Master of Science Topics

Autonomous Driving

Title: Pedestrian detection for autonomous driving

Coordinator: *Prof. Adina Magda Florea (adina.florea@upb.ro)*

Description:

Pedestrian detection is used in many vision-based applications ranging from video surveillance to autonomous driving. Despite achieving high performance, it is still largely unknown how well existing detectors generalize to unseen data.

Advances in pedestrian detection systems can dramatically improve the performance and robustness of applications, which in some cases (e.g. accident avoidance in autonomous vehicles) may even save human lives.

Convolutional Neural Networks (CNNs) have become the dominant paradigm in generic object detection and were also applied for pedestrian detection. Some of the pioneer works for CNN based pedestrian detection used the R-CNN framework and RPN+BF (Region Proposal Network). Afterwards, Faster RCNN became the most popular framework. Some of the recent state-of-the-art pedestrian detectors include ALF [1], CSP [2] and MGAN [3]. ALF is based on Single Shot MultiBox Detector (SSD), it stacks together multiple predictors to learn a better detection from default anchor boxes. MGAN uses the segmentation mask of the visible region of a pedestrian to guide the network attention and improve performance on occluded pedestrians. CSP is an anchor-less fully convolutional detector, which utilizes concatenated feature maps for predicting pedestrians.

Several datasets have been proposed from the context of autonomous driving such as KITTI [4], Caltech [5], CityPersons [6] and ECP [7]. Typically, these datasets are captured by a vehicle-mounted camera navigating through crowded scenarios. A recent large dataset is Wider Pedestrian [8].

Tasks:

- investigate the properties of the different DNN architectures for pedestrian detection;
- implement at least 2 different models;
- propose improvements;
- perform evaluations of the implementation on at least 3 datasets.

[1] Liu, W., Liao, S., Hu, W., Liang, X., Chen, X.: Learning efficient single-stage pedestrian detectors by asymptotic localization fitting. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 618–634 (2018).

[2] Liu, W., Liao, S., Ren, W., Hu, W., Yu, Y.: High-level semantic feature detection: A new perspective for pedestrian detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2019).

[3] Pang, Y., Xie, J., Khan, M.H., Anwer, R.M., Khan, F.S., Shao, L.: Mask-guided attention network for occluded pedestrian detection (2019).

[4] Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3354–3361. IEEE (2012).

[5] Dollar, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: An evaluation of the state of the art. IEEE transactions on pattern analysis and machine intelligence 34(4), 743–761 (2012).

[6] Zhang, S., Benenson, R., Schiele, B.: Citypersons: A diverse dataset for pedestrian detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3213–3221 (2017).

[7] Braun, M., Krebs, S., Flohr, F., Gavrila, D.M.: Eurocity persons: A novel benchmark for person detection in traffic scenes. IEEE transactions on pattern analysis and machine intelligence 41(8), 1844–1861 (2019).

[8] Wider pedestrian 2019, https://competitions.codalab.org/competitions/20132.

Title: Pedestrian tracking for autonomous driving

Coordinator: *Prof. Adina Magda Florea (adina.florea@upb.ro)*

Description:

The purpose of multi-target tracking is to provide accurate trajectories of moving targets from given observations. The produced trajectories are used for position prediction or re-identification. For instance, in autonomous vehicle application, it prevents traffic accidents by predicting movement of pedestrians or vehicles.

Many online multi-target trackers follow a Bayesian tracking process. It predicts a state of each track using previously assigned observations. Based on this prediction, likelihoods between tracks and new observations are calculated to form a cost matrix.

Some years ago, the trend on people detection and tracking from video sequences was to find strong, preferably optimal methods to solve the data association problem. Linking detections in a set of consistent trajectories (matching two detections based on either simple distances or weak appearance models) was solved by various methods such as Conditional Random Fields or as a variational Bayesian model; performances were not very good.

More recently, the focus is on building robust pairwise similarity costs, mostly based on strong appearance cues, leading to better tracker performances and more complex scenarios. Some good approaches use sparse appearance models [1] or integral channel feature appearance models [2] or aggregated local flow of long-term interest point trajectories [3] to improve detection affinity. Still, most of the available tracking approaches do not include a learning algorithm to determine the set of model parameters for a dataset.

Some recent approaches use deep learning, such as recurrent neural networks, to encode appearance, motion, and interactions [4] or deep matching to improve the affinity measure [5] or tracking in occluded scenes [6].

Tasks:

- investigate different methods for pedestrian tracking;
- implement at least 2 different models;
- propose improvements;
- perform evaluations of the implementation on at least 3 datasets, at least one of MOT Challenge [7].

[1] L. Fagot-Bouquet, R. Audigier, Y. Dhome and F. Lerasle, "Improving multi-frame data association with sparse representations for robust near-online multi-object tracking". In: Proc. of European Conf. on Computer Vision, pp.774-790, Springer, Cham, 2016.

[2] H. Kieritz, S. Becker, W. H¨ubner and M. Arens, "Online multi-person tracking using integral channel features". In: Proc. of Conf. on Advanced Video and Signal Based Surveillance, pp.122-130, 2016.

[3] W. Choi, "Near-online multi-target tracking with aggregated local flow descriptor". In: Proc. of Conf. on computer vision, 2015, pp.3029-3037, 2015.

[4] A. Sadeghian, A. Alahi, and S. Savarese, "Tracking the untrackable: Learning to track multiple cues with long-term dependencies". In: Proc. of Conf. on Computer Vision, pp.300-311, arxiv:1701.01909, 2017.

