
SIGN2VOICE

Our team

Maximilian Scheel

Benedikt Gramß

Felix Sattler

Vanessa Türker

Oliver Pürner

SIGN2VOICE

Lost in translation

ca. 80,000

< 1,000

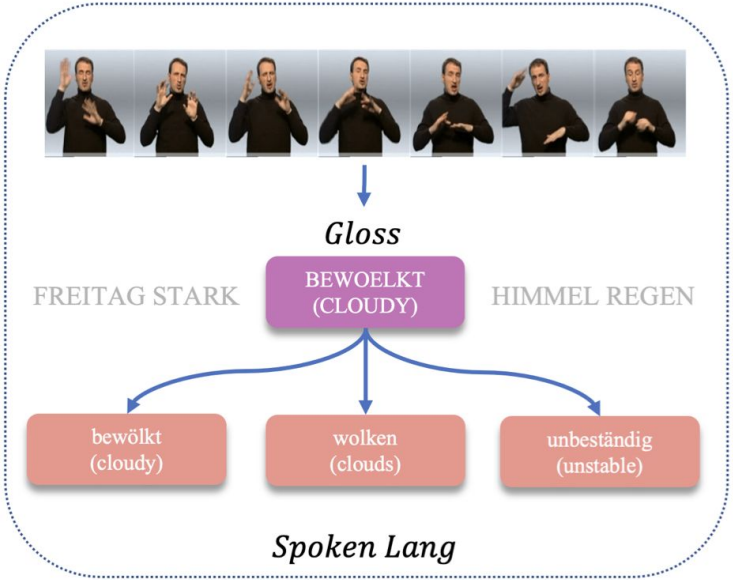
SIGN2VOICE

Giving the unheard a voice

*With **Sign2Voice** we offer **sign language video to audio translation** to to enable everyone to interact with sign language*

SIGN2VOICE

What the hand ...

