ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

Trends and Challenges for Sign Language Recognition with Machine Learning

Jérôme Fink[1], Mathieu De Coster[2], Joni Dambre[2], Benoît Frénay[1] *

1- University of Namur - NaDI - Faculty of Computer Science - PReCISE
Rue Grandgagnage 21, 5000 Namur - Belgium

2- IDLab-AIRO – Ghent University – imec
Technologiepark-Zwijnaarde 126, 9052 Ghent - Belgium

Abstract. Research in natural language processing has led to the creation of powerful tools for individuals, companies... However, these successes for written languages have not yet affected signed languages (SLs) to the same extent. The creation of similar tools for signed languages would benefit deaf, hard of hearing, *and* hearing people by making SL content, learning, and communication more accessible for everyone. SL recognition and translation are related to AI, but require collaboration with linguists and stakeholders. This paper describes related challenges from an AI researcher's point of view and summarizes the state of the art in these domains.

1 Introduction

Signed languages (SLs) are the primary means of communication for many deaf and hard-of-hearing (DHH) people. They are natural languages that evolve in SL communities. Thus, there is no such thing as a unified SL. Instead, there are multiple languages and dialects, just like spoken languages.

In the past few years, we have seen extraordinary progress in natural language processing (NLP). Nowadays, it is easy to find automatic translators for many spoken languages, such as DeepL[1] and Google Translate[2]. It has become common practice to rely on grammar or spell checkers when writing. More recently, so-called large language models paved the way to even more powerful tools, the most well-known of which is probably ChatGPT[3]. In contrast, SL recognition (SLR) and translation (SLT) are still in their infancy and tools such as DeepL, let alone ChatGPT, do not exist for SLs. This work explains the specificities of SLs to a technical audience willing to tackle the problem of SLR or SLT. It also provides some necessary historical and ethical context.

From a technological perspective, applying the recent advances in artificial intelligence (AI) to SLs would allow the creation of several applications for SL users. Perhaps the "holy grail" of sign language technology, is a sign language

*Mathieu De Coster's research is funded by the Research Foundation Flanders (FWO Vlaanderen): file number 77410. This work has been conducted within the SignON project. This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 101017255.

[1] `https://www.deepl.com/en/translator`

[2] `https://translate.google.com/`

[3] `https://chat.openai.com/`

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

translation app that could be used to facilitate communication between hearing and DHH people. Other possible applications of such technology include learning tools or online SL dictionaries.

There is a historical issue of hearing people making decisions on behalf of DHH people, related to the concept of "audism" [1]. Instead, it is more ethical to communicate with SL users and ask them what technologies they want or not. For instance, there is some resistance against SL avatars [2, 3, 4]. Recent research projects such as SignON [5] and EASIER [6] aim to enable communication between researchers in AI, who are often hearing, SL linguists, and the end users of potential SL applications. They present a "co-creation" pipeline, that aims to continuously involve DHH and hearing end users in the design and development of technology. As such, the hope is to converge on useful and desired applications.

This work aims to be a good starting point for new researchers in the field. Section 2 presents a high-level overview of available datasets and their properties. Next, Section 3 shows the state of the art in SLR and SLT. The papers of this special session are listed in Section 4. Finally, Section 5 offers a conclusion that summarizes the remaining challenges and gives perspective on future work.

2 Sign Language Data

The advances in NLP are powered by DL and large amounts of data. Lack of data is the main reason why SLR and SLT do not witness the same progress as speech and text. While there are large amounts of SL data available, they originate from various sources (Section 2.1) and often are only partly annotated, with annotation conventions differing (Section 2.2). Challenges related to the collection, organization, and annotation of these data are major roadblocks on the way toward SL applications [7].

2.1 Data Sources

SL data can be found "in the wild", on social media and video-sharing platforms where SL communities gather to share vocabularies or vlogs [8, 9]. Other sources from which existing SL data can be collected are television broadcasts [10, 11].

There has also been a multitude of targeted SL data collection efforts, the first of which was by SL linguists who wanted to study specific SLs [12, 13, 14]. The success of DL in related fields has also led to the collection of SL datasets specifically for use in DL [15, 16].

