Understanding image classification based on deep neural networks – 2 topics

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

While deep neural networks have had a large impact on a variety of different tasks, explaining their predictions is still challenging. The lack of tools to inspect the behavior of these models makes DNNs less trustable for those domains where interpretability and reliability are crucial, like autonomous driving, medical applications, and finance.

One of the main approaches to understand the behaviour of a DNN is assigning an attribution value, also called "relevance" or "contribution", to each input feature of a network. In this context, we propose 2 research topics:

1. Understanding image classification tasks using Layer-wise Relevance Propagation

Layer-wise Relevance Propagation (LRP) allows to visualize the contributions of pixels to predictions for multilayered neural networks. These pixel contributions can be visualized as heatmaps and are provided to a human expert who can intuitively not only verify the validity of the classification decision, but also focus further analysis on regions of potential interest. LRP is computed with a backward pass on the network. The method assigns a relevance to units of different layers of the network. It starts at the output layer L and assigns the relevance of the target neuron equal to the output of the neuron itself and the relevance of all other neurons to zero. Then it proceeds layer by layer, redistributing the prediction score until the input layer is reached.

Tasks

- investigate the properties of LPR to understanding deep learning classification tasks;
- implement the approach;
- propose improvements;
- perform evaluations of the implementation, both technical and nontechnical.

S. Bach, A. Binder et.al. On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation, pp.1-46, PLOS One, 2015.

A. Binder, G. Montavon, S. Lapuschkin, K.R. Müller and W. Samek, "Layer-wise relevance propagation for neural networks with local renormalization layers". In: Proc. of Conf. on Artificial Neural Networks, pp.63-71, Springer, 2016.

2. Understanding image classification tasks by Class Activation Mapping

Class Activation Mapping (CAM) applies to CNN and uses global average pooling, which acts as a structural regularizer, preventing overfitting during training. The global average pooling also allows the network to retain its localization ability until the final layer. This allows identifying the discriminative image regions and create activation maps that act as a detector for different patterns in the image, localized in space. By inspecting these activations maps, correct and incorrect classifications can be analyzed.

Tasks

- investigate the properties of this approach to understanding deep learning classification tasks, including Gradient-weighted Class Activation Mapping (Grad-

CAM), which extends CAM by producing discriminative heatmaps that achieve object localization without training;

- implement the approach;
- propose improvements;
- perform evaluations of the implementation.

B. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba, "Learning deep features for discriminative localization". In: Proc. of Conf. on Computer Vision and Pattern Recognition, pp.2921-2929, 2016.

R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization". In: Proc. of Conf. on Computer Vision, pp.618-626, 2017.

https://alexisbcook.github.io/2017/global-average-pooling-layers-for-object-localization/

Pedestrian detection for autonomous driving

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

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 was also applied for the pedestrian detection. Some of the pioneer works for CNN based pedestrian detection used R-CNN framework and RPN+BF (Region Proposal Network). Afterwaords, 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

Pedestrian tracking for autonomous driving

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

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

ML methods for space surveillance

Coordinators: Prof. Adina Magda Florea (adina.florea@upb.ro), Assist. Prof. Mihai Trascau (mihai.trascau@upb.ro)

Tasks

- Develop ML algorithms able to process images from optical telescopes, to identify possible objects of interest, correlate the measurements in successive images according to the observational scheme, and extract accurate information on the position and magnitude of detected objects
- Detect and recognize interesting object in images taken by telescopes, such as stars, satelite or cosmic ray.
- Tracklet generation. Use ML algorithms to identify and predict the trend of the tracklet (tracklet pattern), in other words connecting consecutive images in an observational sequence to generate object tracklets.

IntAli: Food recognition, food recommendation and autonomous detection of food in a refrigerator

Coordinators: Prof. Adina Magda Florea (adina.florea@upb.ro), Assist. Prof. Mihai Trascau (mihai.trascau@upb.ro)

Tasks

- Develop an intelligent application to recognize foods and cooks from images, recommend healthy food and recipes, automatically detect foods in a smart refrigerator and recommend necessary supplies.
- Allow the user to personalize the application according to his/her preferences, food dietary restriction, and available ingredients.

Already allocated

The virtual patient

Coordinators: Prof. Adina Magda Florea (adina.florea@upb.ro), Dr. Alex Awada (alex.awada@upb.ro)

Tasks

- Develop an intelligent agent and a virtual patient to provide students in the medical domain with an interactive class, to cope for their limited ability to attend live classes during the pandemic.
- Develop a virtual on-line class
- We want to create a framework system of interaction between the student and the patient, which is developed initially for the symptoms of pulmonary thromboembolism (PET), and then, after validation of this use case, to be extended to other symptoms and to make a complex diagnosis depending on the symptoms.
- Collaboration with prof. med dr. Alexandru Scafa Udriște (UMF and Spitalul de Urgență Floreasca)

Teaching robots with adversarial imitation learning

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

Imitation learning is the process by which one agent tries to learn how to perform a certain task using information generated by another, an expert agent performing that same task. Recently, the specific problem of imitation from observation (IfO) has received much attention. In this case the imitator only has access to the state information (e.g., video frames) generated by the expert. One category of methods is represented by adversarial approaches to IfO, e.g., the generative adversarial imitation learning (GAIL) algorithm [1], which learns explicitly how to act by directly learning a policy and draws an analogy between imitation learning and generative adversarial networks. The approach was further developed in [2] where the authors propose a method for learning joint reward policy options with adversarial methods based on Inverse Reinforcement Learning (IRL). The method can learn underlying reward functions and policies solely from human video demonstrations.

Tasks

- investigate different methods for adversarial imitation learning, in particular IfO methods;
- implement at least 2 different models;
- propose improvements;

- perform evaluations on low dimensional control tasks (cartpole, mountain car) and in the MuJoCo simulated environment [3].

[1] Jonathan Ho and Stefano Ermon. Generative adversarial imitation learning. In NIPS, pages 4565–4573, 2016.

[2] Peter Henderson, Wei-Di Chang, Pierre-Luc Bacon, David Meger, Joelle Pineau, and Doina Precup. Optiongan: Learning joint reward-policy options using generative adversarial inverse reinforcement learning. In Association for the Advancement of Artificial Intelligence, 2018

[3] http://www.mujoco.org/index.html