Sign Language Interpreter

Generalization of a pre-trained TwoStream Model

Graduate

René Mair

Initial Situation: The development of automatic sign language systems requires methods from computer vision, machine learning, and linguistics. The aim of this work was to deploy a pre-trained TwoStream Sign Language Recognition and Translation (SLRT) model from an open-source repository and adapt it for gloss recognition. The focus was not on translating into continuous text, but on the robust recognition of individual sign units.

Approach / Technology: The RWTH-PHOENIX-Weather-2014T dataset served as the primary data source. While it contains both video data and text translations, only still image sequences and gloss annotations were used in this work. As a result, steps such as keypoint extraction or motion normalization were not required. Compared to the SIGNUM dataset, only minor adjustments were necessary, as both formats are gloss-based.
The TwoStream model combines an RGB stream and a keypoint stream; in this work, only gloss recognition was prepared and methodically implemented. The extracted features were not to be translated into text but directly compared to reference glosses. Evaluation was planned to use the Word Error Rate (WER).

Result: The evaluation showed that the model achieved solid results on PHOENIX data and, after a SIGNUM integration, also delivered reliable performance. This allowed a first statement on generalizability: it is in principle achievable but requires careful preprocessing and gloss harmonization. As a next step, extending the system to continuous text translations is recommended to additionally evaluate translation quality with the BLEU metric and thus assess generalizability in a more systematic way.
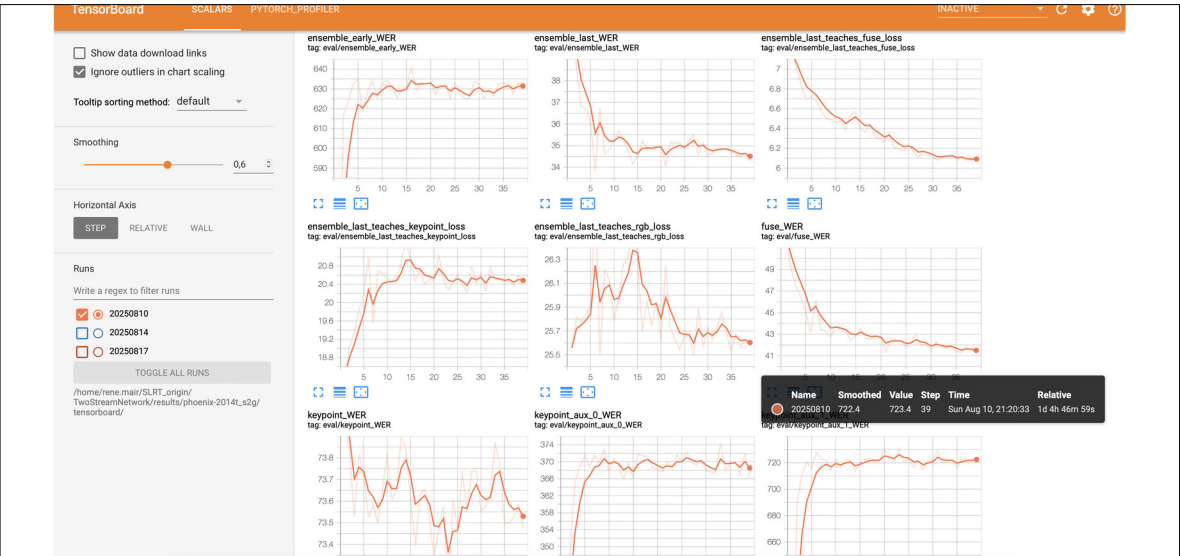
TwoStream Model Architecture
https://doi.org/10.1109/ACCESS.2020.3025165
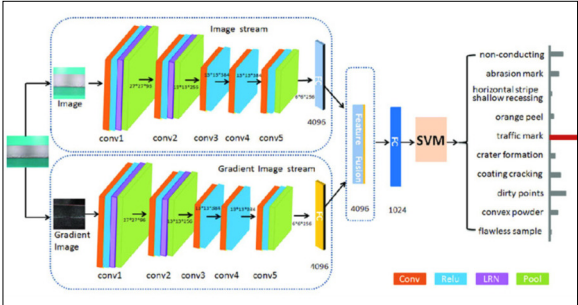
TwoStream SLR Evaluation
Own presentment
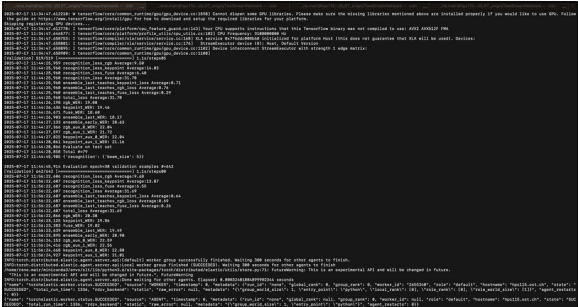
TwoStream SLR WER
Own presentment

Advisor
Prof. Dr. Carlo Bach

Co-Examiner
Prof. Dr. Christoph Würsch

Subject Area
Computer Science