The data's source has an influence on the signing quality. For example, news broadcasts are typically interpreted directly from auto-cue or speech, leading to information loss or non-spontaneous signing. Some datasets were collected by instructing people to record specific signs. Signers tend to sign less naturally when instructed to sign specific sentences. The data source could also introduce representative bias if the number of signers involved in the dataset is too low but also vocabulary bias depending on the task performed by the signers. For instance, the PHOENIX-weather dataset is based on German weather forecasts,

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

Name	Data source	Annotations	Type
ASL [17]	SL lexicon	G	I
ASL-LEX [18]	SL lexicon	P	I
BOBSL [19]	TV broadcast	G, S	I, C
BeCoS [11]	TV broadcast	S	C
LSFB [20, 21]	SL corpus	G, S	I, C
MS-ASL [9]	YouTube	G	I
PHOENIX [10]	TV broadcast	S	C
VGT-Corpus [14]	SL corpus	G, S	I, C
NGT-Corpus [13]	SL corpus	G, S	I, C
LSA64 [15]	DL dataset	G	I
Google ASL [16]	DL dataset	G	I

Table 1: This table shows a non-exhaustive list of existing SL datasets, illustrating their variety (both on the data and annotation level). G: Glosses, P: Phonemes, S: Sentences. I: Isolated, C: Continuous.

thus the vocabulary is more specific than other datasets. It is important to choose the correct dataset for the task and to take into consideration the bias of the used dataset [3]. Table 1 provides a list of some existing SL datasets with some of their characteristics.

2.2 Dataset Annotations

SL datasets have either pre-existing annotations or are annotated in varying ways. There is no consensus about how to annotate SL data. Several conventions co-exist [7]. The main ones are:

- **Phonology**, e.g., [18]: Each video is annotated using the smallest component of sign movement called phonemes. However, this kind of annotation is time-consuming and the exhaustive list of SL phonemes is still discussed among SL linguists.
- **Gloss**, e.g., [12, 13, 16]: Every sign is associated with a unique written indicator of its meaning called a gloss. It is the most popular annotation method. Annotation is also time-consuming and it requires domain experts to associate each sign with the correct unique gloss.
- **Utterances**, e.g., [13, 14, 22]: Each segment of the video is associated with its translation into a written language [22]. This form of annotation is perhaps the least time-consuming of the three mentioned here.

Other annotation methods are also investigated to better represent how speech is structured in SL [23].

Regardless of the methods selected, the annotation process is the main bottleneck when creating SL datasets as it is a slow process requiring a highly

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

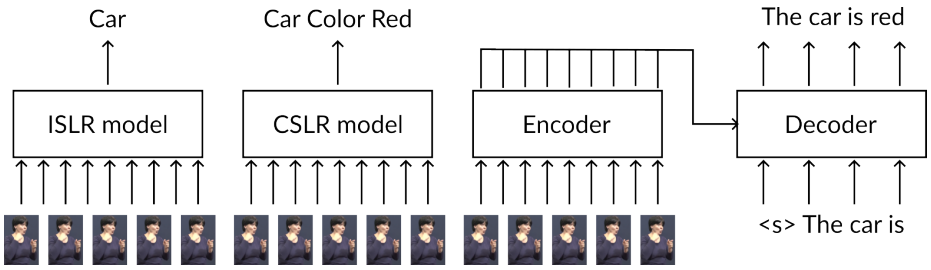

Fig. 1: The ISLR, CSLR, and SLT tasks.

specific skillset. Developing methods for SL recognition could ease this process and create a positive feedback loop.

Despite those standards, every dataset tends to use its own annotation standards. For instance, the NGT corpus [24] contains two kinds of translation annotations: *free* and *narrow*, whereas the VGT corpus only contains a single translation annotation. Despite both corpora being annotated with glosses based on Dutch, conventions on how to annotate concepts like numbers, uncertainties or repetitions differ. Quite some effort can be required to go through the annotation conventions of specific datasets, especially SL corpora, and some linguistic background knowledge may be required to understand these conventions.

3 Sign Language Technology

The umbrella term "SL technology" covers, among others, isolated SLR (ISLR), continuous SLR (CSLR) and SLT. These tasks, illustrated in Figure **??**, each have their own target applications and requirements and they all require processing an SL input signal.

3.1 Input Signal

Signers use visual and spatial information to construct utterances. Thus, video is a natural choice to represent SL data. The signal-to-noise ratio of video-recorded signing is problematic, as these videos contain irrelevant information such as skin color or variable background and lighting. Therefore, pre-processing is often applied to video data to extract relevant features.

