

First Truly End-to-End SLT

Bérard et al. (2018) presents the first truly end-to-end SLT:

- Speech encoder:
 - 2 layers converting n -dim input into n' -dim.
 - 2 layers of convolution
 - 3-layer bidirectional LSTM
- Attention
- Char-level decoder
 - Used either to predict English transcription ($|V| = 46$),
 - or French translation ($|V| = 167$)