Quantitative Survey of the State of the Art in Sign Language Recognition

Oscar Koller*
Speech and Language
Microsoft
Munich, Germany
oscar.koller@microsoft.com

August 22, 2020

Abstract

This work presents a meta study covering around 300 published sign language recognition papers with over 400 experimental results. It includes most papers between the start of the field in 1983 and 2020. Additionally, it covers a fine-grained analysis on over 25 studies that have compared their recognition approaches on RWTH-PHOENIX-Weather 2014, the standard benchmark task of the field. Research in the domain of sign language recognition has progressed significantly in the last decade, reaching a point where the task attracts much more attention than ever before. This study compiles the state of the art in a concise way to help advance the field and reveal open questions. Moreover, all of this meta study's source data is made public, easing future work with it and further expansion. The papers have been manually labeled with a set of categories. The data reveals many insights, such as shifts in the field from intrusive to non-intrusive capturing, from local to global features and the lack of non-manual parameters included in medium and larger vocabulary recognition systems. Surprisingly, RWTH-PHOENIX-Weather with a vocabulary of 1080 signs represents the only resource for large vocabulary continuous sign language recognition benchmarking world wide.

Keywords Sign Language Recognition · Survey · Meta Study · State of the Art Analysis

1 Introduction

Since recently, automatic sign language recognition experiences significantly more attention by the community. The number of published studies, but also the quantity of available data sets is increasing. This work aims at providing an overview of the field following a quantitative meta-study approach. For that, the author covered the most relevant 300 published studies, since the earliest known work [Grimes, 1983]. The 300 analyzed recognition studies have been manually labeled based on their basic recognition characteristics such as modeled vocabulary size, the number of contributing signers, the tackled sign language and additional details, such as the quality of the employed data set (e.g. if it covers isolated or continuous sign language), the available input data type (e.g. if provides colors as well as depth information or specific measuring devices for tracking body parts) and the employed sign language modalities and features (e.g. which of the manual and non-manual sign language parameters have been explicitly modeled and which additional features are employed). Based on this data, extensive analysis is presented by creating graphics and tables that relate specific characteristics, visualize correlations, highlight short-comings and allow to create proven hypotheses. Beyond that, this work focuses on the RWTH-PHOENIX-Weather data set, which has evolved to currently be the standard benchmark data set of the sign language recognition field. We provide a detailed structured view comparing over 25 research studies that have evaluated their approaches on the RWTH-PHOENIX-Weather corpus. We track the employed neural architectures, the training style, the employed losses and the data augmentation of all covered studies

*https://www.microsoft.com/en-us/research/people/oskoller/

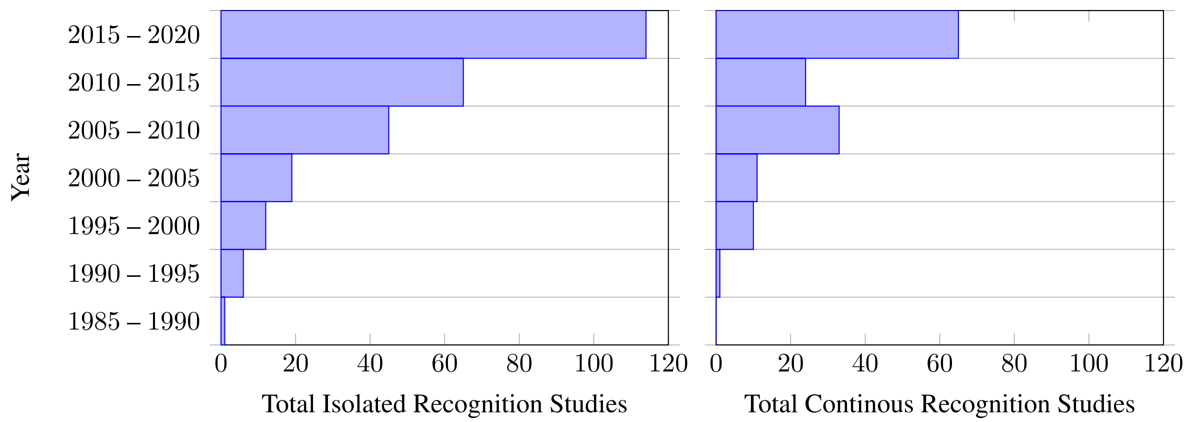

Figure 1: Showing the number of published recognition results between 1983 and 2020.

and present it in a unified table jointly with the achieved performance. The raw data of this work is made publicly available[2]. As such, this paper makes the following contributions:

- Extensive quantitative structured data covering a large part of the sign language recognition research is made publicly available.
- First sign language recognition meta study, providing quantitative insights and analysis of the state of the art.
- First overview and in-depth analysis of all published papers that have compared their proposed recognition systems on PHOENIX 2014, the standard benchmark of the field.

In the following, we will start in Section 2 to dive into the analysis and present the general development of the field, followed by looking into the available input data used for modeling in Section 2.1 and the chosen sign language modalities and features to be modeled in Section 2.2. In Section 2.4, we point out the differences of the research landscape before and after 2015. We compare the studies and investigate general sign language recognition trends as manifested on the RWTH-PHOENIX-Weather 2014 benchmark data set in Section 3. Finally, we conclude this paper with Section 4. The full data table can be found in the appendix.

2 Analysis of the State of the Art

Figure 1 shows the number of published isolated and continuous recognition results in blocks of five years up until 2020. We see that growth looks exponential for isolated studies, while being close to linear for continuous studies. This may reflect the difficulty of the continuous recognition scenario and also the scarcity of available training corpora. On average it seems that there are at least twice as many studies published using isolated sign language data.

However, Figure 2, which shows the number of isolated and continuous recognition results aggregated by vocabulary size, reveals that the vast majority of the isolated sign language recognition works model a very limited amount of signs only (i.e. below 50 signs). This is not the case when comparing continuous sign language recognition, where the overall studies more or less evenly spread across all sign vocabularies (with exception of 500-1000 signs due to lack of available corpora).

Table 1 provides a more detailed perspective on the same data: Here, the number of published results is shown per year and per vocabulary range. In the middle and lower parts of the table, we see this information for isolated and continuous results, respectively, while in the top part of the table it is provided jointly for both data qualities. As in Figure 1 and 2, we note that overall the number of studies increases over the years. However, we also see that this trend is true for the smallest and medium vocabulary (below 50 signs and between 200 and 1000 signs) only. The large vocabulary tasks (over 1000 signs) have been low until year 2015 and following. When looking at the continuous studies only (lower part of Table 1), we see that large vocabulary ($>$ 1000 signs) and 50-200 vocabulary tasks have experienced a large gain in the number of published results since 2015. This can be explained with the community focusing on two benchmark corpora since then ([Koller et al., 2015] with a vocabulary of 1080 signs and [Huang et al., 2018b] with a vocabulary of 178).