An early work extracted 8 features about the hand of the signers in SL videos to train a Hidden Markov Model [25]. The results were encouraging but it was clear that the number of features captured was too low to discriminate an entire SL vocabulary. To ease pre-processing SL data, some researchers make signers wear colored gloves [15] or gloves with sensors [26]. Despite encouraging results, relying on such hardware is unnatural and intrusive for the user [27]. Some works also rely on 3D camera recordings such as Microsoft's Kinect to help track the hands [28] but few devices are equipped with such cameras making the system less accessible.

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

Authors	Year	Vocab.	Signers	Top-1	Method
Albanie et al. [22]	2020	1000	40	65.57%	I3D
Du et al. [37]	2022	1000	40	72.4%	Transformers
Liao et al. [38]	2019	500	8	89.8%	CNN + LSTM
Liao et al. [38]	2019	500	8	89.8%	CNN + LSTM

Table 2: This table summarizes some recent results obtained in ISRL with several architectures on several datasets.

Recently, progress achieved in pose estimation [29, 30] allows to extract landmarks tracking joints and other keypoints of the signers. This facilitate the design of tools leveraging off-the-shelf recording devices to recognize sign languages. However, these pose estimation tools are not always robust and their errors are propagated to SLR models [31, 32].

The processing of SL video data remains an area of active research.

3.2 Isolated SLR

ISLR models aim to classify videos showing a single sign: it is a *many-to-one* classification task. ISLR can be applied to create tools to query dictionaries [33, 34] or help the annotation of datasets [22]. Current ISLR approaches are limited mainly by data and they typically exclude important constructs such as non-lexical signs and spatiality [23].

Next to hand shape, movement plays an important role in signing. Typically, ISLR is tackled by using a spatial (2D) or spatio-temporal (3D) feature extractor (convolutional neural networks or vision transformers) followed by a sequential model such as recurrent neural networks or transformers. The feature extractors are often pre-trained to offset the lack of SL data; for instance, VGG or ResNet pre-trained on ImageNet [35, 36] or I3D pre-trained on Kinetics [22, 20]. Recently, transformers achieved competitive performance for isolated SLR [36]. Table 2 reports results obtained by some recent works.

3.3 Continuous SLR

In CSLR the goal is to predict all the glosses corresponding to a sequence of signs, in the correct order. It is a *many-to-many* classification task. The main additional challenge compared to ISLR is movement epenthesis, a transient movement occurring between two signs and that could easily be mistaken as a sign. The applications of CSLR are automatic dataset annotation [22] and using it as a first step in an SLT pipeline [39].

The models used for CSLR are quite similar to those used for ISLR, but the difference in approach lies in the optimization objective. CSLR is typically tackled with a similar approach to automatic speech recognition, i.e., to use Connectionist Temporal Classification (CTC) [40]. A popular dataset for CSLR is the RWTH-PHOENIX-Weather 2014T dataset [39]. The Word Error Rate

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

Authors	Year	Dataset	WER
Chen et al. [41]	2022	PHOENIX-2014-T	19.3
Zhou et al. [42]	2022	PHOENIX-2014	21.1
Gan et al. [43]	2023	PHOENIX-2014-T	20.1

Table 3: This table summarizes some recent results obtained in CSLR.

(WER) is the most popular metric for this task. Table 3 reports results obtained by some recent works.

3.4 Sign Language Translation

Whereas SLR is concerned with a single signed language, SLT crosses language and modality boundaries. Theoretically, one can translate between any signed and any written language given enough data. This section focuses on translation from signed to written languages, as this task is related to SLR.

The first step in an SLT pipeline is to extract information and/or semantics from video data, i.e., SLR. This information, commonly glosses or embeddings, is then used as input to the actual translation model.

First results for video-based SLT were achieved in 2018 using the RWTH-PHOENIX-Weather 2014T dataset [39]. Five years later, the BLEU-4 scores—quantifying the translation quality—have increased from 9.58 to 25.59. This dataset is rather limited in scope, however. Müller et al. [44] discuss the need for a standardized evaluation method to better track the progress made in the field. RWTH-PHOENIX-Weather 2014T only contains 7096 sentence pairs. Achieving acceptable SLT results on general topics requires larger and more varied datasets. The recently held WMT-SLT22 challenge [45] provides datasets and a standardized evaluation method. Currently, the scores achieved on this benchmark are unsatisfactory.

These low scores are likely due to a lack of training data. In typical translation tasks between pairs of written languages, one would have millions of parallel sentences; in SLT, a more challenging task, we only have access to thousands [46]. Hence, data collection is crucial to improve performance.

