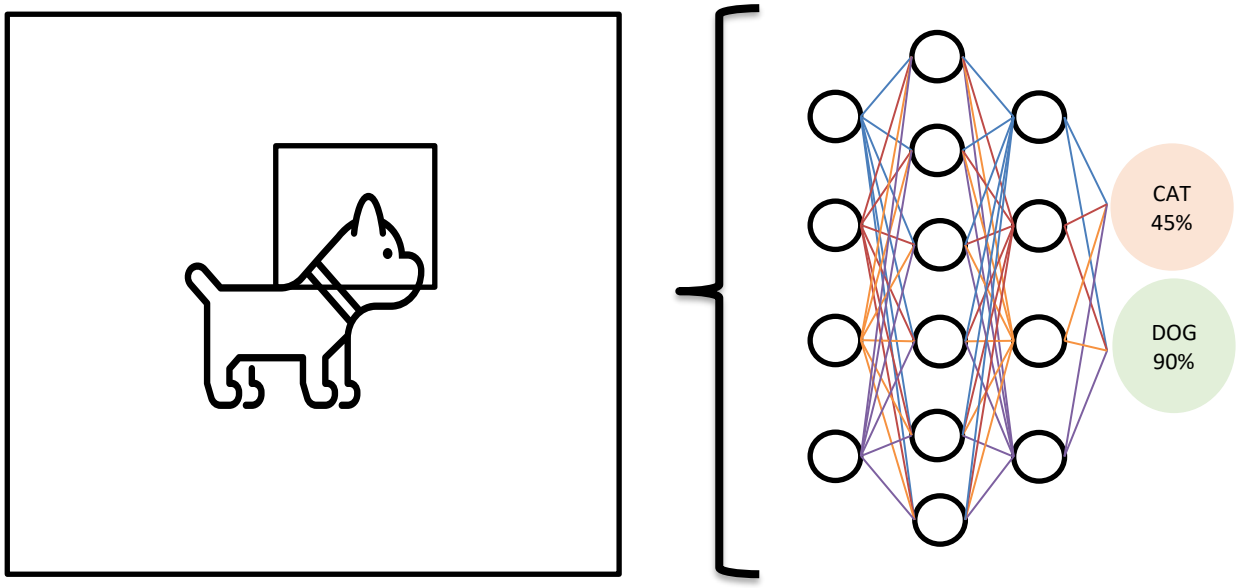


Figure 6. Object identification through full connection.

Image segmentation

Image segmentation is the process of yielding clustered subregions to identify sections of interest in imagery based on common pixel properties (Zhu et al., 2020).

20	21	20	21	17	11	10	10
21	20	21	20	16	11	12	11
20	21	22	21	17	10	11	10
21	15	16	15	16	9	10	12
20	16	15	14	15	6	4	5
1	1	1	2	3	5	4	5
2	3	1	1	2	5	6	6
3	2	2	3	1	1	2	5

Figure 7. Example of image segmentation according to difference value.

Object recognition

Object recognition is defined as a set of subtasks that provide a semantic understanding of digital images and footages aiming to ascertain the identity of elements based on known predefined labels (Yang, 2009; Z.-Q. Zhao et al., 2019).

Two main approaches of Object Recognition:

1) Object detection
2) Instance segmentation

Object detection

Figure 8. Example of object detection. Source: Redmon et al. (2016).

Instance segmentation

Figure 9. Example of instance segmentation. Source: Bolya et al. (2019).

Image enhancement

Image enhancement oversees the process of making certain features of interest in an image more obvious to the observer (Kaur, 2013).

Ways to accentuate objects in images:

- Through highlighting the object itself
- Through downgrading other objects that might be striking

Highlighting techniques

Highlighting can be understood as the alteration of image characteristics to visually emphasize certain regions (Murphy, 2015).

Operations to perform image highlighting:

- Contrast enhancement
- Brightness enhancement
- Saturation enhancement

Downgrading techniques

The downgrading methods are applied over the object to be debased, so their attributes are harder to identify and perceived as part of the background (Murphy, 2015).

Operations to perform object downgrading:

- Blurriness
- Covering layers (haze effect)