[2] https://github.com/oskoller/sign-language-state-of-the-art

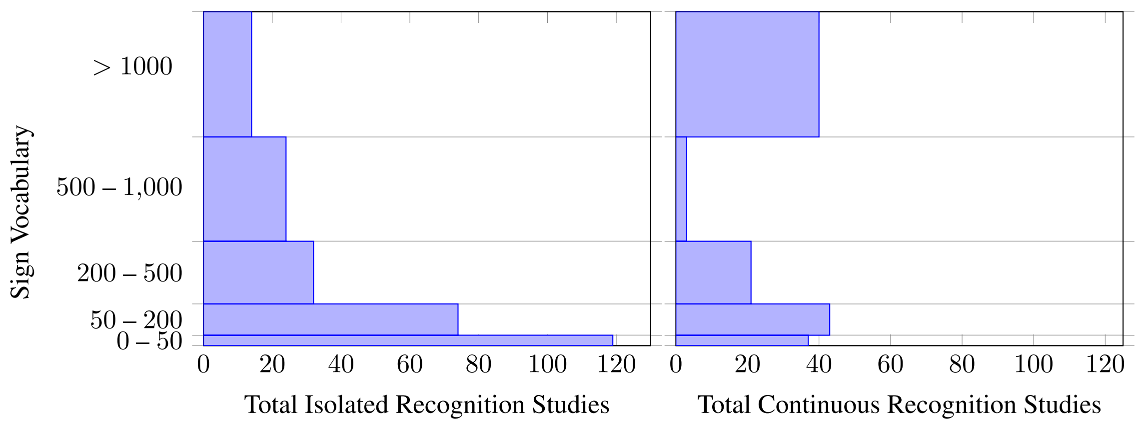

Figure 2: Showing the number of published results between 1983 and 2020 and the size of their modeled sign vocabulary.

Table 1: Shows the number of recognition results that were published in a specific range of years, modeling a specific vocabulary size. The top part of the table show all studies jointly, while the middle and the bottom part of the table show isolated and continuous studies, respectively. E.g. this table reads like: "After 2015, there were 43 results published tackling vocabularies larger than 1000 signs."

		Year						
	Vocabulary	> 2015	$2010 - 2015$	$2005 - 2010$	$2000 - 2005$	$1995 - 2000$	$1990 - 1995$	< 1990
All Studies	> 1000	40	4	3	6	1	0	0
	$500 - 1000$	13	13	1	0	0	0	0
	$200 - 500$	25	15	7	2	3	1	0
	$50 - 200$	51	22	27	10	6	1	0
	$0 - 50$	50	35	40	12	12	5	2
	Vocabulary	> 2015	$2010 - 2015$	$2005 - 2010$	$2000 - 2005$	$1995 - 2000$	$1990 - 1995$	< 1990
Isolated Studies	> 1000	6	2	2	3	1	0	0
	$500 - 1000$	12	11	1	0	0	0	0
	$200 - 500$	19	6	3	1	2	1	0
	$50 - 200$	34	17	12	8	2	1	0
	$0 - 50$	43	29	27	7	7	4	2
	Vocabulary	> 2015	$2010 - 2015$	$2005 - 2010$	$2000 - 2005$	$1995 - 2000$	$1990 - 1995$	< 1990
Continuous Studies	> 1000	34	2	1	3	0	0	0
	$500 - 1000$	1	2	0	0	0	0	0
	$200 - 500$	6	9	4	1	1	0	0
	$50 - 200$	17	5	15	2	4	0	0
	$0 - 50$	7	6	13	5	5	1	0

Table 2: Shows the fraction in [%] of published sign language recognition results that make use of a specific input data type (e.g. 'RGB', 'Depth', etc.) relative to all published results that fall in the same modeled vocabulary range (top part of the table) and that have been published in a similar range of years (bottom part of the table)E.g. this table reads like: "86% of all results with a modeled vocabulary above 1000 signs employ RGB input data. 88% of all results published after 2015 also use depth as input data."

Vocabulary	RGB	Depth	Color Glove	Elect. Glove	Mocap
> 1000	85	4	0	17	13
$500-1000$	93	41	0	4	4
$200-500$	77	23	6	12	12
$50-200$	73	11	6	16	15
$0-50$	72	24	13	10	8
Year	**RGB**	**Depth**	**Color Glove**	**Elect. Glove**	**Mocap**
> 2015	87	22	3	4	6
$2010-2015$	85	38	4	7	7
$2005-2010$	72	1	18	10	6
$2000-2005$	33	0	10	50	57
$1995-2000$	36	0	18	41	23
$1990-1995$	29	0	0	71	29
< 1990	50	0	0	50	0

2.1 Type of Employed Input Data

Table 2 shows in the top part of the type of employed input data across different sizes of modeled vocabulary. The input data refers to the data that is consumed by the recognition algorithms to extract features from and perform computation. We can observe that RGB is the most popular type of input data both for small and larger scale vocabulary ranges. Colored gloves have only ever been applied to small and medium vocabulary tasks and did never get significant attention. The lower part of Table 2 shows the type of employed input data relative to all results published in the same range of years. We can see that RGB data attracts most attention since 2005. Depth as input modality became only popular after the release of the Kinect sensor in 2010. There was one work that employed depth data before [Fujimura and Xia Liu, 2006] which had access to early time-of-flight sensors. Colored gloves got some traction between 1995 and 2010, which looks like a transition phase from electronic measuring devices to pure vision based processing.

Table 3 displays the input data aggregated into the categories 'non-intrusive' and 'intrusive'. Intrusiveness refers to the need to interfere with the recognition subject in order to perform body pose estimation and general feature extraction. As such, 'RGB' and 'Depth' are non-intrusive capturing methods, while 'Color Glove', 'Electronic Glove' and 'Motion Capturing' are intrusive techniques. As can be seen in Table 3 on the left, intrusives capturing methods can be encountered in about one quarter of all experiments with a vocabulary of up to 500 signs. They are more rare in larger vocabulary sizes, possibly due to the fact that those have mainly been researched after 2010 (compare Table 1). We clearly see a paradigm shift after 2005, when the formerly dominating intrusives capturing methods were less and less used and their prevalence decreased from around 70% to less than 30% with a tendency to further reduce over time.