The way signers structure their dialog is also a challenge. Signing relies heavily on the usage of the 3D signing space to construct a scene or conceptual objects. Depicting and enacting [23] are also often used instead of lexical signs. Facial expression can also affect the meaning of a sign or sentence. SLT methods should adapt to these particularities. To model those particularities, AZee was proposed as a formal SL grammar [47].

4 Overview of the Papers

This section introduces the five papers accepted in the special session on Machine Learning Applied to Sign Language.

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

Firstly, the paper **Large-scale dataset and benchmarking for hand and face detection focused on sign language** of Leandro et al. introduces a dataset for hand and face segmentation. They evaluate their dataset using various models for image segmentation. This work could help improve current pre-processing pipelines.

Two papers investigate the use of multimodal methods for sign language:

- **Multimodal Recognition of Valence, Arousal, and Dominance via Late-Fusion of Text, Audio and Facial Expressions**: Nunnari et al. explore the usage of multimodal channels such as speech, image, and text to infer the emotion of speech. The detected emotion could be used to generate a realistic avatar in the context of sign language generation.
- **Disambiguating Signs: Deep Learning-based Gloss-level Classification for German Sign Language by Utilizing Mouth Actions** by Nam Pham et al. evaluates the gain in performance of a multimodal architecture taking mouth region information into consideration to predict signs. They find that lip movements convey important information to discriminate similar signs.

Finally, there are two papers on sign language phonology and how to model it in an SLR context:

- **Exploring Strategies for Modeling Sign Language Phonology** by Kezar et al. explores several learning strategies to create models able to better model sign language phonemes.
- **Exploring the Importance of Sign Language Phonology for a Deep Neural Network** by Martinez et al. investigates if deep neural networks trained on sign language data have learned phoneme representation of the language.

5 Conclusion

The field of SLR is still young and constantly evolving. The recent increase in interest gives reason to hope that, soon, powerful tools will appear. However, more high-quality data is needed in order to build a system working correctly when signers sign fluently. The data acquisition and annotation process is the main bottleneck of the field. Developing methods to speed up the annotation process or leverage SL video in the wild is a remaining challenge. Also, the construction of a standard benchmark dataset adopted by the whole research community would be beneficial to accurately track the progress in the field.

Another challenge is to bring useful tools to SL communities. A majority of SLR works focuses on creating algorithms for SLR but does not seek to apply them to solve real-life issues. Such tools must be designed in co-creation with deaf stakeholders. UX researchers must also be involved in this process. The creation and deployment of tools matching the community's needs is a challenge that could have a big impact on society.

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

Finally, current works in SLR and SLT only focus on well-established lexical signs. However, in free-speech settings, signers may choose to use the productive lexicon to describe complex scenes. Such signs are generally created for the situation and imitate the scene that is described. Such constructions constitute an essential part of SL and automated translation of such structures has, to the best of our knowledge, not been tackled yet.

References

[1] H-Dirksen L Bauman. Audism: Exploring the metaphysics of oppression. *Journal of deaf studies and deaf education*, 9(2):239–246, 2004.

[2] Rosalee Wolfe, John C Mcdonald, Eleni Efthimiou, Evita Fotinea, Frankie Picron, Davy Van Landuyt, Tina Sioen, Annelies Braffort, Michael Filhol, Sarah Ebling, et al. The myth of signing avatars. In *1st International Workshop on Automatic Translation for Signed and Spoken Languages*, 2021.

[3] Maartje De Meulder. Is "good enough"' good enough? ethical and responsible development of sign language technologies. In *Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)*, pages 12–22, 2021.

[4] Maria Kopf, Rehana Omardeen, and Davy Van Landuyt. Representation matters: The case for diversifying sign language avatars. In *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, pages 1–5. IEEE, 2023.

[5] Dimitar Shterionov, Vincent Vandeghinste, Horacio Saggion, Josep Blat, Mathieu De Coster, Joni Dambre, Henk Van den Heuvel, Irene Murtagh, Lorraine Leeson, and Ineke Schuurman. The SignON project: a sign language translation framework. In *the 31st Meeting of Computational Linguistics in the Netherlands*, 2021.

[6] EASIER PROJECT. Easier – intelligent automatic sign language translation, 2021.

[7] Mirella De Sisto, Vincent Vandeghinste, Santiago Egea Gómez, Mathieu De Coster, Dimitar Shterionov, and Horacio Saggion. Challenges with sign language datasets for sign language recognition and translation. In *13th International Conference on Language Resources and Evaluation (LREC)*. European Language Resources Association, 2022.

