References: Deep Neural Network ASR

Deep Speech 2

The following presentation, slides, and paper from Baidu on *DeepSpeech 2* were important resources for the development of this course and its capstone project:

- Amodei, Dario, et al. "Deep speech 2: End-to-end speech recognition in english and mandarin." International Conference on Machine Learning. 2016.
- Presentation
- Slides

Language modeling with CTC

Gram-CTC from Baidu on integrating a language model into CTC for better performance:

- Liu, Hairong, et al. "Gram-CTC: Automatic Unit Selection and Target Decomposition for Sequence Labelling." arXiv preprint arXiv:1703.00096 (2017).

Language modeling with CTC based on weighted finite-state transducers (WFSTs):

- Miao, Yajie, Mohammad Gowayyed, and Florian Metze. "EESEN: End-to-end speech recognition using deep RNN models and WFST-based decoding." Automatic Speech Recognition and Understanding (ASRU), 2015 IEEE Workshop on. IEEE, 2015.
- Slides