Universität Osnabrück
Fachbereich Humanwissenschaften
Institute of Cognitive Science

Bachelorthesis - Expose

Image Segmentation

Sven Groen
970219
Bachelor's Program Cognitive Science
Starting month and year - end month and year

First supervisor: Manuel Kolmet
IMANOX GmbH
Berlin

Second supervisor: Prof. Dr. Someone Else
Institute of Cognitive Science
Osnabrück

Contents

List of Figures

List of Tables

List of Algorithms

1 Goal and background information

1.1 Cooperation with Imanox

This project is realized in cooperation with Imanox . Imanox is a Berlin-based Startup that developed a smart photo booth for expositions, events and promotions. This photo booth enables customers virtual product placements using augmented and mixed reality. Main features are hand-tracking, digital masks and changing virtual backgrounds.

1.2 Virtual backgrounds

Currently, the photo booth has a build in depth sensor that measures the distance of objects by casting illumination onto the scene and indirectly measuring the time it takes to travel back to the camera. The camera struggles with correctly predicting the depth in certain situations. Pixels are rendered as invalid and no depth information is provided. The reasons for this are numerous. Pixels might get under saturated (signal is not strong enough) or over saturated (signal is too strong). Other artifacts occur due to the geometry of the scene. The sensors of the camera might receive signals from multiple locations in the scene, leading to an ambiguous depth. Especially around the edges and borders of objects pixels contain mixed signals from fore-and background leading to blurred outlines. In the current version of the photo booth alpha values (0 to 1) are calculated based on the data from the depth sensor. Objects in the foreground receive high alpha values and the background is considered to have an alpha value of 0, making it transparent. In this way the background can be virtually replaced without affecting the objects in the foreground. Due to the described inaccurate data that is given by the depth sensor the result is of low quality. The edges and borders of the objects/people in the scene are not sharp and often misclassified. Especially for small / thin objects, e. g. hair, the camera hardly recognizes it and parts of the hair are therefore considered as background and are also replaced by the virtual background. For more detailed information on this issue

I refer to https://docs.microsoft.com/de-de/azure/Kinect-dk/depth-camera. [1]

1.3 Goal of the thesis

The goal of this bachelor thesis is to improve the quality of the semantic segmentation of the current Imanox photo booth using machine learning techniques. We will use ...

2 Segmentation

Szeliski [2] refers to image segmentation as "the task of finding groups of pixels that 'go together' " (p. 237). In the following semantic segmentation refers to a pixel-wise classification of an image [3]. In the classical image classification tasks the task is to name the objects that can be seen in an image. Semantic segmentation extends this problem further. Each pixel in an image is assigned to one category label given a set of categories. However, individual instances of an object in one image are not differentiated. When individual instances in an image should be recognized, object detection is necessary. For single objects this would be a classification + localization task. Object detection is usually realized by framing the object with a box and assigning a category label to each box. Lastly, there is instance segmentation. Instance segmentation extends the problem of object detection by a pixel-wise classification (similar to semantic segmentation) but with instances being differentiated [3]. See (Figure 2.1) for an overview of the described tasks. Given the goal of this project only semantic segmentation is necessary. Detailed information which objects are in the scene is not required.

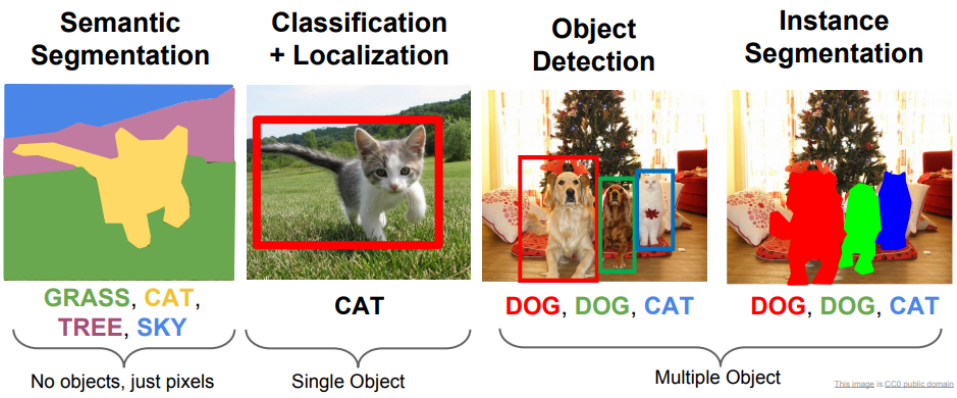

Figure 2.1: Overview of Computer Vision Tasks. (Li *et al.* [4], Slide 17).