[8] Dongxu Li, Cristian Rodriguez, Xin Yu, and Hongdong Li. Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 1459–1469, 2020.

[9] Hamid Reza Vaezi Joze and Oscar Koller. Ms-asl: A large-scale data set and benchmark for understanding american sign language. *arXiv preprint arXiv:1812.01053*, 2018.

[10] Oscar Koller, Jens Forster, and Hermann Ney. Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. *Computer Vision and Image Understanding*, 141:108–125, 2015.

[11] Vincent Vandeghinste, Bob Van Dyck, Mathieu De Coster, Maud Goddefroy, and Joni Dambre. BeCoS Corpus: Belgian Covid-19 Sign Language Corpus. A Corpus for Training Sign Language Recognition and Translation. *Computational Linguistics in the Netherlands Journal*, 12:7–17, 2022.

[12] Trevor Johnston. The lexical database of Auslan (Australian Sign Language). *Sign Language & Linguistics*, 4(1-2):145–169, December 2001.

[13] Onno A Crasborn and IEP Zwitserlood. The corpus ngt: an online corpus for professionals and laymen. 2008.

[14] Mieke Van Herreweghe, Myriam Vermeerbergen, Eline Demey, Hannes De Durpel, Hilde Nyffels, and Sam Verstraete. Het Corpus VGT. Een digitaal open access corpus van videos and annotaties van Vlaamse Gebarentaal, ontwikkeld aan de Universiteit Gent i.s.m. KU Leuven. `www.corpusvgt.be`, 2015.

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

[15] Franco Ronchetti, Facundo Quiroga, César Armando Estrebou, Laura Cristina Lanzarini, and Alejandro Rosete. Lsa64: An argentinian sign language dataset. In *XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016)*., 2016.

[16] Ashley Chow, Glenn Cameron, Mark Sherwood, Phil Culliton, Sam Sepah, Sohier Dane, and Thad Starner. Google - isolated sign language recognition, 2023.

[17] Vassilis Athitsos, Carol Neidle, Stan Sclaroff, Joan Nash, Alexandra Stefan, Quan Yuan, and Ashwin Thangali. The american sign language lexicon video dataset. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, June 2008.

[18] Zed Sevcikova Sehyr, Naomi Caselli, Ariel M Cohen-Goldberg, and Karen Emmorey. The ASL-LEX 2.0 project: A database of lexical and phonological properties for 2, 723 signs in american sign language. *The Journal of Deaf Studies and Deaf Education*, 26(2):263–277, February 2021.

[19] Samuel Albanie, Gül Varol, Liliane Momeni, Hannah Bull, Triantafyllos Afouras, Himel Chowdhury, Neil Fox, Bencie Woll, Rob Cooper, Andrew McParland, and Andrew Zisserman. BOBSL: BBC-Oxford British Sign Language Dataset. 2021.

[20] Jérôme Fink, Benoît Frénay, Laurence Meurant, and Anthony Cleve. Lsfb-cont and lsfb-isol: Two new datasets for vision-based sign language recognition. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2021.

[21] Laurence Meurant. Corpus LSFB. Corpus informatisé en libre acces de vidéo et d'annotations de langue des signes de Belgique francophone. Namur: Laboratoire de langue des signes de Belgique francophone (LSFB Lab), FRS-FNRS, Université de Namur, 2015.

[22] Samuel Albanie, Gül Varol, Liliane Momeni, Triantafyllos Afouras, Joon Son Chung, Neil Fox, and Andrew Zisserman. Bsl-1k: Scaling up co-articulated sign language recognition using mouthing cues. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, pages 35–53. Springer, 2020.

[23] Lindsay Ferrara and Rolf Piene Halvorsen. Depicting and describing meanings with iconic signs in norwegian sign language. *Gesture*, 16(3):371–395, December 2017.

[24] Onno Crasborn and Inge Zwitserlood. The Corpus NGT: an online corpus for professionals and laymen. *Construction and Exploitation of Sign Language Corpora. 3rd Workshop on the Representation and Processing of Sign Languages*, 01 2008.

[25] Thad Starner. *Visual recognition of american sign language using hidden markov models*. PhD thesis, Massachusetts Institute of Technology, 1995.

[26] Priyanka Lokhande, Riya Prajapati, and Sandeep Pansare. Data gloves for sign language recognition system. *International Journal of Computer Applications*, 975:8887, 2015.