Table 4 shows the number of recognition results per per sign language and employed type of input data. We note that experiments recognizing American sign language (ASL) are clearly dominated by RGB data. Chinese sign language (CSL) has most results using RGB-D (color with depth) data or just RGB data. Gloves make up a significant number of published results in both sign languages as well. German Sign Language (Deutsche Gebärdensprache) (DGS) and most other sign languages focus mainly on RGB based recognition.

2.2 Modeled Sign Language Parameters

In the previous section, we have looked at what kind of input data is being employed for sign language recognition studies. Now, we will investigate the sign language parameters and features that are extracted based on the input data. Therefore, we tagged which sign language parameters are covered by the modeled features. We distinguish manual parameters (i.e. hand shape, movement, location and orientation) and non-manual parameters (i.e. head, mouth, eyes, eye blink, eye brows and eye gaze). For non-manual parameters, it needs to be pointed out that we focused on studies that explicitly target sign language recognition and also include non-manuals. There are many works that focus on non-manual marker recognition for sign language, but these works typically do not model a sign language

Table 3: Shows the fraction in [%] of published sign language recognition results that make use of non-intrusives data input capturing methods (i.e. ‘RGB’ or ‘Depth’) and those that are intrusives (i.e. ‘Color Glove’, ‘Elect. Glove’ or ‘Mocap’) relative to all published results that fall in the same modeled vocabulary range (left table) and relative to a year range (right table). E.g. this table reads like: “84% of all published results with a modeled vocabulary larger than 1000 signs employ non-intrusives input data capturing methods.”

Vocabulary	non-Intrusive	Intrusive
> 1000	83	17
$500 - 1000$	93	7
$200 - 500$	77	23
$50 - 200$	73	27
$0 - 50$	74	26

Year	non-Intrusive	Intrusive
> 2015	89	11
$2010 - 2015$	87	13
$2005 - 2010$	72	28
$2000 - 2005$	30	70
$1995 - 2000$	27	73
$1990 - 1995$	29	71
< 1990	50	50

Table 4: Shows the number of published recognition results per sign language and type of input data. The sign language abbreviations are mentioned in the appendix. The sign languages are ordered by result counts. This table reads like: “99 results were published for ASL that used RGB input data.”

Input Data	ASL	CSL	DGS	BSL	ArSL	JSL	IndianSL	GSL	TSL	NGT	FlemishSL	LIS	Auslan	ArgentSL	TaiwanSL	PolishSL	Libras	IrishSL	KSL	IndoSL	CzSL	MalaySL	K-RSL	DSGS	TamisSL	PersianSL	MexicanSL	LSE	KurdishSL	ISL
RGB	93	53	59	18	7	6	7	12	5	7	7	4	0	0	1	4	4	1	1	3	3	2	2	2	1	1	1	1	1	0
Depth	14	38	3	0	2	1	1	1	4	0	0	0	0	0	0	2	2	0	0	1	1	0	0	2	0	0	1	0	0	1
Color Glove	4	1	4	2	4	0	0	0	4	2	0	0	1	5	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
Elect. Glove	16	17	1	0	4	2	0	1	0	0	0	1	4	0	3	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
Mocap	14	15	0	0	4	3	4	0	0	0	0	0	1	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0

recognition problem. Additionally, we track common features that capture a global view of the signers (i.e. body joints, fullframe RGB images covering the full signer, fullframe depth images and fullframe motion images with optical flow). Table 5 shows the employed sign language parameters and features relative to all results published using a similar sign vocabulary (top of the table) and relative to all results published during a similar time (lower part of the table).

We note that hand shape is the most covered parameter, while location and movement are the next popular parameters across all vocabulary sizes below 1000 signs. Fullframe features followed by hand shapes are most frequently encountered in large vocabulary tasks beyond 1000 signs. The lower part of Table 5 confirms that since 2015 fullframe features have become the most frequently encountered feature (while being very close to hand shape features). Furthermore, it can be noticed that since 2015 hand shape are tackled by a much larger fraction of published results. It needs to be pointed out that while most studies that have been published after 2015 employ a cropped hand patch as input to their recognition systems, we tagged that with the hand shape parameter. However, using deep learning based feature extractors, such hand inputs may implicitly learn hand posture / orientation parameters. Similarly, global input features such as fullframe inputs may implicitly help to learn location and movement parameters and, to a lesser degree, all other parameters as well as the full image comprises all available information.

Table 6 aggregates hand location, movement, shape and orientation into manual parameters. Head, mouth, eyes, eye blink, eyebrows and eye gaze are referred to as non-manual parameters. Body joints, fullframe, depth and motion are all computed on the full image and hence we call them global features. We can see that with larger modeled vocabularies the trend goes from manual to global features (left side of Table 6), where the latter increase from 18% usage across all published results with vocabularies of up to 50 signs to 62% with large vocabularies above 1000 signs. The increase of global features may have two reasons:

1. The availability of body joints and full depth image features with the release of the Kinect in 2010.

2. The shift towards deep learning and trend to input fullframes instead of manual feature engineering.

Table 5: Shows the fraction in [%] of published sign language recognition results that make use of a specific sign language parameter (e.g. 'Loc.', 'Mov.', etc.) relative to all published results that fall in the same vocabulary range (top part of the table), or in the same range of years (lower part of the table). 'Loc.', 'Mov.', 'Shape' and 'Orient.' stand for hand location, movement, shape and orientation (manual parameters). 'Joints' refers to tracked body joint locations. 'Fullframe' and 'Depth' are the full RGB and depth image, respectively, while 'Motion' unites all types of motion estimation on the full image (often optical flow). E.g. this table reads like: "27% of all results with a modeled vocabulary above 1000 signs include the location modality."