METHODOLOGY

Methodoly outline

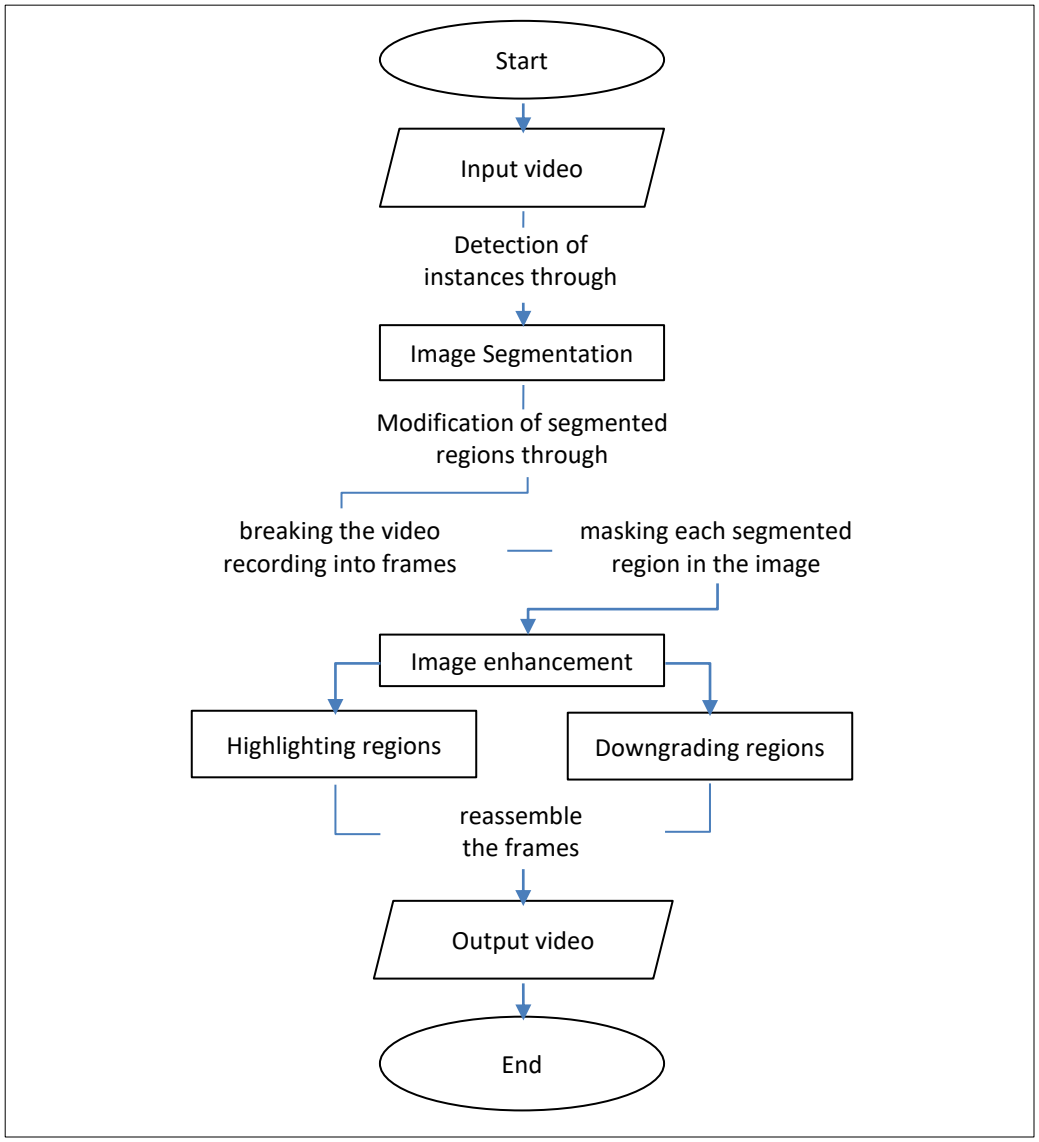


Figure 10. Methodology workflow chart.

Object recognition model

Selection criteria

- Identify and label objects
- Obtain object instances
- Process video files
- Yield high resolution video

Object recognition model

YOLO (You Only Look Once)

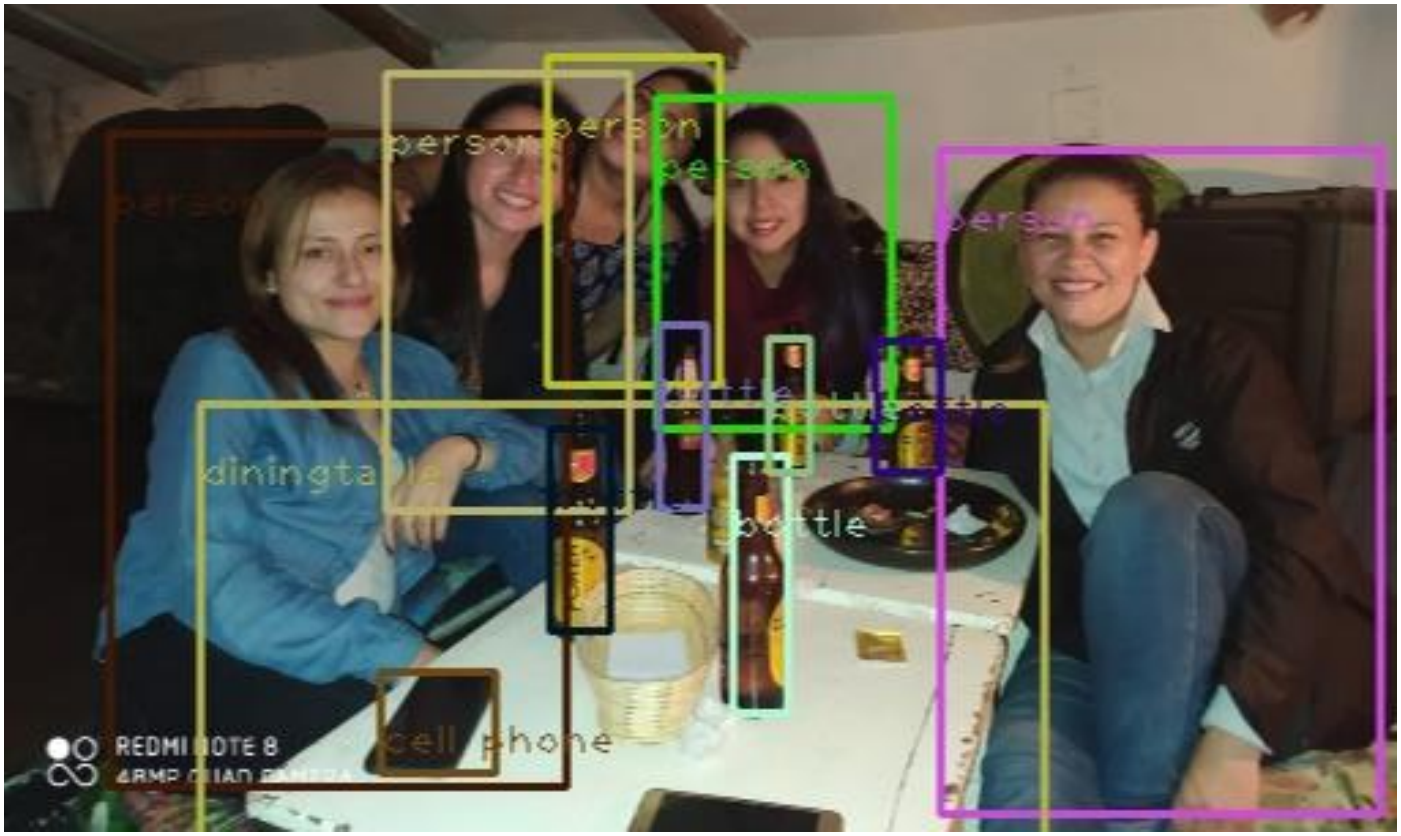


Fig. 11. Example of object detection implementing YOLO (You Only Look Once).

Object recognition model

YOLACT (You Only Look At Coeficient)

