BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Master's Thesis Assignment

164401

Institut:	Department of Computer Graphics and Multimedia (DCGM)
Student:	**Horník Matej, Bc.**
Programme:	Information Technology and Artificial Intelligence
Specialization:	Machine Learning
Title:	**Effective Training of Neural Networks for Automatic Speech Recognition**
Category:	Speech and Natural Language Processing
Academic year:	2024/25

Assignment:

1. Familiarize yourself with the fundamentals of machine learning and automatic speech recognition.
2. Train an Encoder-Decoder model for speech recognition using one of the freely available datasets (e.g., VoxPopuli).
3. Utilize pre-trained acoustic models (e.g., XLS-R) and language models (e.g., GPT-2) to initialize the network and compare whether it is advantageous to initialize the model from pre-trained models.
4. Determine which network parameters are suitable for updating and which can be left unchanged without significant loss in recognition accuracy.
5. Analyze whether the models converge faster if individual components of the model (the acoustic and language model) are first fine-tuned on the given corpus and only then combined.
6. Publish at least one of the trained models.

Literature:

- Baevski, A., Zhou, Y., Mohamed, A. a Auli, M. Wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. In: Advances in Neural Information Processing Systems. Curran Associates, Inc., 2020, sv. 33, s. 12449–12460.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. et al. Language models are unsupervised multitask learners. In: OpenAI blog. 2019, sv. 1, č. 8, s. 9.

Requirements for the semestral defence:
points 1 and 2 completed, point 3 in progress

Detailed formal requirements can be found at https://www.fit.vut.cz/study/theses/

Supervisor:	**Polok Alexander, Ing.**
Head of Department:	Černocký Jan, prof. Dr. Ing.
Beginning of work:	1.11.2024
Submission deadline:	21.5.2025
Approval date:	12.11.2024