Vocabulary	Loc.	Mov.	Shape	Orient.	Head	Mouth	Eyes	Blink	Brows	Gaze	Joints	Fullframe	Depth	Motion
> 1000	28	17	47	19	11	9	6	0	6	0	11	55	0	4
$500 - 1000$	46	58	58	4	0	0	0	0	0	0	42	21	0	4
$200 - 500$	44	35	73	17	21	8	2	0	2	0	25	27	0	4
$50 - 200$	56	52	56	17	7	2	1	0	0	0	14	23	0	1
$0 - 50$	55	52	66	17	6	3	2	0	2	0	12	8	1	1
Year	**Loc.**	**Mov.**	**Shape**	**Orient.**	**Head**	**Mouth**	**Eyes**	**Blink**	**Brows**	**Gaze**	**Joints**	**Fullframe**	**Depth**	**Motion**
> 2015	23	22	43	9	6	5	3	0	3	0	22	46	1	4
$2010 - 2015$	65	67	81	12	17	2	2	0	2	0	31	6	0	0
$2005 - 2010$	73	71	65	8	10	6	0	0	0	0	0	1	0	0
$2000 - 2005$	87	57	80	53	0	0	0	0	0	0	0	0	0	0
$1995 - 2000$	68	45	77	55	5	0	0	0	0	0	0	0	0	0
$1990 - 1995$	57	57	86	71	0	0	0	0	0	0	0	0	0	0
< 1990	50	50	100	50	0	0	0	0	0	0	0	0	0	0

Both hypotheses can be confirmed by looking at the right side of Table 6. There, we see that global features started gaining traction just after 2010 (release of the Kinect) and also coincides with when deep learning for sign language took off in 2015.

While for the previous tables each sign language parameter has been looked at separately and tagged when present, Table 7 shows the frequency of combinations of features over different vocabularies. Hence, if a study models two types of parameters their combination will appear in this table. Inline with previous results, we see that fullframe features alone are by far the most popular on large vocabulary (> 1000 signs) tasks. They are followed by hand shape features and bodyjoints. On very small vocabulary (< 50 signs) tasks, a preference on hand shape features can be noticed.

Table 8 shows the number of published results with employed parameters broken down per sign language. In the top part of the table all studies are reflected, while the lower part of the table only shows studies with a vocabulary of at least 200 signs. We see that while ASL has the most published results overall, non-manual parameters (e.g. head, mouth or eyes) are most frequently included in studies on DGS. It is also striking that despite the fact that CSL is the second most frequently researched sign language, there is only a single study that includes non-manual parameters like the face [Zhou et al., 2020a]. We also note that there are studies on smaller sign languages such as Kazakh-Russian sign language (K-RSL) that explicitly focus on non-manual parameters [Mukushev et al., 2020, Sabyrov et al., 2019]. Eyes and specifically eyebrows have only been tackled in few studies [Koller et al., 2016a, Koller et al., 2015, Koller et al., 2016b, Mukushev et al., 2020, Sabyrov et al., 2019, Yang and Lee, 2011, Zhang et al., 2016a], while, to the best of our knowledge, no single work has explicitly included eye gaze or eye blinks for sign language recognition. In the lower part of Table 8 studies are limited to have at least a vocabulary of 200 signs. Besides two British sign language (BSL) studies [Albanie et al., 2020, von Agris et al., 2008b], all others are works on DGS, covering the 450 sign vocabulary corpus SIGNUM [Oberdörfer et al., 2012, von Agris et al., 2008a] and the 1080 sign vocabulary corpus RWTH-PHOENIX-Weather [Forster et al., 2013a, Forster et al., 2013b, Koller et al., 2016a, Koller et al., 2015, Koller et al., 2016b, Zhou et al., 2020a]. [Zhou et al., 2020a] is the first work that uses the face in a deep learning based large vocabulary task.

2.3 Analysis by Sign Language

Table 9 and Table 10 show the number of published recognition results per sign language over time and per modeled vocabulary range, respectively. ASL has usually been the sign language with the most results published. However, we see in Table 10 that this is only true for vocabularies below 200 signs. On larger vocabularies CSL is leading and on vocabularies above 1000 signs DGS has significantly more research published. Table 10 further reveals that it is

Table 6: Shows the fraction in [%] of published sign language recognition results that employ manual, non-manual or global features relative to all published results that fall in the same vocabulary range (left side) or the same range of years (right side). Manual parameters refer to hand location, movement, shape and orientation. Non-manual parameters are head, mouth, eyes, eyeblink, eyebrow and eyegaze features. Global features refer to body joints, fullframe, depth and motion features. E.g. this table reads like: "46% of all results with a modeled vocabulary above 1000 signs include manual parameters."

Vocabulary	Manual	non-Manual	Global
> 1000	49	15	64
$500 - 1000$	67	0	62
$200 - 500$	77	23	52
$50 - 200$	74	7	35
$0 - 50$	90	7	20

Year	Manual	non-Manual	Global
> 2015	47	8	66
$2010 - 2015$	100	17	37
$2005 - 2010$	99	13	1
$2000 - 2005$	100	0	0
$1995 - 2000$	100	5	0
$1990 - 1995$	100	0	0
< 1990	100	0	0

Table 7: Shows the 26 most frequently used modality combinations and their relative frequency of use as. This is displayed as fraction in [%] of published sign language recognition results that make use of the specific combination of sign language parameters relative to all published results that fall in the same vocabulary range. 'Loc.', 'Mov.', 'Shape' and 'Orient.' stand for hand location, movement, shape and orientation (manual parameters). 'Joints' refers to tracked body joint locations. 'Fullframe' and 'Depth' are the full RGB and depth image, respectively, while 'Motion' unites all types of motion estimation on the full image (often optical flow). E.g. this table reads like: "39% of all results with a modeled vocabulary above 1000 signs rely fully on the fullframe modality, while 7% rely on the hand shape modality."

	Vocabulary				
Modality Combination	> 1000	$500 - 1000$	$200 - 500$	$50 - 200$	$0 - 50$
Fullframe	39	15	13	20	6
Shape	7	0	9	5	18
Loc.-Mov.-Shape	4	11	0	14	12
Loc.-Mov.	0	4	0	9	11
Loc.-Mov.-Shape-Orient.	6	4	4	8	7
Mov.-Shape	2	11	2	3	7
Loc.-Shape	0	0	9	8	3
Joints	6	11	8	5	1
Loc.-Shape-Orient.	9	0	4	3	4
Loc.-Mov.-Shape-Joints	0	15	4	3	3
Loc.-Shape-Joints	2	7	8	2	1
Mov.	0	4	2	3	2
Loc.	0	0	0	2	4
Mov.-Shape-Orient.	0	0	2	3	2
Shape-Orient.	0	0	4	0	3
Mov.-Shape-Head-Fullframe	0	0	9	0	0
Loc.-Mov.-Shape-Head	0	0	0	1	3
Shape-Fullframe	4	0	0	1	1
Shape-Head-Joints-Fullframe	4	0	0	1	0
Mov.-Shape-Joints	0	4	2	0	1
Loc.-Shape-Head	0	0	4	0	1
Loc.-Mov.-Shape-Head-Mouth-Eyes-Brows	4	0	2	0	0
Loc.-Mov.-Shape-Head-Mouth	0	0	4	1	0
Loc.-Mov.-Head	0	0	0	2	1
Loc.-Joints	0	0	0	0	2

Table 8: Shows the number of published recognition results per sign language and employed sign language modality. The top part of the table shows all studies, while the lower part only shows studies with a vocabulary of at least 200 signs. The sign language abbreviations are mentioned in the appendix. This table reads like: "There are 62 results published that use the location modality in the recognition of ASL."

