# Semi-automated Approach for Annotating Data in Perception Network (EN)

Nienburger Straße 1, 30167 Hannover
Fakultät für Bauingenieurwesen und Geodäsie

Institut für Photogrammetrie
und GeoInformation

Prof. Dr.-Ing. habil. Christian Heipke

Tuan Nguyen

Tel. +49 511 762 2488
Fax +49 511 762 2483
E-Mail:
tuan.nguyen@ipi.uni-hannover.de

12 December 2023

The perception network using this annotation data aims at the development of methods for improving collaborative positioning of vehicles in areas without adequate GNSS coverage on the basis of stereo image sequences. Focusing on an improvement of the perception of the environment, the poses and shapes of vehicles in the surrounding of the ego-position are of particular interest. Thus, the CNN-based method is developed utilizing data annotation from our proposed method, that integrates the detection and semantic classification of objects via panoptic segmentation (Alexander Kirillov, et al., 2019) with the reconstruction of the 3D geometry of these objects. Assuming that the scene is dynamic, with respect to ego-motion as well as moving objects, temporal information is considered to be beneficial and is taken into account in the network. The end-to-end architecture is depicted in Figure 1, where the red rectangle represents the anticipated output generated by the network, trained on data produced by our methodology.

The anticipated annotation serves as the training data for our method, including three key subtasks: panoptic segmentation, video panoptic segmentation, and shape and pose reconstruction. Initially, panoptic segmentation is annotated in each image within the sequence. The annotation involves creating masks for each instance, followed by assigning a unique ID to the masks of the same instance across images in the sequence, thereby supervising the subsequent video panoptic segmentation task. While manual annotation for segmentation tasks demands considerable human effort, recent advancements in techniques, such as Segment Anything (Alexander Kirillov et al., 2023), or the application of Active Learning methods (Soufiane Belharbi et al., 2021), allow for a reduction in this effort. Additionally, the automation of 3D shape and pose reference is achievable through the utilization of lidar points corresponding to the object.

Within the framework of a master thesis, we explore a semi-automated annotation method applied to our own dataset, comprising stereo camera and lidar data. The initial phase involves experimentation with each available support technique to scrutinize their advantages and drawbacks. Subsequently, the analysis extends to exploring potential improvements or combinations of different methods, aiming to enhance the performance of the semi-automated framework. Another approach involves considering data from various sensors; given the high accuracy of lidar points, their synchronization with images can be leveraged, and recent segmentation methods based on lidar can be employed for additional support.

This thesis will be supervised by Tuan Nguyen.

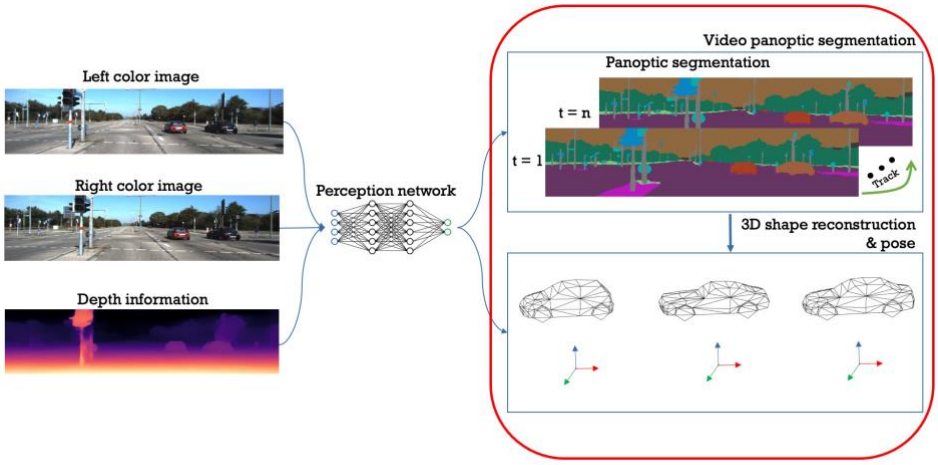


Figure 1. The perception network with an input as sequence of color and depth image. The expected output is in the red rectangle, which will use data annotation from the proposed method.

## References

Kirillov, A., He, K., Girshick, R., Rother, C. and Dollár, P., 2019. Panoptic segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9404-9413).

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y. and Dollár, P., 2023. Segment anything. arXiv preprint arXiv:2304.02643.

Belharbi, S., Ben Ayed, I., McCaffrey, L. and Granger, E., 2021. Deep active learning for joint classification & segmentation with weak annotator. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 3338-3347).