



October 22, 2021

MASTER'S THESIS
for
Zhifan Ni
Student ID 03683316, Degree EI

Human-Object Interaction prediction in videos through Gaze-Following

Problem description:

In this thesis project, the student will build a system that can predict the Human-Object Interaction (HOI) in the near future from videos, by considering the human eye-gaze information. This work is inspired by several biological human behavior researches, [1] which claim that humans tend to gaze at an object before interacting with that object.

The student will explore the use of gaze-following methodologies [2] as an additional novel step towards HOI benchmark in videos [3]. Regarding this, the student would be required to propose a deep learning-based HOI prediction model which will learn the spatio-temporal relationship between human and objects. Compared to the existing similar related works that use eye-gaze information for HOI detection in a single image [4], the main contribution of this work would be using the gaze information for HOI prediction in videos.

Tasks:

- Literature review on Human Object Interaction and Gaze-Following.
- Integration of Tracking-Detection system for locating Human and Objects through time with Gaze-Following existing models.
- Design and implement an attention-based Neural network for Gaze-Following.
- Design and implement a deep-learning model to combine Gaze-Focus and human/object representations for HOI prediction.
- Evaluation and ablation study based on baseline models

Bibliography:

- [1] Michael F. Land and Mary Hayhoe. In what ways do eye movements contribute to everyday activities? *Vision Research*, 41(25):3559–3565, 2001.
- [2] Eunji Chong, Yongxin Wang, Nataniel Ruiz, and James M. Rehg. Detecting attended visual targets in video, 2020.
- [3] Meng-Jiun Chiou, Chun-Yu Liao, Li-Wei Wang, Roger Zimmermann, and Jiashi Feng. St-hoi: A spatial-temporal baseline for human-object interaction detection in videos. *Proceedings of the 2021 Workshop on Intelligent Cross-Data Analysis and Retrieval*, Aug 2021.
- [4] Bingjie Xu, Junnan Li, Yongkang Wong, Mohan S. Kankanhalli, and Qi Zhao. Interact as you intend: Intention-driven human-object interaction detection, 2019.

Supervisor: Dipl.-Ing. Esteve Valls and Dr. Hyemin Ahn
Start: 01.11.2021
Intermediate Report: XX.02.2022
Delivery: XX.05.2022

(D. Lee)
Univ-Professor