

Advanced Signal Processing Seminar WS2018/19

Resource-efficient Neural Networks

First Meeting: 8th October 2018, 4pm,
Seminarroom Inffeldgasse 16b, IST (IC02062)

Instructor: Franz Pernkopf

Regular Meeting: 16:00-18.00, Monday (starting in November)

Where: **Inffeldgasse 16c, ground floor (IDEG 134)**

Language: English

Abstract:

While machine learning is traditionally a resource intensive task, embedded systems, autonomous navigation and the vision of the Internet-of-Things fuel the interest in resource efficient approaches.

These approaches require a carefully chosen trade-off between performance and resource consumption in terms of computation and energy. On top of this, it is crucial to treat uncertainty in a consistent manner in all but the simplest applications of machine learning systems.

In particular, a desideratum for any real-world system is to be robust in the presence of outliers and corrupted data, as well as being "aware" of its limits, i.e. the system should maintain and provide an uncertainty estimate over its own predictions. These complex demands are among the major challenges in current machine learning research and key to ensure a smooth transition of machine learning technology into every day's applications.

In this seminar we aim to discuss recent advances for reducing the computational burden of deep neural networks, while maintaining the level of recognition performance. This enables to use deep models in mobile devices and embedded systems with limited power-consumption and computational resources.

Organization:

Each student selects a topic (in the topic area, some suggestions of topics/papers are below). He/she should give an in-depth presentation of this topic (45 minutes per student). In particular, the methods used in the papers should be presented in depth.

Step 0: First Meeting, 8th October, discussion of the organization

Step 1: Assignment of topics. Each student should suggest a topic/papers in the area of the seminar.

Some suggestions are below in the tutorial paper. The topic is fixed with me.

Step 2: The student should look for relevant literature in the assigned topic. The student prepares a concept for the presentation. This concept + relevant literature is discussed in a meeting with me (2 weeks before the presentation). In this meeting we select the most interesting work for the presentation.

Step 3: The main (1-3) papers are sent by the presenting student to all students. Each student should read these papers. So that everybody can participate in the discussion.

Step 4: Prepare presentation and present at a scheduled date.

Grading: Grades are given based on the presentation and the participation in discussions (50% - 50%). The presentation slides have to be sent to pernkopf@tugraz.at.

Suggested Topics – see overview Paper:

F. Pernkopf, W. Roth, M. Zöhrer, L. Pfeifenberger, G. Schindler, H. Fröning, S. Tschatschek, R. Peharz, M. Mattina, Z. Ghahramani, "Efficient and Robust Machine Learning for Real-World Systems", Technical Report, 2018