	Modalities	ASL	CSL	DGS	BSL	ArSL	JSL	IndianSL	GSL	TSL	NGT	FlemishSL	LIS	Auslan	ArgentSL	TaiwanSL	PolishSL	Libras	IrishSL	KSL	IndoSL	CzSL	MalaySL	K-RSL	DSGS	TamisSL	PersianSL	MexicanSL	LSE	KurdishSL	ISL
All Studies	Location	62	38	16	12	12	8	3	9	7	3	0	3	3	2	2	4	1	3	2	2	2	0	0	0	0	0	1	0	0	1
	Movement	63	32	16	14	7	7	2	11	5	2	0	2	2	2	2	2	2	4	1	1	1	0	0	1	0	1	0	0	0	1
	Shape	68	49	34	15	8	7	5	10	7	2	0	2	5	4	3	3	3	1	3	2	3	2	0	1	1	1	0	0	1	1
	Orientation	18	19	1	4	3	2	1	0	0	2	0	2	4	1	3	0	1	0	2	0	0	0	0	0	0	0	0	0	0	1
	Head	8	1	15	2	0	1	1	0	3	0	0	1	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
	Mouth	2	0	8	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
	Eyes	2	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
	Eyeblink	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Eyebrows	1	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
	Gaze	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Bodyjoints	11	26	7	2	0	0	5	4	1	0	1	0	0	2	0	2	0	0	0	0	0	0	2	0	0	0	1	0	0	0
	Fullframe	21	11	34	1	0	0	2	2	1	5	6	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
	Fullframedepth	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Fullframemotion	0	1	4	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Studies with vocabulary ≥ 200	Location	2	25	8	3	2	0	2	6	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Movement	2	20	10	1	0	0	1	6	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Shape	3	34	26	3	2	0	2	5	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Orientation	1	14	1	1	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Head	0	0	14	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Mouth	0	0	7	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Eyes	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Eyeblink	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Eyebrows	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Gaze	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Bodyjoints	6	14	3	1	0	0	4	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Fullframe	13	0	33	1	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Fullframedepth	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Fullframemotion	0	1	3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

RWTH-PHOENIX-Weather with a vocabulary of over 1000 signs that represents the only resource for large-scale continuous sign language world wide.

This can partly be explained by the public availability of sign language data sets, which represent after all still extremely low resource languages. However, there are corpora available for ASL that have larger vocabularies (e.g. ASLLRP [Neidle and Vogler, 2012]). It seems there is necessity for the corpora to be packaged for reproducible sign language recognition research. Fixed partitions into train, development and test and an easily accessible download method are required. Also, the licenses under which the corpora are provided may impact dissemination.

2.4 Change of Continuous Recognition Landscape After 2015

Figure 3 shows counts of published continuous sign language recognition experiments and modeled vocabularies before and after 2015. Prior 2015, there were 80 results published, while after 2015 66 results can be found. We note that prior 2015 most studies model different vocabulary sizes, while after 2015 there is a large peak of close to 30 published results at a 1080 vocabulary and a smaller peak at a vocabulary size of 178.

Table 9: Shows the number of published recognition results per sign language and year. The sign language abbreviations are mentioned in the appendix. This table reads like: "There are 50 results published after 2015 that use ASL."

	Year	ASL	CSL	DGS	BSL	ArSL	JSL	IndianSL	GSL	TSL	NGT	FlemishSL	LIS	Auslan	ArgentSL	TaiwanSL	PolishSL	Libras	IrishSL	KSL	IndoSL	CzSL	MalaySL	K-RSL	DSGS	TamisSL	PersianSL	MexicanSL	LSE	KurdishSL	ISL
All Studies	> 2015	46	35	40	2	4	1	9	2	4	5	7	1	1	5	0	0	1	0	1	3	1	0	2	2	1	1	1	1	1	1
	$2010 - 2015$	21	21	17	2	7	1	2	10	0	0	0	2	0	0	0	2	2	1	0	0	0	0	0	0	0	0	0	0	0	0
	$2005 - 2010$	30	4	4	13	6	2	0	1	5	2	0	2	0	0	0	1	1	3	0	0	2	2	0	0	0	0	0	0	0	0
	$2000 - 2005$	11	10	2	3	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$1995 - 2000$	5	2	1	0	0	4	0	0	0	2	0	0	3	0	3	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
	$1990 - 1995$	5	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	< 1990	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Isolated Studies	> 2015	43	20	1	2	1	0	7	1	4	5	7	1	1	5	0	0	1	0	1	3	1	0	2	2	1	1	1	1	1	0
	$2010 - 2015$	13	19	4	2	6	1	2	10	0	0	0	2	0	0	0	2	2	1	0	0	0	0	0	0	0	0	0	0	0	0
	$2005 - 2010$	9	3	1	10	5	1	0	1	4	2	0	1	0	0	0	0	1	3	0	0	2	2	0	0	0	0	0	0	0	0
	$2000 - 2005$	6	6	2	3	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$1995 - 2000$	1	2	0	0	0	3	0	0	0	1	0	0	3	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	$1990 - 1995$	4	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	< 1990	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Continuous Studies	> 2015	3	15	39	0	3	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	$2010 - 2015$	8	2	13	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$2005 - 2010$	21	1	3	3	1	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$2000 - 2005$	5	4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$1995 - 2000$	4	0	1	0	0	1	0	0	0	1	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	$1990 - 1995$	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	< 1990	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 10: Shows the number of published recognition results per sign language and modeled vocabulary. The sign language abbreviations are mentioned in the appendix. This table reads like: "There are 6 recognition results of ASL published."