[5] S. Tang, B. Andres, M. Andriluka and B. Schiele, "Multiperson tracking by multicuts and deep matching". In: Proc. of European Conf. on Computer Vision, Amsterdam, The Netherlands, pp.100-111, 2016.

[6] E. Haq, H. Jianjun, K. Li, H. Haq. Human detection and tracking with deep convolutional neural networks under the constrained of noise and occluded scenes. Multimedia tools and applications, 2020.

[7] https://motchallenge.net/

https://towardsdatascience.com/people-tracking-using-deep-learning-5c90d43774be

Title: Pedestrian trajectory analysis in outdoor environments

Coordinators: *Prof. Adina Magda Florea (adina.florea@upb.ro)*
Assist. Prof. Mihai Trascau (mihai.trascau@upb.ro)
drd. ing. Ștefania Ghiță (stefania.a.ghita@upb.ro)

Description:

The trajectory prediction problem is an important part of autonomous systems, in both robot navigation and autonomous driving applications. This problem improves tracking accuracy and prevents collision between autonomous robotic platforms and moving objects (people or other autonomous platforms). This project focuses on analyzing pedestrian trajectories based on visual data extracted from videos. The analysis can include:

- a tracking component applied on individuals or groups of people;
- an observation component which computes the general flows of pedestrians and constructs heat maps based on them;
- an estimation component to predict possible trajectories of the tracked targets.

The solution will be adapted to outdoor environments, which may imply very crowded places

Title: Car crash detection in video using neural networks

Coordinators: *Assist. Prof. Mihai Trascau (mihai.trascau@upb.ro)*
drd. ing. Mihai Nan (mihai.nan.cti@gmail.com)

Description:
Increasing the number of cars and excessive traffic congestion in cities is a major problem in the current time. Statistics show that more and more accidents happen daily, and many of these could be avoided.
The aim of this research topic is to develop a system capable of detecting the possibility of an accident by analyzing a video sequence. Starting from the available datasets (ex. CADP, CarCrashDetector), we propose to implement a system that combines Computer Vision and Machine Learning techniques to identify possible car crash.

Title: Depth estimation for improved scene understanding in autonomous driving scenarios

Coordinators: *dr. ing. Alexandru Sorici (alexandru.sorici@upb.ro)*
drd. ing. Iancu David (davidiancudvd@gmail.com)
drd. ing. Vlad Florea (vlad.florea1709@upb.ro)

Description:
Autonomous driving is a hot research topic today but there are still a lot of challenges to solve until we can trust a completely autonomous car. One of the biggest problems is how to estimate accurate distances to surrounding objects, such as other cars, to approximate the braking time and distance, to avoid collisions, etc.
This is still an open problem as all existing depth estimation methods are susceptible to noise due to reflective surfaces, occlusions, and weather conditions such as heavy rain or other particles in the air. These issues are present for LiDAR sensors as well as for stereo and monocular cameras. In order to achieve accurate predictions state of the art literature uses cutting edge Deep Learning techniques to identify patterns in the data and mitigate these problems by introducing common world knowledge into the depth estimation process, much like a human driver would think (e.g. we can only see a small part of a truck in front of us due to heavy rain, but we assume its rough size because we know this is a truck and not some random object). An insight on how such models might “think” can be seen in [1].

While there are open datasets with LiDAR ground truth information on which students can evaluate their models (https://waymo.com/open/, https://level-5.global/data/ , etc.), the ultimate goal of this project is to evaluate state of the art depth estimation methods on our own dataset, which contains about 140 km driven through the University Politehnica of Bucharest campus (50% driven at night) and no ground truth LiDAR information, such that all proposed techniques should leverage computer vision based methods.

As a starting point you can explore techniques such as [2] and [3] and then extend to more powerful models which make use of multiple sensors on the car (e.g. the IMU) to improve precision [4]. Your research can also branch out to domains like 3D scanning of the UPB campus streets (by integrating depth predictions over time into a growing point cloud), visual odometry (https://www.youtube.com/watch?v=M4v_-XyYKHY) or attempt to build accurate vector space representations of the real world, to be used for inference by self-driving algorithms, with methods such as the ones presented by Tesla's Andrej Karpathy this year: https://youtu.be/j0z4FweCy4M?t=3409.
In the end students have the opportunity to publish their solutions as open source software and contribute to the continually growing NemoDrive project run by our laboratory: https://nemodrive.cs.pub.ro/research/.

[1] Tom van Dijk and Guido de Croon. How do neural networks see depth in single images In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 2183–2191, 2019. https://arxiv.org/pdf/1905.07005
[2] Tinghui Zhou, Matthew Brown, Noah Snavely, and David G Lowe. Unsupervised learning of depth and ego-motion from video. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1851–1858, 2017. https://arxiv.org/pdf/1704.07813
[3] Clément Godard, Oisin Mac Aodha, Michael Firman, and Gabriel J Brostow. Digging into self-supervised monocular depth estimation. In Proceedings of the IEEE international conference on computer vision, pages 3828–3838, 2019. https://arxiv.org/pdf/1806.01260
[4] Vitor Guizilini, Rares Ambrus, Sudeep Pillai, Allan Raventos, and Adrien Gaidon.3D packing for self-supervised monocular depth estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2485–2494, 2020. https://arxiv.org/pdf/1905.02693