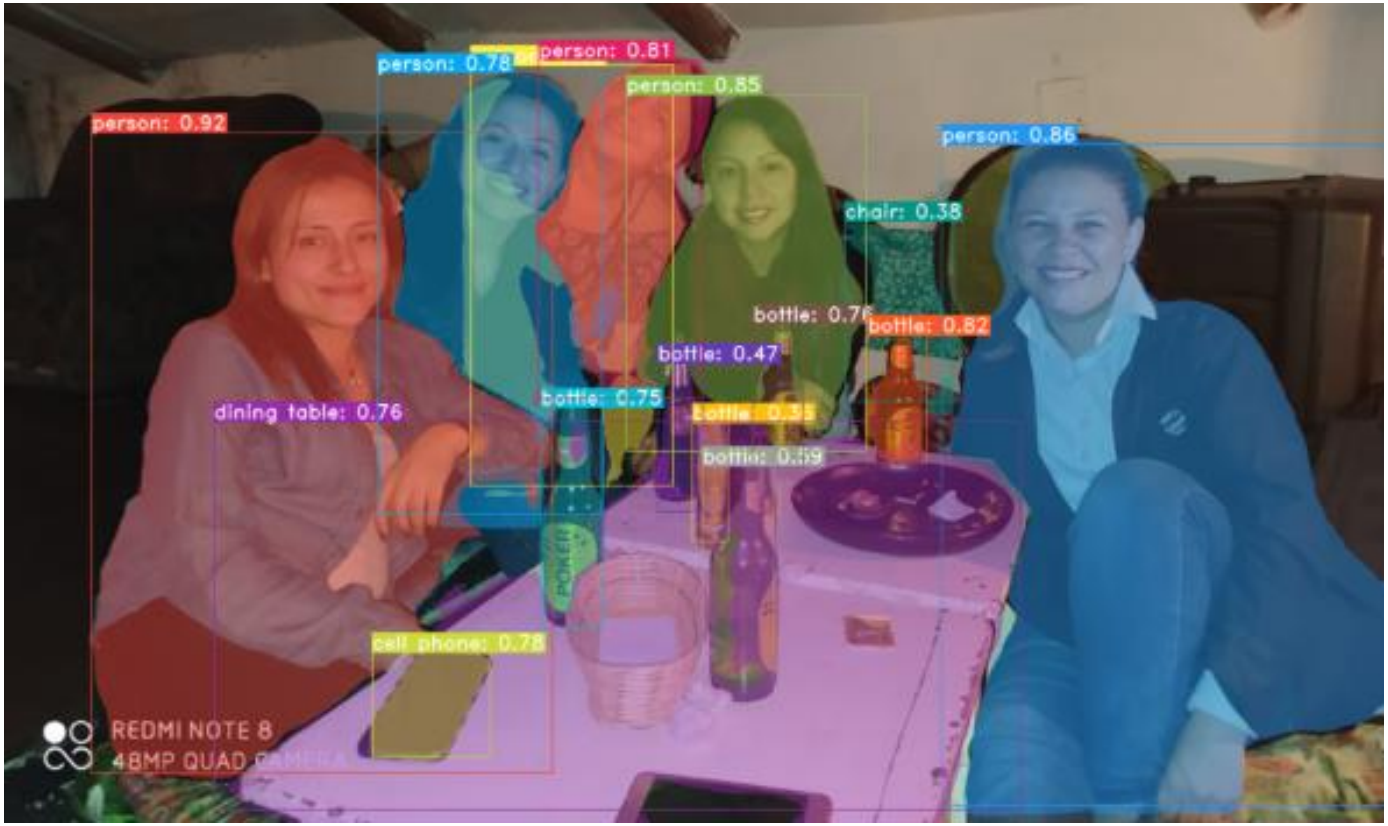


Fig. 12. Example of instance segmentation employing YOLACT (You Only Look at Coefficient).

Object recognition model
Selection

Object detection YOLO	Instance segmentation YOLACT
Demarcates the spatial extension	Yields the spatial extension and object mask
Overlooks object features	Focuses on object shape
Low resolution output	High resolution output

Table 1. Comparison between YOLO and YOLACT.