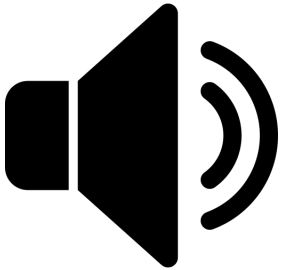


SIGN2VOICE

… and the complexity of sign language processing

=

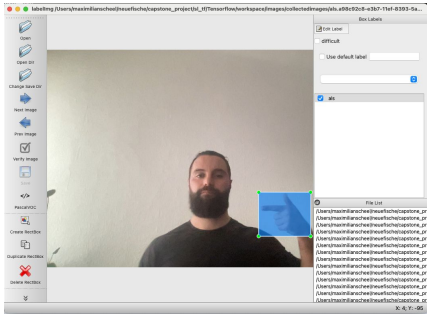
=

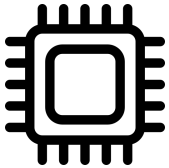


SIGN2VOICE

SLR - identifying glosses

tensorflow





create

train

test

SIGN2VOICE

SLR - identifying glosses

CorrNet

encoder

input

context vector

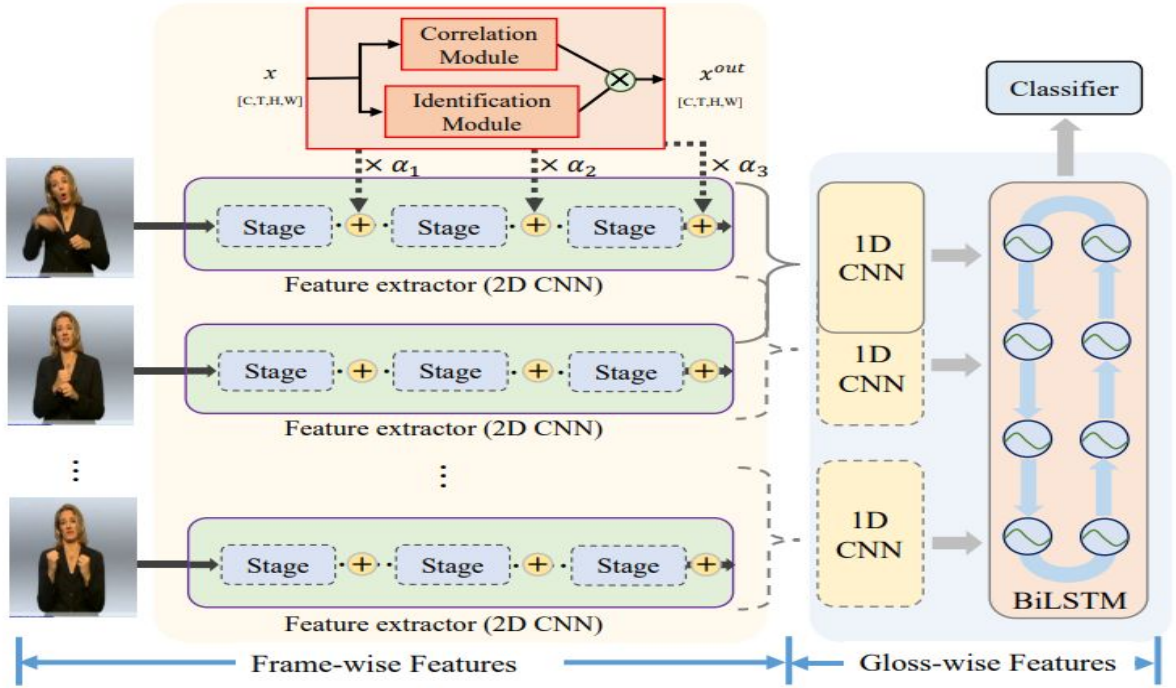
decoder

output

SIGN2VOICE

Transformer models - CorrNet

An overview for our CorrNet. It first employs a feature extractor (2D CNN) to capture frame-wise features, and then adopts a 1D CNN and a BiLSTM to perform short-term and longterm temporal modeling, respectively, followed by a classifier to predict sentences. We place our proposed identification module and correlation module after each stage of the feature extractor to identify body trajectories across adjacent frames (Hu, et al. 2023).

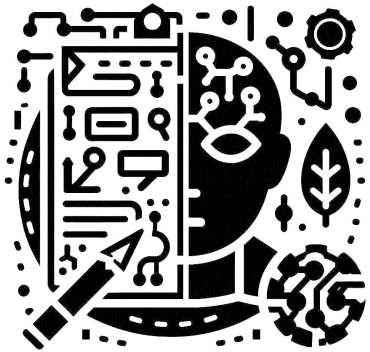


SIGN2VOICE

G2T - making sense of glosses

Gloss2Text

Multilingual LLM



Paraphrasing & back translation

Semantic relationships

SIGN2VOICE

Sign Language Translation (G2T)

Gloss2Text

To improve the translation of glosses into text our model uses

- **NNLB-200 - a Large Language Model (LLM)** pre-trained on expansive and diverse corpora
- **tailored data augmentation techniques** like paraphrasing & back translation
- fine-tuning through **semantically aware label smoothing techniques (SALS)** on the target dataset, based on the semantic relationships between the classes

