Tutorial 11: Challenge (40P)

In this tutorial we present you the *EIHW Deep Learning Challenge* – a common practice in the machine learning community. Typically, several groups compete against each other to achieve the highest score (e.g., accuracy) on the same data set. The task for this challenge is image-based environment classification.
For our Challenge the results will be presented in the final tutorial on 05/02/2020 (Bonus points available for the presentation).
The first place team will win a small price!
Note: In many cases the test labels are not provided with the challenge data. However, as the dataset we will be working on is publicly available, you will be provided with test labels.

Please submit your code, results, and your presentation by 03/02/2020 23:59 to manuel.milling@informatik.uni-augsburg.de and thomas.wiest@informatik.uni-augsburg.de. You can also download the files from the following link: `https://megastore.uni-augsburg.de/get/GHws5B0Kt4/`.
We recommend you to work in teams of 2 (please indicate all names with the submission), as it is usually an advantage to explore multiple ideas/architectures.

Data Set

The dataset contains roughly 25 000 images of six different scenes, divided into Train, Prediction and Test partition. The labels of the data set have to be created based on the given folder structure. The validation/development set can be a subset of the train set. `https://www.kaggle.com/puneet6060/intel-image-classification`

1 Task and Rules (30P)

The task is to train a classifier on the data that achieves the highest possible accuracy (i.e., number of correctly predicted samples from the test set divided by the total number of samples in the test set, micro accuracy) on the test set. Any deep learning approach is allowed including transfer learning, data augmentation, etc. Use a random seed of 42 to make your results reproducable. Report your accuracy with 3 digits (e.g., 90.4%) on training development and test set.
Important: Only evaluate the final configuration of your network on the test data after training it on the train data and evaluating it (and its hyperparameters) on the validation data. Do NOT improve your network on the test data. Do NOT alter the test set in any way (e.g., upsampling).
Note: Usually it is preferable to run the training (experiment) several times and average the results. However, in order to be less resource-exhausting, you have to run the training only once.

2 Bonus: Beat the Baseline (10P)

Even though there is no specific baseline (results of a rather simple approach which are supposed to be outperformed) given with the database, we set the baseline based on online approaches to 90% accuracy.

3 Presentation (10P)

Present and overview of your methodology, results, and possible challenges and further steps in a 5-min presentation. Submit the slides together with your code and your results.
Note: An interesting result of such a classification problem is the confusion matrix, which shows likely mistakes between two classes are. This can be a hint to the similarity between classes.