Object recognition model

YOLACT implementation

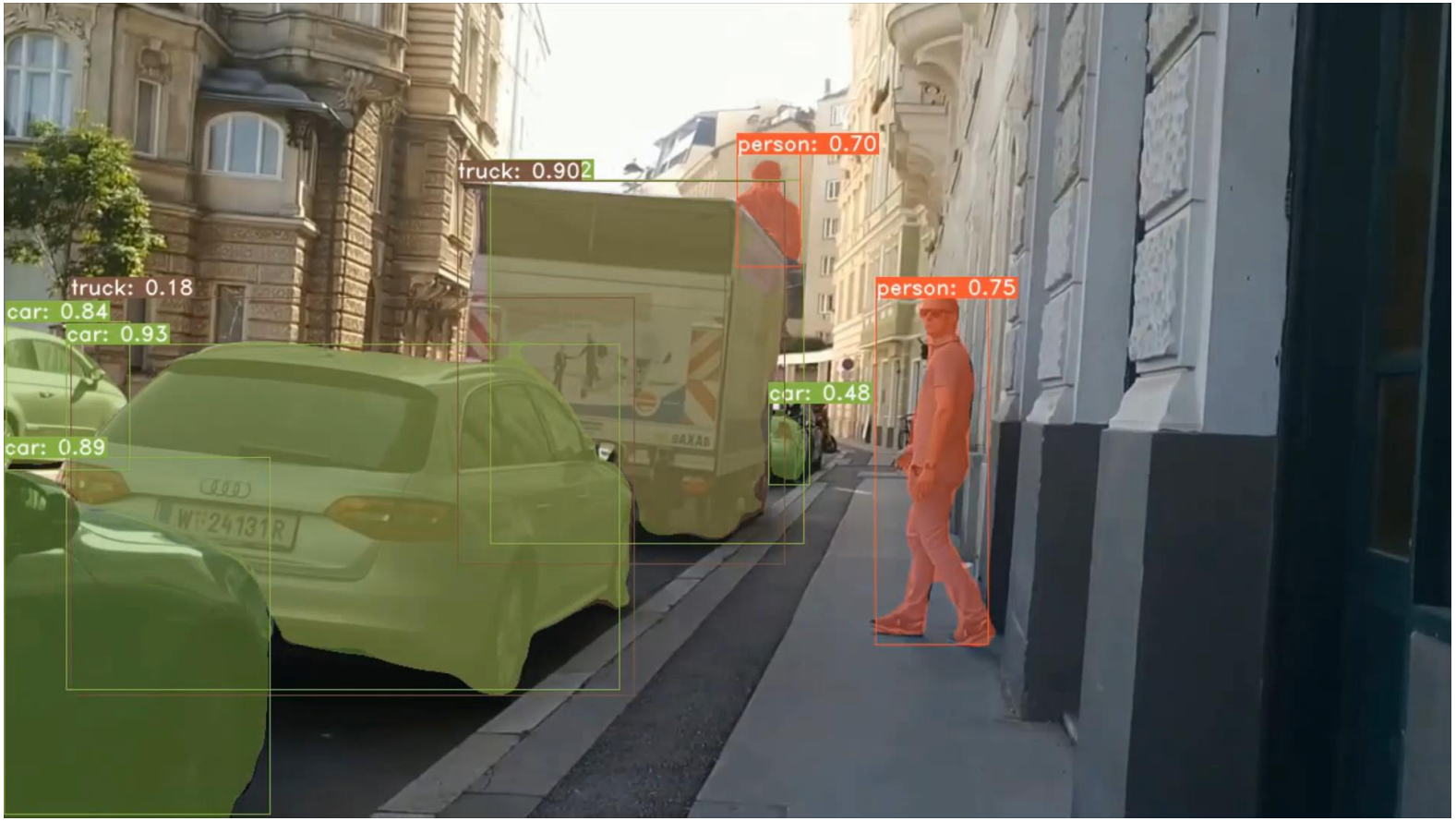


Image enhancement

Indicator of sriking categories

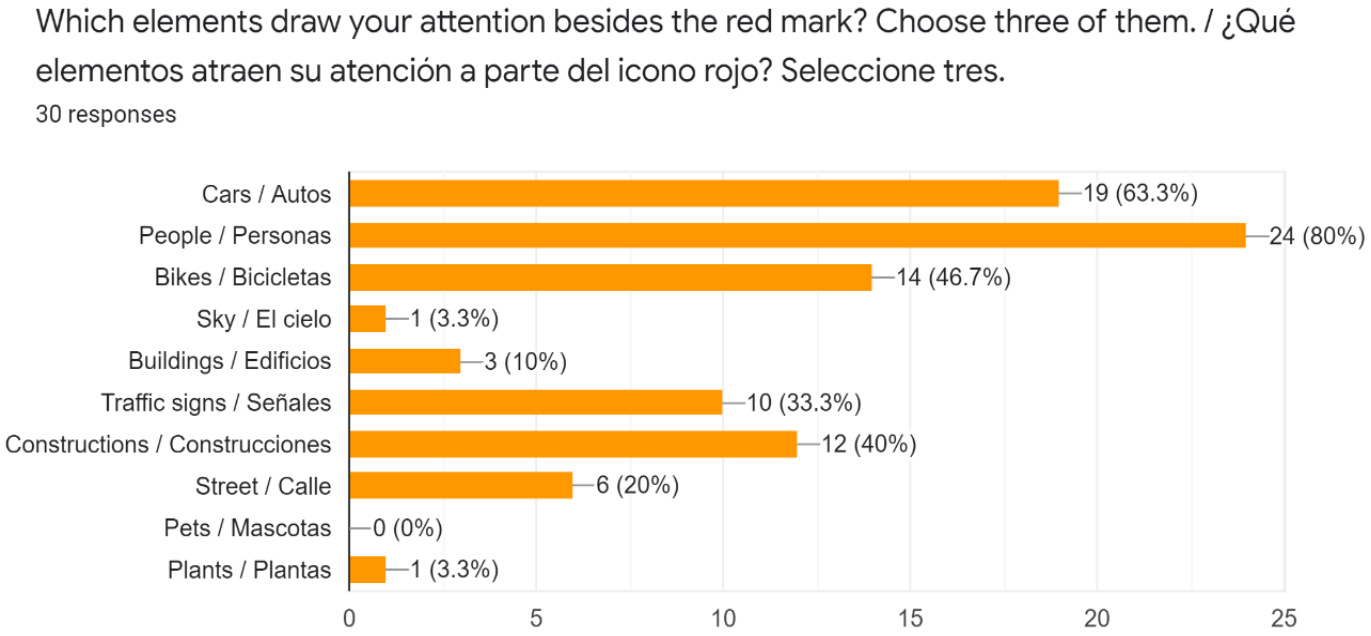


Fig. 13. Indicator of striking objects.

Image enhancement
Masking process

One-minute video segment -> 1,818 frames

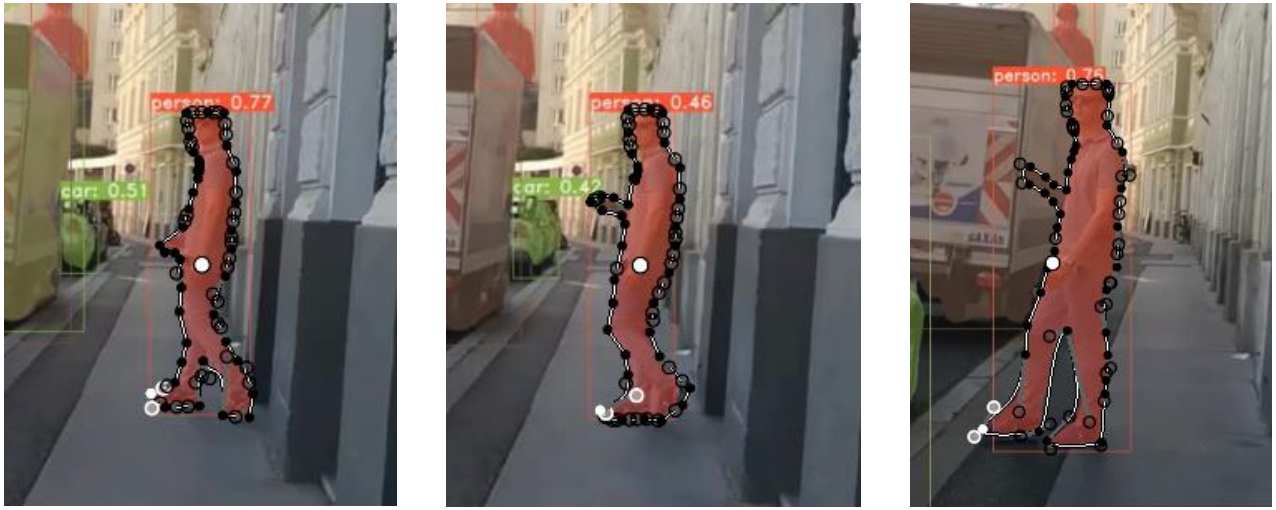


Figure 14. Elaboration of masks in BLENDER.

Image enhancement
Higlighting techniques

Original

Detected object

Contrast enhancement

Brightness enhancement

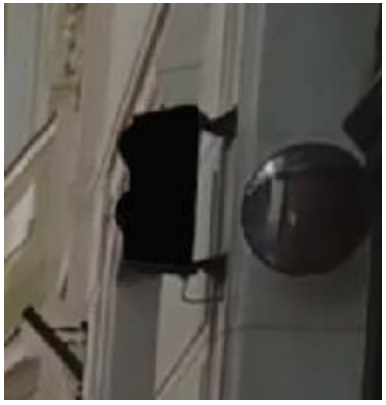
Saturation enhancement

Figure 15. Image enhancement techniques applied to traffic signs.

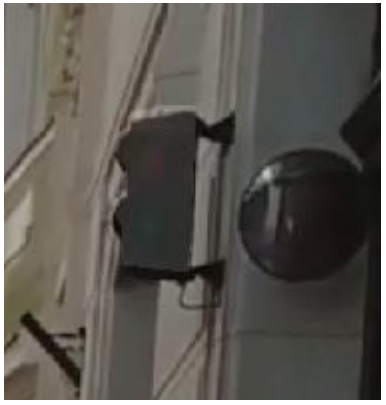
Image enhancement
Higlighting techniques

Original

Detected object



Contrast enhancement



Brightness enhancement

Saturation enhancement

Figure 16. Image enhancement techniques applied to traffic lights.