[27] Michael Erard. Why sign-language gloves don't help deaf people. *The Atlantic*, 9, 2017.

[28] Zahoor Zafrulla, Helene Brashear, Thad Starner, Harley Hamilton, and Peter Presti. American sign language recognition with the kinect. In *Proceedings of the 13th international conference on multimodal interfaces*, pages 279–286, 2011.

[29] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[30] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, et al. Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172*, 2019.

[31] Mathieu De Coster, Ellen Rushe, Ruth Holmes, Anthony Ventresque, and Joni Dambre. Towards the extraction of robust sign embeddings for low resource sign language recognition. *arXiv preprint arXiv:2306.17558*, 2023.

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

[32] Amit Moryossef, Ioannis Tsochantaridis, Joe Dinn, Necati Cihan Camgoz, Richard Bowden, Tao Jiang, Annette Rios, Mathias Muller, and Sarah Ebling. Evaluating the immediate applicability of pose estimation for sign language recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3434–3440, 2021.

[33] Jerome Fink, Pierre Poitier, Maxime ANDRE, Laurence Meurant, Benoît Frénay, and Anthony Cleve. Dictionnaire contextuel langue des signes belge francophone vers français. 2022.

[34] Mathieu De Coster and Joni Dambre. Querying a sign language dictionary with videos using dense vector search. In *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, pages 1–5. IEEE, 2023.

[35] Ozge Mercanoglu Sincan, Anil Osman Tur, and Hacer Yalim Keles. Isolated sign language recognition with multi-scale features using LSTM. In *2019 27th Signal Processing and Communications Applications Conference (SIU)*. IEEE, April 2019.

[36] Mathieu De Coster, Mieke Van Herreweghe, and Joni Dambre. Sign language recognition with transformer networks. In *12th international conference on language resources and evaluation*, pages 6018–6024. European Language Resources Association (ELRA), 2020.

[37] Yao Du, Pan Xie, Mingye Wang, Xiaohui Hu, Zheng Zhao, and Jiaqi Liu. Full transformer network with masking future for word-level sign language recognition. *Neurocomputing*, 500:115–123, 2022.

[38] Yanqiu Liao, Pengwen Xiong, Weidong Min, Weiqiong Min, and Jiahao Lu. Dynamic sign language recognition based on video sequence with blstm-3d residual networks. *IEEE Access*, 7:38044–38054, 2019.

[39] Necati Cihan Camgoz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. Neural sign language translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7784–7793, 2018.

[40] Alex Graves and Alex Graves. Connectionist temporal classification. *Supervised sequence labelling with recurrent neural networks*, pages 61–93, 2012.

[41] Yutong Chen, Ronglai Zuo, Fangyun Wei, Yu Wu, Shujie Liu, and Brian Mak. Two-stream network for sign language recognition and translation. *Advances in Neural Information Processing Systems*, 35:17043–17056, 2022.

[42] Hao Zhou, Wengang Zhou, Yun Zhou, and Houqiang Li. Spatial-temporal multi-cue network for sign language recognition and translation. *IEEE Transactions on Multimedia*, 24:768–779, 2022.

[43] Shiwei Gan, Yafeng Yin, Zhiwei Jiang, Kang Xia, Lei Xie, and Sanglu Lu. Contrastive learning for sign language recognition and translation. In Edith Elkind, editor, *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, pages 763–772. International Joint Conferences on Artificial Intelligence Organization, 8 2023. Main Track.

[44] Mathias Müller, Zifan Jiang, Amit Moryossef, Annette Rios, and Sarah Ebling. Considerations for meaningful sign language machine translation based on glosses. *arXiv preprint arXiv:2211.15464*, 2022.

[45] Mathias Müller, Sarah Ebling, Eleftherios Avramidis, Alessia Battisti, Michele Berger, Richard Bowden, Annelies Braffort, Necati Cihan Camgöz, Cristina Espana-Bonet, Roman Grundkiewicz, et al. Findings of the first wmt shared task on sign language translation (wmt-slt22). In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 744–772, 2022.

[46] Mathieu De Coster, Dimitar Shterionov, Mieke Van Herreweghe, and Joni Dambre. Machine translation from signed to spoken languages: State of the art and challenges. *Universal Access in the Information Society*, pages 1–27, 2023.

[47] Camille Challant and Michael Filhol. A first corpus of azee discourse expressions. In *Language Resources and Evaluation Conference*, 2022.