Experimental Machine Translation of the Swiss German Sign Language via 3D augmentation of body keypoints

Lorenz Hufe, Eleftherios Avramidis
German Research Center for Artificial Intelligence (DFKI), Berlin, Germany

Introduction

Challenges for MT of signed languages:
(a) **multimodal and multilateral** nature
-> different ML architectures than classical text-to-text MT
(b) **lack of data**
-> hard for end-to-end deep learning
- experiments in few SLs and domains
- open questions for generalization

WMT SLT 2022:
- new language pair (DSGS-German), data sources: former Deaf TV channel, SRF
- dataset based on originally signed content (not interpretation-based)

Motivation

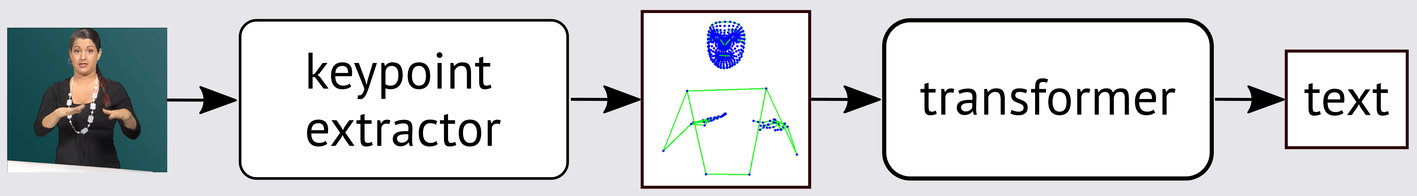
Keypoint extraction:
- lack of data: employ external knowledge from computer vision models
- possibility of utilizing additional data thanks to anonymize of speakers

Geometric transformation:
- models should be robust to speakers seen from different angles
- more valid training instances
- avoid spurious feature correlation during training

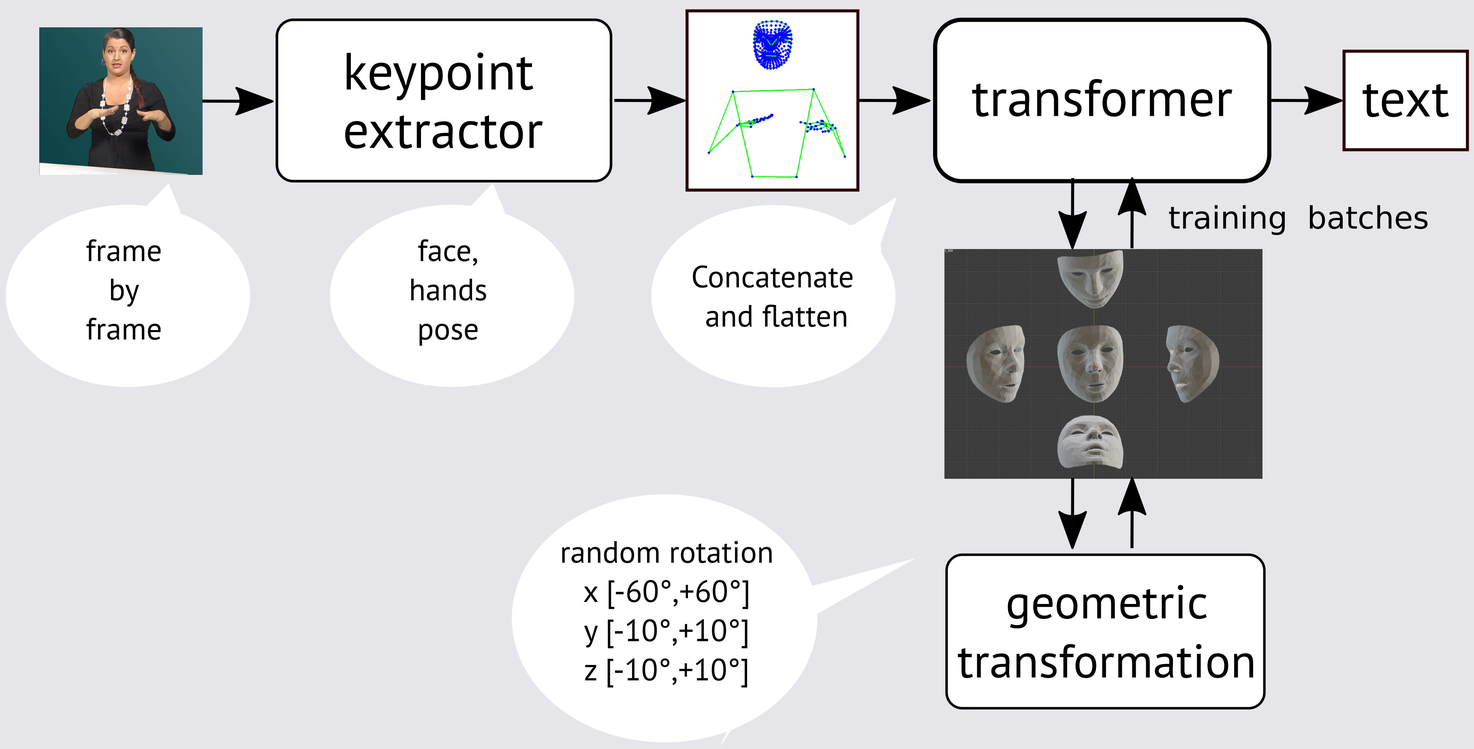
Method: Body Keypoints + 3D augmentation

Classical text-to-text translation

SL translation via keypoints extraction



SL translation via keypoints extraction and 3D transformation



Experiment setup

training dataset: FocusNews (10,000 sent.)
pre-extracted keypoints: Mediapipe holistic
features: face, hands, pose (708)
3D transformations: NumPy arrays
MT framework: JoeyNMT
validation metrics: BLEU/chrF (SacreBLEU)

parameter	value
feature size	708
max sentence len.	400
dropout	0,1
FF size	2048
heads	8
embeddings dim.	512
hidden size	512
optimizer	adam
batch size	32
random seed	42
weight decay	0,001
learning rate	0,001
validation freq.	100
beam size	1
beam alpha	-1
translation max len.	30

Results

max rotation +/- (°) x	y	z	LR scheduler patience	metric	layers	scores BLEU-3	BLEU-4	ChrF	runtime (h)
10	10	10	25	BLEU	4	0,28	0,00	15,36	00:21
10	10	10	50	BLEU	4	0,28	0,00	15,36	00:31
60	10	10	50	BLEU	4	0,00	0,00	17,58	00:24
60	10	10	500	ChrF	3	0,310	0,00	16,08	07:44
▶ 60	10	10	500	ChrF	4	0,314	0,00	16,43	04:14

- **very low scores:** repetitive sentences and irrelevant translations, hard to draw conclusions
- **limited time and resources:** not possible to experiment with all possible combinations
- **validation metrics:** zero BLEU-4 could not be used, switched to chrF

Conclusions and Further work

- **dataset difficulty**: bad performance by all systems, better results using external SL data
- **imprecision of keypoints**: lack of details needed for SL, error propagation

Possible next steps:
- **ablation study** in other datasets - comparison with state of the art experiments
- **better keypoint extraction** taking the frame sequences into consideration
- **better combination** of keypoints with frame embeddings and training process
- **more data**: dataset collection, data augmentation techniques,

This work has been supported by the German Federal Ministry of Education and Research (BMBF) through the project SocialWear (01IW2000)