Image enhancement
Downgrading techniques

Original

Detected object

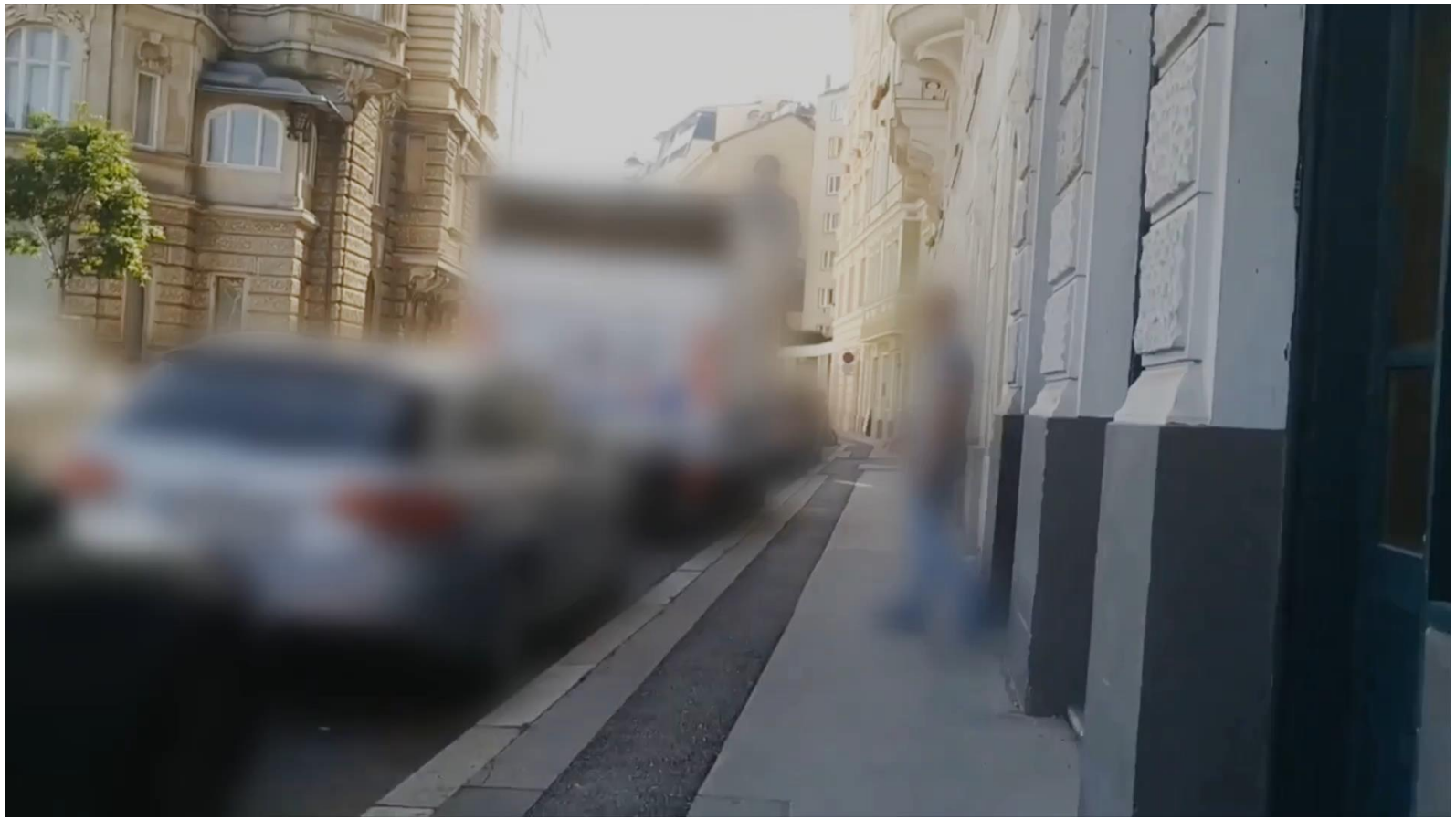
Light cover

Dark cover

Blurriness

Figure 17. Image downgrading techniques.