	Vocabulary	ASL	CSL	DGS	BSL	ArSL	JSL	IndianSL	GSL	TSL	NGT	FlemishSL	LIS	Auslan	ArgentSL	TaiwanSL	PolishSL	Libras	IrishSL	KSL	IndoSL	CzSL	MalaySL	K-RSL	DSGS	TamisSL	PersianSL	MexicanSL	LSE	KurdishSL	ISL
All Studies	> 1000	4	11	36	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$500 - 1000$	7	10	2	0	0	0	1	5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$200 - 500$	8	17	16	3	2	0	2	2	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$50 - 200$	43	20	3	8	7	4	3	4	1	4	4	2	4	2	2	1	1	0	1	0	0	0	0	2	0	0	0	0	0	0
	$0 - 50$	57	14	7	7	8	7	5	1	6	4	3	3	1	3	1	3	3	4	2	3	3	2	2	0	1	1	1	1	1	1
Isolated	> 1000	4	7	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$500 - 1000$	7	9	0	0	0	0	1	5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$200 - 500$	8	16	0	1	2	0	2	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$50 - 200$	24	7	1	8	2	3	3	4	1	4	4	2	4	2	1	0	1	0	0	0	0	0	0	2	0	0	0	0	0	0
	$0 - 50$	34	11	7	6	8	5	3	1	5	3	3	2	1	3	1	2	3	4	2	3	3	2	2	0	1	1	1	1	1	0
Continuous	> 1000	0	4	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$500 - 1000$	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$200 - 500$	0	1	16	2	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$50 - 200$	19	13	2	0	5	1	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	$0 - 50$	23	3	0	1	0	2	2	0	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Since 2015, the sign language recognition community is focusing more on benchmark data sets, which explains these characteristics. RWTH-PHOENIX-Weather 2014 [Koller et al., 2015] has a vocabulary of 1080 and the CSL corpus [Huang et al., 2018b] covers 178 signs. In the following section, we will provide a deep analysis of the research studies that compared their work on the PHOENIX corpus.

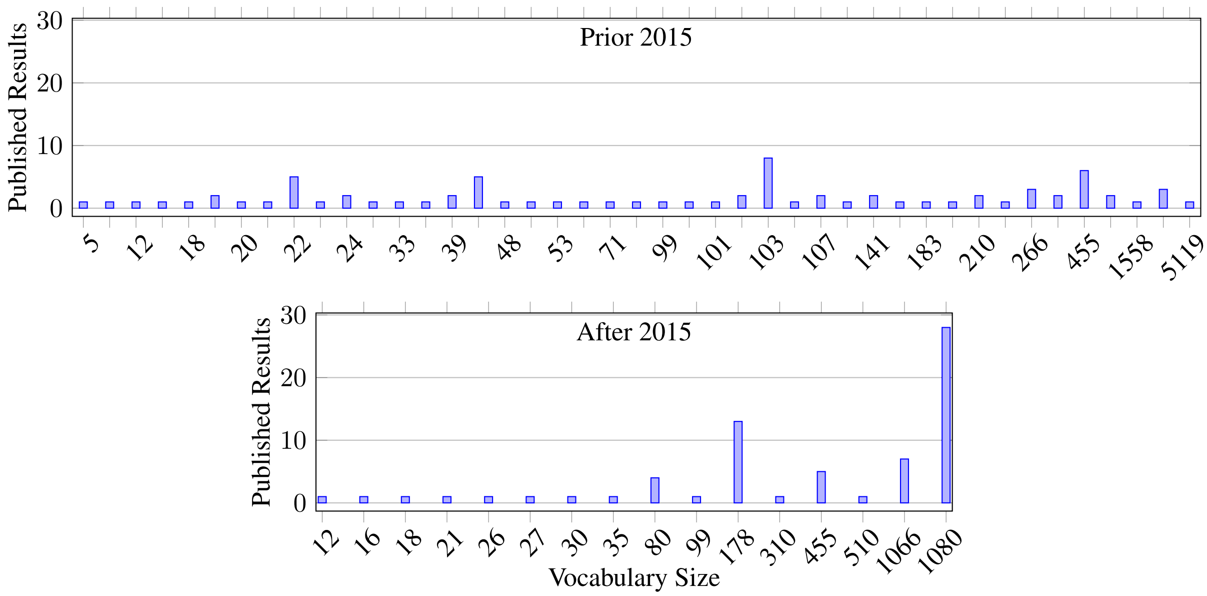

Figure 3: Showing the number of published continuous sign language recognition results per modeled vocabulary (prior to 2015 on the top and 2015-2020 on the bottom plot). This allows to see that after 2015 researcher have started to focus on few benchmark data sets.

3 Analysis of PHOENIX 2014 Benchmark Papers

Table 11 and Table S1 present, to the best of our knowledge, all known results on the RWTH-PHOENIX-Weather 2014 continuous sign language recognition benchmark that have been published as of June 2020. Table 11 provides information on the employed features, the chosen neural architecture and the achieved results, while Table S1 shows what kind of data augmentation was used, if iterative training was employed and what training losses were part of the optimization. Iterative training refers to an expectation maximization (EM) like training procedure where a trained model is used to create pseudo labels on the training data which will then be used to train a part or the full recognition network. Inspired by EM training in Gaussian mixture model (GMM) hidden Markov model (HMM) systems, this way of training was first proposed in [Koller et al., 2016a]. It was then adopted by many teams as can be seen in Table S1. Besides [Cheng et al., 2020], all best performing approaches on PHOENIX with a word error rates (WERs) below 27.0% make use of iterative training procedures. In many works it is described to help overcome vanishing gradients issues when training deep convolutional neural network (CNN) architectures that are succeeded by bi-directional long short-term memory (BLSTM) layers [Zhou et al., 2020a, Papastratis et al., 2020, Cui et al., 2019].

The employed losses are very diverse, as can be seen in Table S1. However, most networks that achieve below 30.0% WER are trained with cross-entropy (CE) loss and also use connectionist temporal classification (CTC) loss. Additionally, a variety of different loss terms are reported ranging from Kullback-Leibler (KL) divergence, over squared error to smooth-L1 loss and others.

Table 11 shows that [Cui et al., 2017] first suggested the use of 2D convolutions followed by 1D convolutions on PHOENIX. Later, [Tran et al., 2018] did a detailed analysis for action recognition. All best performing approaches on PHOENIX with WERs below 25.0% employ 2D+1D convolutions [Cheng et al., 2020, Cui et al., 2019, Papastratis et al., 2020, Zhou et al., 2020a].

