DETIC-TRACK: ROBUST DETECTION AND TRACKING OF OBJECTS IN VIDEO

Hannes Fassold

JOANNEUM RESEARCH - DIGITAL

ABSTRACT

The automatic detection and tracking of objects in a video is crucial for many video understanding tasks. We propose a novel deep learning based algorithm for object detection and tracking, which is able to detect more than 1,000 object classes and tracks them robustly, even for challenging content. The robustness of the tracking is due to the usage of optical flow information. Additionally, we utilize only the part of the bounding box corresponding to the object shape for the tracking.

***Index Terms*—** object detection, tracking, optical flow, deep learning

1. ALGORITHM

The automatic detection and tracking of general objects in a video provides semantic information which is crucial for many high-level computer vision tasks in various application areas like surveillance, autonomous driving, automatic video annotation and brand monitoring. Our proposed *Detic-Track* algorithm for object detection and tracking is based upon the OmniTrack algorithm [1], which combines the YoloV4 object detector with TV-L1 optical flow and is real-time capable. For the *Detic-Track* algorithm, we have extended this algorithm in several ways. Firstly, we replace the YoloV4 object detector with the *Detic* detector proposed in [2]. In contrast to YoloV4 (which detects only the 80 MS-COCO classes), Detic is able to detect significantly more object categories. Specifically, we employ a pretrained Detic model which detects the $1,203$ object classes from the LVIS dataset [1]. Furthermore, instead of using the whole bounding box for tracking a detected object, we utilize only the part of the bounding box corresponding to the object mask. This improves the tracking considerably, especially for moving objects whose bounding box contains significant background areas.

2. DEMO APPLICATION

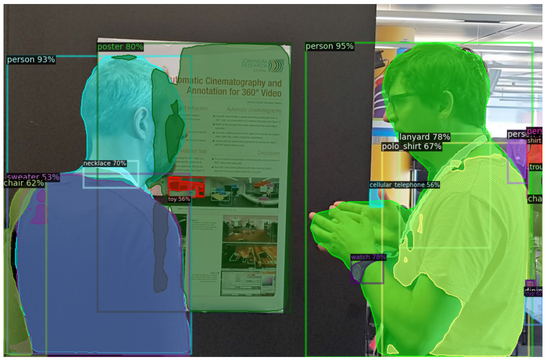

Fig. 1. Result of *Detic* object detector [2] (bounding boxes and instance segmentation for detected objects).

The demo application is a mixed C++/Python application, which runs on a Intel PC equipped with a CUDA-capable NVIDIA GPU. The Detic object detector [2] runs in Python using the implementation provided at [2], whereas the other components of the detection and tracking framework (see [1] for details) run in C++. The communication and synchronisation between the python and C++ components is done via inter-process communication, using shared memory. The demo application opens a video, runs the *Detic-Track* algorithm on it and visualizes the detected and tracked objects in a GUI window. The demo application is not real-time capable, as the Detic object detector needs roughly 400 milliseconds for one image in resolution $896x480$ on the A4000 GPU.

Acknowledgments

This work was supported by European Union´s Horizon 2020 research and innovation programme under grant number 951911 - AI4Media.

3. REFERENCES

[1] Hannes Fassold and Ridouane Ghermi, “OmniTrack: Real-time detection and tracking of objects, text and logos in video,” in *Proc. ISM*, 2019.

[2] Xingyi Zhou, Rohit Girdhar, Armand Joulin, Philipp Krähenbühl, and Ishan Misra, “Detecting twenty-thousand classes using image-level supervision,” *CoRR*, vol. abs/2201.02605, 2022.

[1] `https://www.lvisdataset.org/`

[2] `https://github.com/facebookresearch/Detic`