Implementation

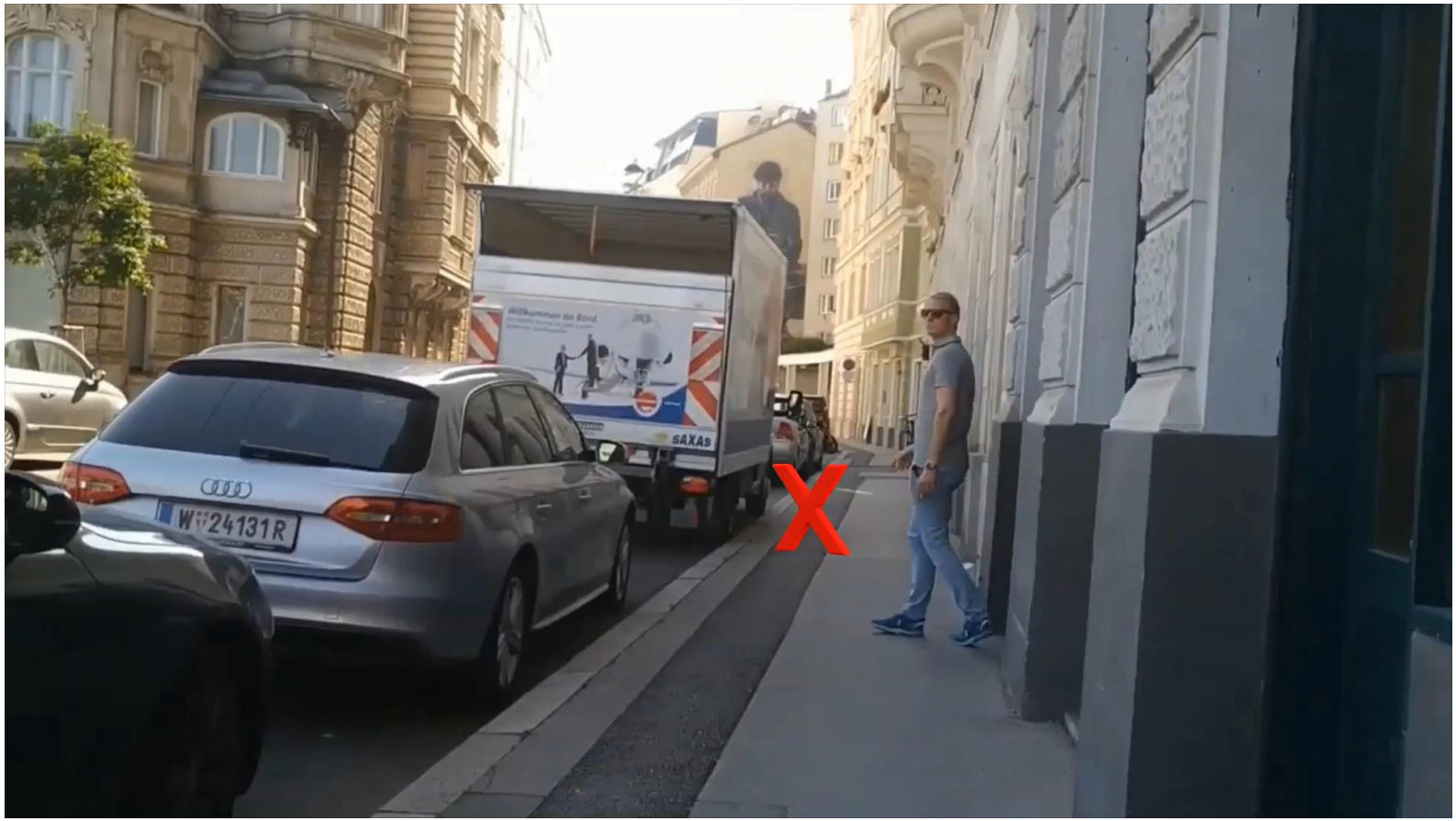


USER TESTS

Generalities

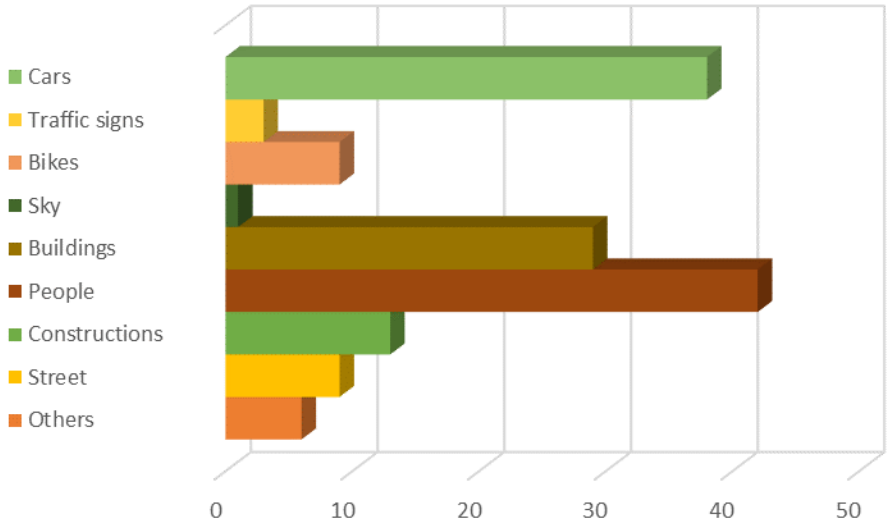
- 2 user tests were conducted
- 50 participants for each test
- Questionaries answered through Google Forms
- An X-shaped rotating three-dimensional object tracing the trajectory through the sidewalk was added to create a MR context
- Users were required to focus on the digital object and answer:
  - Besides the red object, which elements drew your attention the most?
  - Which elements called your attention the least?

User test I: Identification of striking objects



User test I: Identification of striking objects

Besides the red object, which elements drew your attention the most?



Which elements called your attention the least?

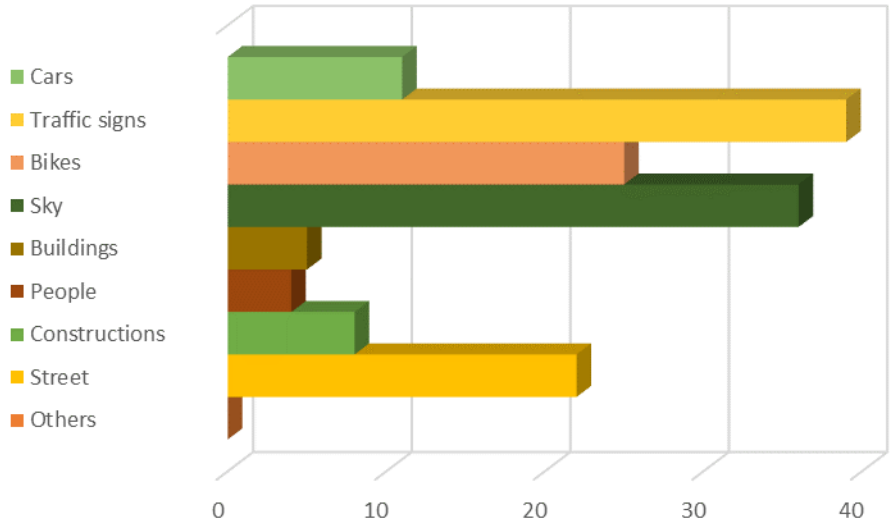
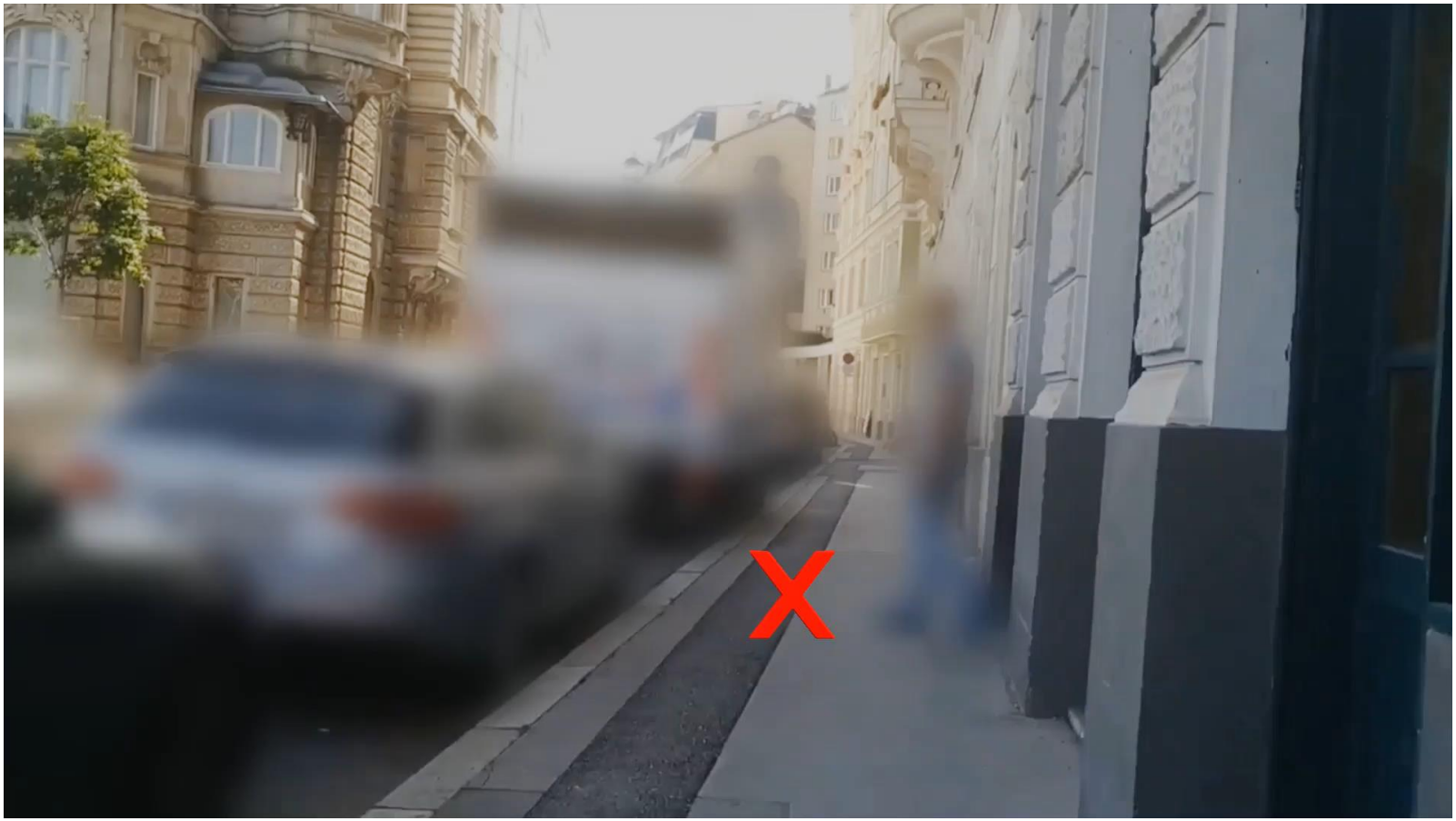