While the early works on PHOENIX all relied on tracked and cropped hand shape features [Koller et al., 2017] first proposed to train the CNNs directly on the fullframe input image. This trend continues and all recent studies rely on this feature (e.g. [Adaloglou et al., 2020, Cheng et al., 2020, Papastratis et al., 2020, Zhou et al., 2020b, Zhou et al., 2020a]).

In terms of data augmentation, the most popular choice seems to be random cropping, followed by temporal scaling (re-sampling or random frame drop) as can be seen in Table S1. However, many papers do not specify any augmentation methods leaving the reader without a clear understanding of what happens. While augmentation certainly has significant impact on the results, no study has yet analyzed the effect of the various augmentation options.

Table 11: The table covers (to the best of our knowledge) all published sign language recognition works until mid 2020 that reported results on the RWTH-Phoenix Weather 2014 [Koller et al., 2015] task. It allows to compare the type of employed features (manual, non-manual and fullframe features), the employed neural architectures and the achieved WER on the development and test partition of the corpus.

			Manuals				Non-M.				Fullframe			Neural Architecture						WER	
Reference	Group	Short Title	Location	Movement	Shape	Orientation	Head	Mouth	Eyes	Eyebrows	Bodyjoints	RGB	Motion	CNN	BLSTM	LSTM	BGRU	Attention	Transformer	Dev	Test
[Koller et al., 2015]	RWTH	CSLR	x	x	x		x	x	x	x										55.0	53.0
[Koller et al., 2016a]	RWTH/Surrey	Align Hamnosys	x	x	x	x	x	x	x	x				2d						49.6	48.2
[Koller et al., 2016b]	RWTH/Surrey	1 Million Hands	x	x	x		x	x	x	x				2d						47.1	45.1
[Koller et al., 2016c]	RWTH/Surrey	Deep Sign			x									2d						38.3	38.8
[Camgoz et al., 2017]	Surrey/RWTH	SubUNets			x							x		2d	x					40.8	40.7
[Cui et al., 2017]	Tsinghua	Staged Optimization			x									2d-1d	x					39.4	38.7
[Koller et al., 2017]	RWTH	Re-Align										x		2d	x					27.1	26.8
[Huang et al., 2018b]	USTC/Here	Without Segmentation			x							x		3d	x			x		-	38.3
[Wang et al., 2018]	Hefei Tech/USTC	Temporal Fusion										x		3d-1d			x			37.9	37.8
[Pu et al., 2018]	USTC	Dilated Convolutions										x		3d-Dilated						38.0	37.3
[Koller et al., 2018]	RWTH/Surrey	Hybrid CNN-HMMs			x									2d						31.6	32.5
[Pei et al., 2019]	Hefei Tech	Pseudo Supervised Learning										x		3d			x			40.9	40.6
[Song et al., 2019]	Hefei Tech	Parallel Temp. Encoder										x		3d	x					38.1	38.3
[Zhang et al., 2019]	USTC	Reinforcement Learning										x		3d					x	38.0	38.3
[Cui et al., 2019]	Tsinghua	Iterative Training											x	2d-1d	x					37.9	37.6
[Pu et al., 2019]	USTC	Iterative Alignment Network										x		3d	x	x		x		37.1	36.7
[Guo et al., 2019]	Hefei Tech/Huawei	Dense Temporal Conv.										x		3d-1d						35.9	36.5
[Yang et al., 2019]	Tencent/HKUST	SF-Net										x		3d-2d	x					35.6	34.9
[Zhou et al., 2019]	USTC	Pseudo Label Decoding										x		3d-1d			x			35.6	34.5
[Cui et al., 2019]	Tsinghua	Iterative Training			x									2d-1d	x					31.7	31.5
[Koller et al., 2019]	RWTH/Surrey	Multi-Stream CNN-HMMs			x			x				x		2d	x					26.0	26.0
[Cui et al., 2019]	Tsinghua	Iterative Training										x		2d-1d	x					23.8	24.4
[Cui et al., 2019]	Tsinghua	Iterative Training										x	x	2d-1d	x					23.1	22.9
[Zhou et al., 2020b]	HKBU/HKU/BJTU/Nvidia	Fully-Inception Networks										x		2d-1d					x	31.7	31.3
[Adaloglou et al., 2020]	CERTH/Patras	Comprehensive Study										x		2d-1d	x					28.9	29.1
[Cheng et al., 2020]	HKUST/Tencent/Kwai	Fully Conv Networks										x		2d-1d						24.6	24.6
[Papastratis et al., 2020]	CERTH	Cross-Modal Alignment										x		2d-1d	x					23.9	24.0
[Zhou et al., 2020a]	USTC	ST Multi-Cue Network			x		x				x	x		2d-1d	x					21.1	20.7

Table 12: The table covers (to the best of our knowledge) all published sign language recognition works until mid 2020 that reported results on the RWTH-Phoenix Weather 2014 [Koller et al., 2015] task. It allows to compare the type of employed data augmentation and the employed loss. Additionally, it can be seen if a paper performed an iterative training and the achieved performance in WER

