Synopsis for 02456 Project "Various Deep Learning Architectures for Urban Sound Classification"

s161041	Sébastien Demortain	s161027	Péter Semság
s161174	Lorenzo Belgrano	s161463	Benjamin Jüttner

Background and Motivation:

Sound classification is a task commonly solved by RNNs rather than CNNs, which in turn are rather suitable for image data. However, since a spectrogram of an audio sequence can be interpreted as an image, CNNs too can be used for sound data, as was done e.g. in [1]. Therefore, sound data is a good opportunity to work with two of the architectures we learned in 02456, namely CNNs and RNNs. The dataset chosen for the project was the *UrbanSound8K* [2], which is a collection of over 8000, 2 to 7 seconds long, audio clips from urban environments with labels such as dog barking or jackhammer.

Milestones:

1. (also safe plan B) Reproduce the CNN architecture proposed in [1], with each audioclip processed into several 60×41 pixel spectrograms.
2. same architecture and same data as in Milestone 1, but now train the less noisy observations first and the noisier observations afterwards (see *curriculum learning* [3]). See if the performance improves.
3. Implement an architecture combined of CNN and RNN, as done in [4] (will probably involve CTC).
4. Test the networks we have trained, optimzied and evaluted, on mixture audioclips, and see which classes have the strongest representation in the softmax output of the network.

References

[1] K. J. Piczak: "ENVIRONMENTAL SOUND CLASSIFICATION WITH CONVOLUTIONAL NEURAL NETWORKS", in *2015 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING*, Sept. 17–20, 2015, Boston, USA

[2] J. Salamon, C. Jacoby, and J. P. Bello, "A dataset and taxonomy for urban sound research", in *Proceedings of the ACM International Conference on Multimedia*. ACM, 2014, pp. 1041–1044.

[3] Y. Bengio, J. Louradour, R. Collobert, J. Weston: "Curriculum Learning"

[4] Baidu Research – Silicon Valley AI Lab: "Deep Speech 2: End-to-End Speech Recognition in English and Mandarin"