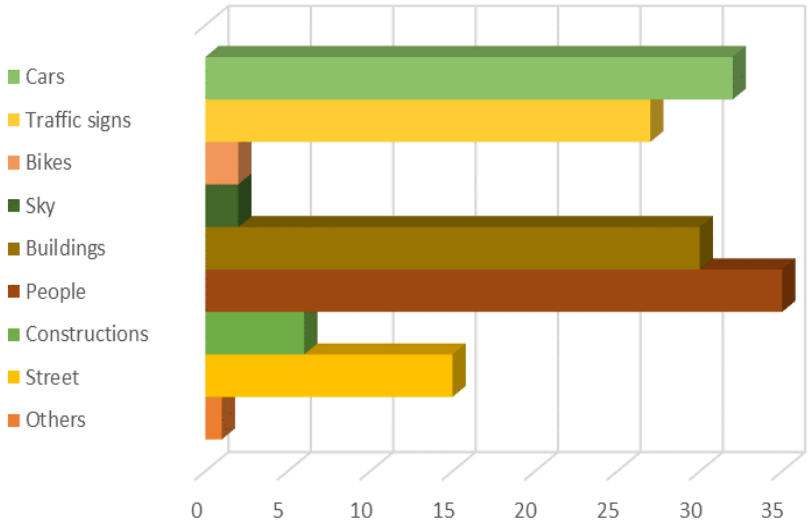
Figure 18. Responses to User test I.

User test II: Identification of striking objects applying image enhancement



User test II: Identification of striking objects applying image enhancement

Besides the red object, which elements drew your attention the most?



Which elements called your attention the least?

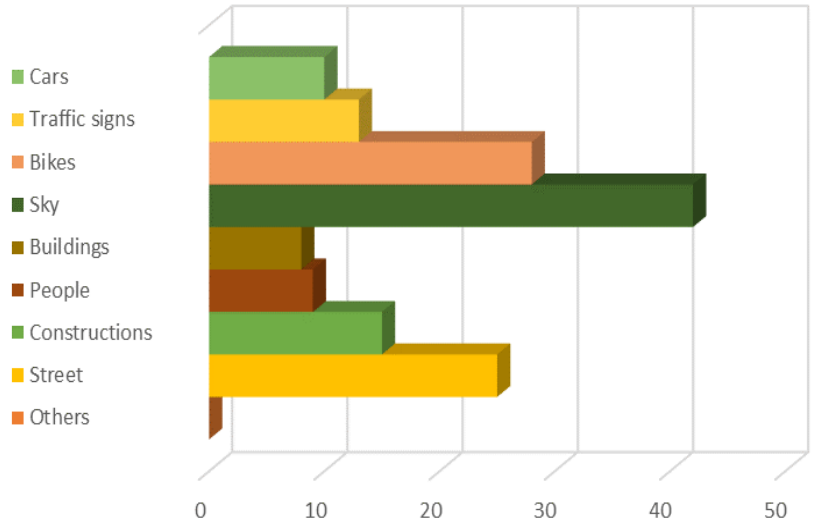


Figure 19. Responses to User test II.

RESULTS

User test findings

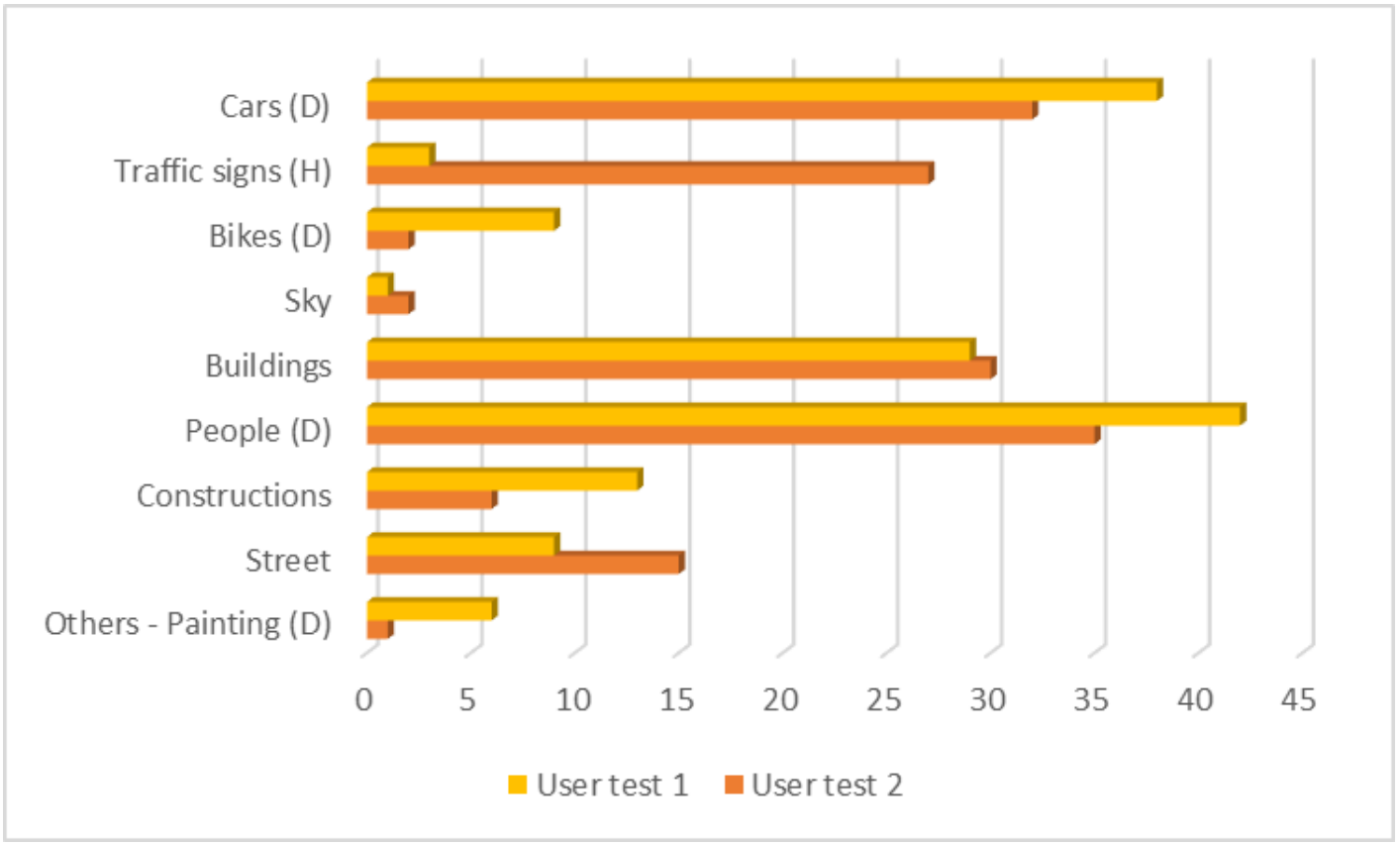


Figure 20. Comparison between user test I and II.