SIGN2VOICE

TTS - turn text into audio

G2T Output

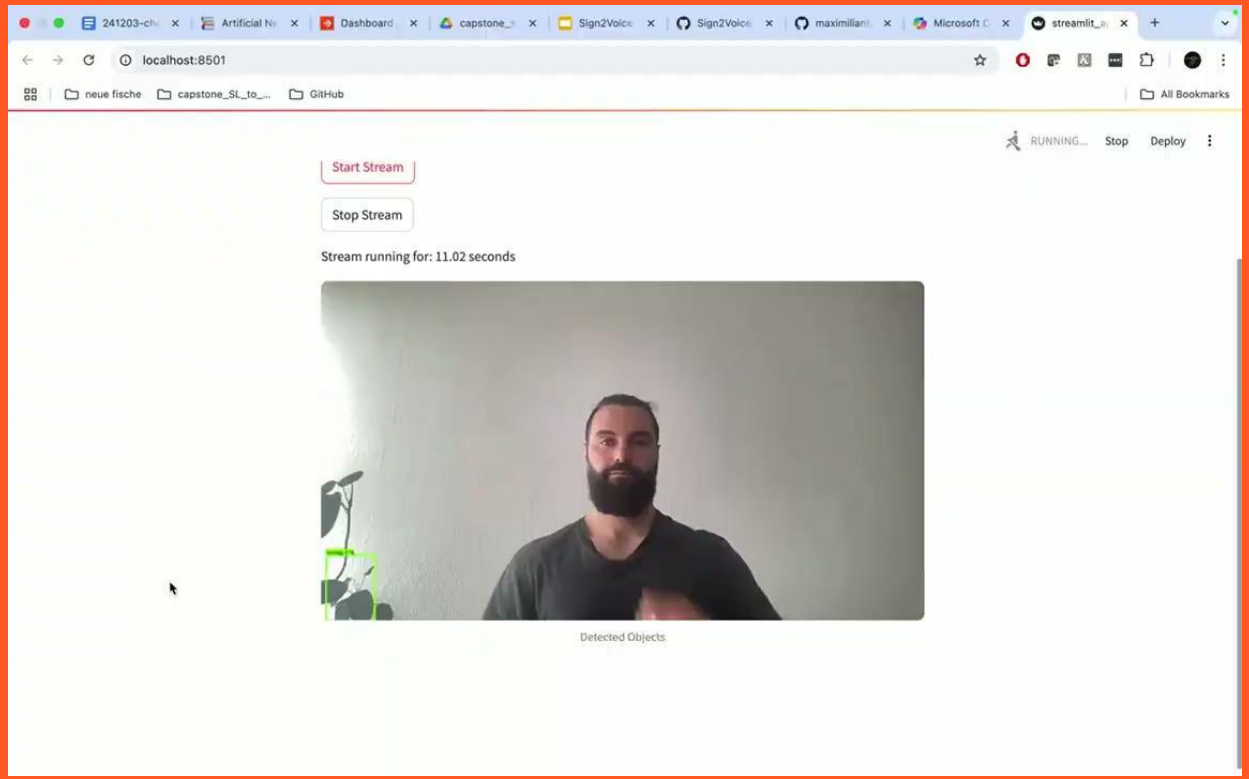
Audio API TTS Model

PyAudio Stream

SIGN2VOICE

TIME FOR A DEMO

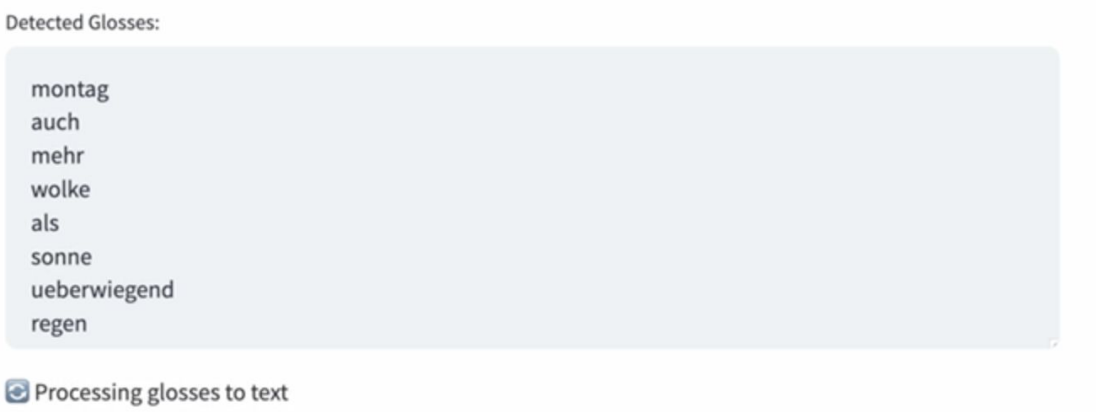
tensorflow


localhost:8501
neue fische
capstone_SL_to_...
GitHub
All Bookmarks
RUNNING...
Stop
Deploy
Start Stream
Stop Stream
Stream running for: 11.02 seconds
Detected Objects

tensorflow - WER at confidence 0.65

1 of 9 = ca. 11 % error rate

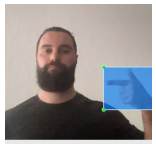
montag	auch	mehr	wolke	als	sonne	ueberwiegend	regen	gewitter



SIGN2VOICE

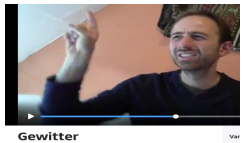
tensorflow - limitations

1. spatial context
   (body, facial expression)



2. timing context

3. gloss differentiation



3. training data quality & generalisability
   (hardware setup, variety of signers, ...)

SIGN2VOICE

Sign2Voice 🗣️

📂 Upload Frames

Upload your frames here (only images)

How is the weather?

WER
(WORD ERROR RATE)

29.37%

The WER measures how different the predicted glosses are from the actual glosses in the Ground Truth

- A WER of **0%** would mean a perfect translation
- A **high WER** means the model makes many mistakes.

SIGN2VOICE

How to improve sign2voice

1. **ACCURACY** — Improve the model pipeline
2. **SPEED** — Increase the processing speed
3. **USABILITY** — Add additional app features

SIGN2VOICE

References

- CORRNET:
  - CorrNet GitHub Repo
- GLOSS2TEXT:
  - Gloss2Text GitHub Repo
- TTS:
  - OpenAI TTS Documentation
  - ReallyEasyAI GitHub Repo

SIGN2VOICE