Effect of Data Augmentation with Multi-View Perspectives of Signers on the DGS-Fabeln-1 Dataset

Fabian Renner, Daksitha Withanage Don, Elisabeth André, Cristina Luna-Jiménez

Chair for Human-Centered Artificial Intelligence, University of Augsburg, Augsburg, Germany

Abstract

- **Gap: Sign Language Translation paradigm still requires further research to overcome current limitations.** The SLT research community is driving their efforts to investigate **Continuous Sign Language Translation** (CSLT). This new paradigm **is more natural to the translation task,** receiving videos in Sign Language and translating them into Spoken Languages.

Our Contributions:

1) Given the success of transfer learning in multiple tasks, we investigated the effect of the transferability of knowledge in the UniSign transformer architecture in DGS-Fabeln-1.

2) As training large models with more samples usually results in higher performance, we compared the effect of including only frontal camera perspectives versus 7 perspectives.

Related Work

German Sign Language Datasets:

- **DGS Public Corpus [**Konrad et al. (2022)**]:** One of the largest covering **everyday dialogues**. **50 hours** of video material, **annotated using their double glossing scheme**. Only two video perspectives.
- **DGS-Fabeln-1 [**Nunnari et al. (2024)]: German text and videos in DGS. Material recorded from **different camera perspectives**. Topic: German fairy tales interpreted by a deaf DGS signer.

Transformer Architectures for SLT:

- **UniSign (Li et al.,2024)** is a **Transformer-based** architecture designed for sign language understanding (SLU), which **unifies** within a single framework: **isolated sign language recognition (ISLR) and continuous sign language recognition (CSLR).**

Methodology

Dataset: DGS-Fabeln-1 [Nunnari et al., 2024]

Dataset Configuration:

1) **Only Front-view camera** perspectives of **6 fairy tales** (No "Snow White"). Random sample splits in 80/10/10 Train/validation/test → 338 videos for training, 42 validation and **43 test (only frontal)**.

1) **Only Front-view camera** perspectives **of 7 fairy tales** → 476 videos for the training set, 52 videos in the validation set and **43 in the test set (only frontal).**

1) **All perspectives of the 7 fairy tales** → 3,749 videos (all perspectives) in train, 409 videos (all perspectives) in the validation split and **43 in the test set (only frontal).**

Model: Uni-Sign [Li et al., 2024]

- ✓ Pre-trained on other SL datasets. (SoTa)
- ✓ RGB vs. Pose pipelines

Results

Tales	BLEU-1	BLEU-4	ROUGE
6	13.66	0.97	13.73
7	**18.17**	**2.16**	**15.80**

Effect of Dataset Size: Despite the "originality/creativity" of the domain, adding more fairy tales to the training set results in **higher performance** of the models for predicting next sentences in other fairy tales.

Training	BLEU-1	BLEU-4	ROUGE
Scratch	18.17	2.16	15.80
Transfer-Learning	17.88	**4.15**	**16.09**

Effect of Transfer Learning: Training times were **similar** and **pretrained weights delivered superior performance across the majority of experiments**, with a particularly strong advantage for the Pose + RGB input type.

Batch	Camera Perspectives	BLEU-1	BLEU-4	ROUGE
2	Frontal	17.88	4.15	16.09
2	All	17.88	4.20	**16.12**
4	Frontal	17.88	4.15	16.09
4	All	17.88	4.20	16.12
8	**Frontal**	**18.32**	2.15	11.47
8	**All**	15.21	**4.21**	11.43

- **Data Augmentation with additional Camera Perspectives:** For the experiments with the **Pose + RGB input type, with more perspectives higher BLUE-4 (but not across all metrics) and higher training times.**

Conclusions & Future Work

Key Takeaways

- ✓ Adding **more fairy tales** to the training helped to predict **better** fragments of other fairy tales.
- ✓ Employing **pre-trained weights** of the multi-language model **improved** the final performance of the model.
- ✓ Employing **all the camera perspectives** resulted in an increment in the BLUE-4, however these additional number of training samples resulted in longer training times. **(More experiments are required).**

Future Work

- Further **optimization** steps can be taken by introducing and varying other parameters and more advanced LLMs**.**
- **Extend study to only Pose pipelines and datasets with several camera perspectives.**

References

- [Konrad et al. (2022)] Reiner Konrad, Thomas Hanke, Gabriele Langer, Susanne König, Lutz König, Rie Nishio, and Anja Regen. 2022. Public DGS Corpus: Annotation Conventions / Öffentliches DGS-Korpus: Annotationskonventionen.
- [Li et al., 2024] Zecheng Li, Wengang Zhou, Weichao Zhao,Kepeng Wu, Hezhen Hu, and Houqiang Li. 2024.Uni-Sign: Toward Unified Sign Language Under-standing at Scale. In The Thirteenth InternationalConference on Learning Representations, Sin-gapore, Singapore. URL: https://iclr.cc/virtual/2025/poster/31250.
- [Nunnari et al., 2024] Fabrizio Nunnari, Eleftherios Avramidis, Cristina España-Bonet, Marco González, Anna Hennes, and Patrick Gebhard. 2024. DGS-Fabeln-1: A Multi-Angle Parallel Corpus of Fairy Tales between German Sign Language and German Text. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), pages 4847–4857, Torino, Italia. ELRA and ICCL. URL: https://aclanthology.org/2024.lrec-main.434/.

Acknowledgements

This work was funded by the **BIGEKO** project (BMBF, German Ministry for Education and Research, grant number 16SV9094), and **FORSocialRobots** project (BFS, Bavarian Research Foundation, grant number AZ1594-23).