D: downgraded
H: highlighted

Discussion

- User’s attention
  - Highlighted objects
  - Downgraded objects
- Why did objects continue to be striking?
  - Image motion
  - Changes in perception are subjective

Development of research questions

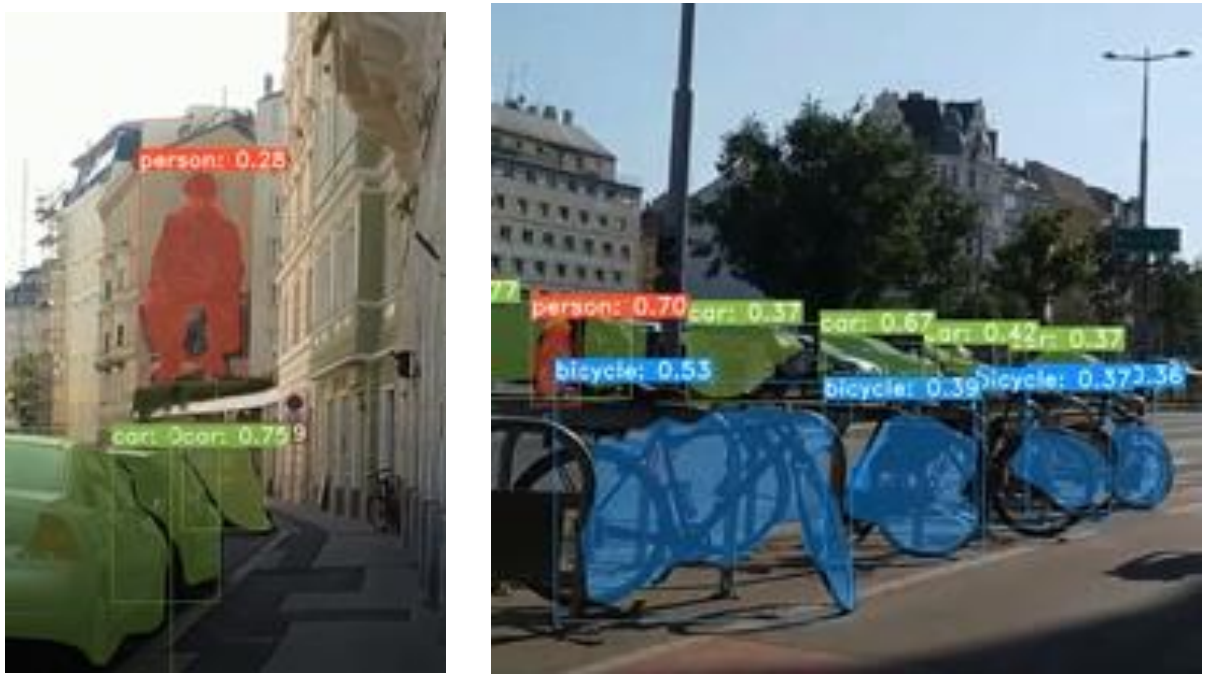
- What is the difference between instance segmentation and object detection?
  - Object detection: bounding-boxes
  - Instance segmentation: object geometries
- Which existing model fits adequately to generate the desired segmentation process?
  - YOLACT
  - Able to process video format
  - Instance segmentation

Development of research questions

- How to integrate an object recognition model with image enhancement techniques to highlight and downgrade elements in a video?
  - Blender version 2.83.0
  - Able to perform image enhancement in each frame
  - Only in segmented regions
- Can a preprocessed image segmentation tool together with image enhancement methods facilitate object recognition in Mixed Reality?
  - User tests demonstrated that highlighted objects were more conspicuous for users

Limitation and recommendations

- Inaccurate segmented regions and misclassification / Training with more classes



(a) (b)

Figure 17. Limitations of YOLACT such as (a) deformed instances and (b) misclassification.

Limitation and recommendations

- Masking is a time-consuming task / Integration with the object recognition model
- User tests
  - Unsatisfactory responses
  - More participants required
- Conduct user tests in different temporalities so that the same users can estimate the improvement of the methodology that was developed

Conclusions

- This thesis aimed at facilitating the navigation experience in the context of Mixed Reality
- Reduced the complexity of recognizing objects by making them more striking
- Downgraded objects continued to be striking
- Integration with navigation apps

Contribution to research

- New approach to visualize object in a MR environment
- Reduced the complexity of detecting objects
- Introduced a visualization technique to visually highlight objects

References

Bolya, D., Zhou, C., Xiao, F., & Lee, Y. J. (2019). YOLACT: Real-time instance segmentation. *Proceedings of the IEEE International Conference on Computer Vision*, *2019-Octob*(1).

Çöltekin, A., Griffin, A. L., Slingsby, A., Robinson, A. C., Christophe, S., Rautenbach, V., Chen, M., Pettit, C., & Klippel, A. (2020). Geospatial Information Visualization and Extended Reality Displays. In H. Guo, M. F. Goodchild, & A. Annoni (Eds.), *Manual of Digital Earth* (pp. 229–264). Springer.

Costanza, E., Kunz, A., & Fjeld, M. (2009). Mixed reality: A survey. *Lecture Notes in Computer Science*, 47–68.

Fast-Berglund, Å., Gong, L., & Li, D. (2018). Testing and validating Extended Reality (xR) technologies in manufacturing ScienceDirect Costing models for capacity optimization in Industry 4.0: Trade-off between used capacity and operational efficiency. *Procedia Manufacturing*, *25*, 31–38.

Murphy, C. (2015). Intellectual highlighting of remote sensing imagery for better image map design. Proceedings of the 27th International Cartographic Conference.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 779–788.

Zhu, Z., Li, D., Hu, Y., Li, J., Liu, D., & Li, J. (2020). Indoor scene segmentation algorithm based on full convolutional neural network. Neural Computing and Applications.

O'Shea, K., & Nash, R. (2015). *An Introduction to Convolutional Neural Networks*.

Thank you for your attention!

Fernando Esquivel
Munich, Germany
29.10.2020


ITC
UNIVERSITY OF TWENTE.
TECHNISCHE
UNIVERSITÄT
DRESDEN
Technical
University
of Munich
TU
WIEN
TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology