FACULTATEA DE AUTOMATICĂ ȘI CALCULATOARE

AI-MAS Group
"AIM AS high as you can"

Research Topics (2020 – 2021)

Mihai Nan (mihai.nan@upb.ro), Mihai Trăscău (mihai.trascau@upb.ro)

---

1 Explainable deep learning for video recognition tasks

Coordinators

Mihai Nan (mihai.nan@upb.ro), Mihai Trăscău (mihai.trascau@upb.ro)

Description

Most state-of-the art solutions [1, 7, 4, 8] that have been proposed so far for the chosen tasks are black-box approaches. These are based on a neural network that predicts an unintentional result, and it is very difficult to find explanations for the critical cases where the predicted results are wrong. A deep neural network that learns millions of parameters and may be regularised by techniques like batch normalisation and dropouts is quite incomprehensible. Given that the problem of recognising human actions is a problem of classifying a temporal sequence, classical solutions that use deep learning are made up of two parts: feature extraction – transforming the input feature space into a different representation; and optimisation – searching for a decision boundary to separate the classes in that representation space. The goal in this context is to extend these approaches by adding an additional step in which the neural network learns to explain the predicted outcome by providing arguments to validate the network's decision. The reasons why it would be important to add a capability of explanation to a model that recognises human action are not limited to user rights and acceptance. This explicability property is also fundamental for people who design, implement and test the system to enhance system robustness and enable diagnostics to prevent bias, unfairness and discrimination, as well as to increase trust by all users in ***why*** and ***how*** decisions are made.

References

[1] Joao Carreira and Andrew Zisserman. "Quo vadis, action recognition? a new model and the kinetics dataset". In: *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 6299–6308.

[2] David Gunning. "Explainable artificial intelligence (xai)". In: *Defense Advanced Research Projects Agency (DARPA), nd Web* 2 (2017), p. 2.

[3] Tae Soo Kim and Austin Reiter. "Interpretable 3d human action analysis with temporal convolutional networks". In: *2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW)*. IEEE. 2017, pp. 1623–1631.

[4] Feng Mao et al. "Hierarchical video frame sequence representation with deep convolutional graph network". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, pp. 0–0.

[5] Chiradeep Roy et al. "Explainable Activity Recognition in Videos." In: *IUI Workshops*. 2019.

[6] Wojciech Samek, Thomas Wiegand, and Klaus-Robert Müller. "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models". In: *arXiv preprint arXiv:1708.08296* (2017).

[7] Xianyuan Wang et al. "I3d-lstm: A new model for human action recognition". In: *IOP Conference Series: Materials Science and Engineering*. Vol. 569. 3. IOP Publishing. 2019, p. 032035.

[8] Chao-Yuan Wu et al. "A Multigrid Method for Efficiently Training Video Models". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 153–162.

2 Temporal convolutional neural networks for sequence modelling

Coordinators

Mihai Nan (mihai.nan@upb.ro), Mihai Trăscău (mihai.trascau@upb.ro)

Description

Sequence-to-sequence learning is about training models to convert sequences from one domain to sequences in another domain. There are multiple ways to handle this task, either using RNNs or using TCNs with 1D convnets and this project will focus on the TCN-based approach. In general, input sequences and output sequences have different lengths and the entire input sequence is required in order to start predicting the target. But not all the information that makes up this sequence is equally important, we want to extract information taking into account the importance of each component. In the original TCN paper [1], the authors conduct a systematic evaluation of generic convolutional and recurrent networks for sequence modeling. Unlike in RNNs where the predictions for later time-steps must wait for their predecessors to complete, convolutions can be done in parallel since the same filter is used in each layer. Therefore, in both training and evaluation, a long input sequence can be processed as a whole in TCN, instead of sequentially as in RNN. A TCN can change its receptive field size in multiple ways. For instance, stacking more dilated (causal) convolutional layers, using larger dilation factors, or increasing the filter size are all viable options (with possibly different interpretations). TCNs thus afford better control of the model's memory size, and are easy to adapt to different domains. The purpose of this project is to analyze the results that can be obtained by an architecture that contains TCN type layers for a sequence modelling problem, by highlighting all the characteristics listed above.

References

[1] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling". In: *arXiv preprint arXiv:1803.01271* (2018).

[2] Colin Lea et al. "Temporal convolutional networks: A unified approach to action segmentation". In: *European Conference on Computer Vision*. Springer. 2016, pp. 47–54.

[3] Alan J Lockett and Risto Miikkulainen. "Temporal convolution machines for sequence learning". In: *Dept. Comput. Sci., Univ. Texas, Austin, Tech. Rep. AI-09-04* (2009).

[4] Annapurna Sharma and Dinesh Babu Jayagopi. "Towards efficient unconstrained handwriting recognition using Dilated Temporal Convolution Network". In: *Expert Systems with Applications* (2020), p. 114004.