2.1 related work

Segmenting an image into its individual parts is a classical problem of computer vision [2]. Early methods involve classical methods like threshold detection [5], while modern approaches like k-means clustering [6] improved the results. Deep learning architectures, especially convolutional neural networks (CNNs) [7], have lead to an improvement in performance whereas classical methods have seem to reach a plateau (Figure 2.2). Shelhamer *et al.* [8] have been the first to proposed a CNN architecture where a pixel-wise supervised training was achieved. This was done by upsampling the class prediction layer to the input image size, leading to a pixel-wise classification. Following papers proposed different architectures. Chen *et al.* [9] proposed a combination of Deep CNNs with fully connected conditional random fields (CRFs) that tries to grasp the semantic context of the image. Noh *et al.* [10] suggested a "Deconvolutional Network" with special unpooling and deconvolution operations. A similar Encoder-Decoder architecture has been proposed by Badrinarayanan *et al.* [11].

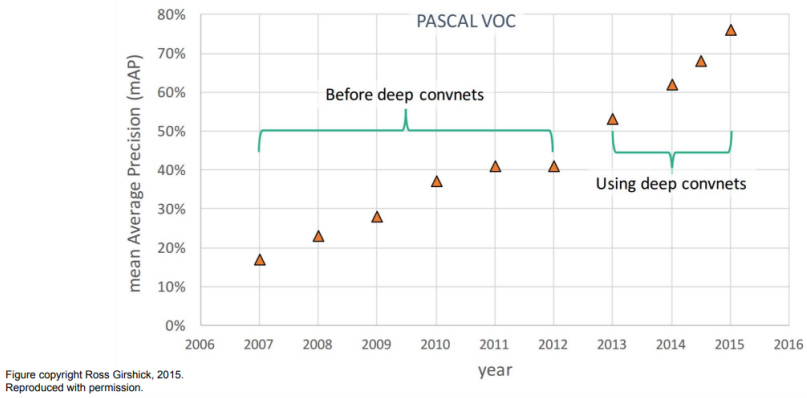

Figure 2.2: Mean average precission (mAP) of object detection before and after using deep learning techniques. (Li *et al.* [4], Slide 54).

2.2 Challenges

3 Related work

4 Methods

4.1

5 Time frame

6 Conclusion

7 Bibliography

1. Sych, T., Brent, A., Phil, M. & Microsoft. *depth-camera @ docs.microsoft.com* 2019.
2. Szeliski, R. *Computer Vision: Algorithms and Applications* 185–186. doi:10.1017/cbo9780511974076.010 (Springer, 2011).
3. Mittal, M., Arora, M., Pandey, T. & Goyal, L. M. Image Segmentation Using Deep Learning : A Survey, 41–63. doi:10.1007/978-981-15-1100-4_3 (2020).
4. Li, F.-F., Johnson, J. & Yeung, S. *Lecture 11: Detection and Segmentation* 2017.
5. Nobuyuki, O. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* **9,** 62–66. doi:10.1109/TSMC.1979.4310076 (1979).
6. Dhanachandra, N., Manglem, K. & Chanu, Y. J. Image Segmentation Using K-means Clustering Algorithm and Subtractive Clustering Algorithm. *Procedia Computer Science* **54,** 764–771. doi:10.1016/j.procs.2015.06.090 (2015).
7. Fukushima, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics* **36,** 193–202. doi:10.1007/BF00344251 (1980).
8. Shelhamer, E., Long, J. & Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39,** 640–651. doi:10.1109/TPAMI.2016.2572683 (2017).
9. Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A. L. Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **40,** 834–848. doi:10.1109/TPAMI.2017.2699184 (2018).

10. Noh, H., Hong, S. & Han, B. Learning deconvolution network for semantic segmentation. *Proceedings of the IEEE International Conference on Computer Vision* **2015 International Conference on Computer Vision, ICCV 2015,** 1520–1528. doi:10.1109/ICCV.2015.178 (2015).

11. Badrinarayanan, V., Kendall, A. & Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39,** 2481–2495. doi:10.1109/TPAMI.2016.2644615 (2017).