				Data Augmentation											Employed Loss						WER	
Reference	Group	Short Title	Iterative Training	Crop	Framedrop	Temporal Scaling	Spatial Scaling	Intensity Noises	Flip	Brightness	Contrast	Hue	Saturation	Not Specified	CE	CTC	KL-Divergence	Squared Error	Reinforce	Other	Dev	Test
[Koller et al., 2015]	RWTH	CSLR																			55.0	53.0
[Koller et al., 2016a]	RWTH/Surrey	Align Hamnosys	x	x					x						x						49.6	48.2
[Koller et al., 2016b]	RWTH/Surrey	1 Million Hands	x	x					x						x						47.1	45.1
[Koller et al., 2016c]	RWTH/Surrey	Deep Sign		x					x						x						38.3	38.8
[Camgoz et al., 2017]	Surrey/RWTH	SubUNets	x											x		x					40.8	40.7
[Cui et al., 2017]	Tsinghua	Staged Optimization	x			x										x	x			x	39.4	38.7
[Koller et al., 2017]	RWTH	Re-Align	x	x					x						x						27.1	26.8
[Huang et al., 2018b]	USTC/Here	Without Segmentation												x						x	-	38.3
[Wang et al., 2018]	Hefei Tech/USTC	Temporal Fusion												x		x					37.9	37.8
[Pu et al., 2018]	USTC	Dilated Convolutions	x											x	x	x					38.0	37.3
[Koller et al., 2018]	RWTH/Surrey	Hybrid CNN-HMMs		x					x						x						31.6	32.5
[Pei et al., 2019]	Hefei Tech	Pseudo Supervised Learning	x											x		x					40.9	40.6
[Song et al., 2019]	Hefei Tech	Parallel Temp. Encoder	x											x		x		x			38.1	38.3
[Zhang et al., 2019]	USTC	Reinforcement Learning												x					x		38.0	38.3
[Cui et al., 2019]	Tsinghua	Iterative Training	x			x	x	x								x	x			x	37.9	37.6
[Pu et al., 2019]	USTC	Iterative Alignment Network	x											x	x	x					37.1	36.7
[Guo et al., 2019]	Hefei Tech/Huawei	Dense Temporal Conv.												x		x					35.9	36.5
[Yang et al., 2019]	Tencent/HKUST	SF-Net		x												x	x				35.6	34.9
[Zhou et al., 2019]	USTC	Pseudo Label Decoding	x											x	x	x	x				35.6	34.5
[Cui et al., 2019]	Tsinghua	Iterative Training	x			x	x	x								x	x			x	31.7	31.5
[Koller et al., 2019]	RWTH/Surrey	Multi-Stream CNN-HMMs	x	x					x						x						26.0	26.0
[Cui et al., 2019]	Tsinghua	Iterative Training	x			x	x	x								x	x			x	23.8	24.4
[Cui et al., 2019]	Tsinghua	Iterative Training	x			x	x	x								x	x			x	23.1	22.9
[Zhou et al., 2020b]	HKBU/HKU/BJTU/Nvidia	Fully-Inception Networks		x		x										x				x	31.7	31.3
[Adaloglou et al., 2020]	CERTH/Patras	Comprehensive Study		x	x					x	x	x	x		x	x				x	28.9	29.1
[Cheng et al., 2020]	HKUST/Tencent/Kwai	Fully Conv Networks		x		x									x	x					24.6	24.6
[Papastratis et al., 2020]	CERTH	Cross-Modal Alignment	x	x	x					x	x	x	x		x	x				x	23.9	24.0
[Zhou et al., 2020a]	USTC	ST Multi-Cue Network	x	x	x				x						x	x				x	21.1	20.7

4 Conclusion and Outlook

In this paper we shared, to the best of our knowledge, the most extensive quantitative study on the field of sign language recognition covering analysis of over 300 publications from 1983 till 2020. All analyzed studies have been manually tagged with a number of categories. This source data is shared in the supplementary materials of this work. Among others, we present following findings in this meta study:

- While many more studies are published on isolated than on continuous sign language recognition, the majority only covers small vocabulary tasks.

- After 2005 there was a paradigm shift in the community abandoning intrusive capturing methods and embracing non-intrusive methods.
- Deep learning led the community towards the predominant use of global feature representations that are based on fullframe inputs. Those are particularly more common for larger vocabulary tasks.
- Non-manual parameters are still very rare in sign language recognition systems, despite their known importance for sign languages [Pfau and Quer, 2010]. No sign recognition work has included eye gaze or blinks yet. Despite being the second most frequently researched sign language, research studies for CSL have hardly incorporated non-manual parameters. DGS is currently the only sign language where non-manuals have been successfully incorporated considering tasks with a vocabulary of at least 200 signs.
- RWTH-PHOENIX-Weather with a vocabulary of 1080 signs represents the only resource for large vocabulary continuous sign language world wide.

Moreover, We also presented the first meta analysis covering all known works that compared themselves on the RWTH-PHOENIX-Weather benchmark data set. Besides many details, we note that the best performing systems typically adopt an iterative training style to overcome vanishing gradients in deep CNN architectures followed by BLSTMs. We also find that 2D convolutions followed by 1D convolutions on fullframe inputs can be encountered in most state-of-the-art systems. Surprisingly, we see that in many studies data augmentation is not carefully described and also an ablation study that details the effect of various augmentation methods is left for coming research.

We hope that in the future more works will include and be led by Deaf researchers, which seems the only viable way to continue on this accelerated path the field is currently on. More efforts are needed to create real-life large vocabulary continuous sign language tasks that should be made publicly accessible with well defined train, development and test partitions.

Acronyms

ArgentSL Argentinian sign language.

ArSL Arabic sign language.

ASL American sign language.

Auslan Australian sign language.

BLSTM bi-directional long short-term memory.

BSL British sign language.

CE cross-entropy.

CNN convolutional neural network.

CSL Chinese sign language.

CTC connectionist temporal classification.

CzSL Czech sign language.

DGS German Sign Language (Deutsche Gebärdensprache).

DSGS Swiss German sign language / Deutschschweizerische Gebärdensprache.

EM expectation maximization.

FlemishSL Flemish sign language.

GMM Gaussian mixture model.

GSL Greek sign language.

HKSL Hong Kong sign language.

HMM hidden Markov model.

IndianSL Indian sign language.

IndoSL Indonesian sign language.

IrishSL Irish sign language.

ISL Irish Sign Language.

JSL Japanese sign language.

KL Kullback-Leibler.

K-RSL Kazakh-Russian sign language.

KSL Korean sign language.

KurdishSL Kurdish sign language.

Libras Brazilian sign language / Lingua Brasileira de sinais.

LIS Italian sign language / Lingua Italiana dei segni.

LSE Spanish sign language / Lengua de signos española.

MalaySL Malaysian sign language.

MexicanSL Mexican sign language.

NGT Dutch sign language / Nederlandse Gebaren Taal.

PersianSL Persian sign language.

PolishSL Polish sign language.

RussianSL Russian sign language.

TaiwanSL Taiwanese sign language.

TamisSL Tamil sign language.

TSL Turkish sign language.

WER word error rate.

Glossary

continuous Specifies the nature of sign language data sets that encompass long phrases or full sentences as opposed to single, isolated signs.

intrusive Specifies the capturing of sign language data sets that requires the signer to wear specific measuring devices such as gloves or trackers..

isolated Specifies the nature of sign language data sets that only encompass single signs as opposed to long phrases or full sentences..

non-intrusive Specifies the capturing of sign language data sets that does not require the signer to wear specific measuring devices such as gloves or trackers..

parameter Each sign is consists of a set of parameters. We distinguish manual and non-manual parameters. Hand shape, orientation, location and movement are the four manual parameter, while non-manual parameters include head and body posture, facial expression, eye gaze and mouth patterns..

vocabulary The set of unique signs (or words) that occur in a dataset. Typically, statistical recognition systems are limited to recognize a specific set of words: the vocabulary..

References

[Adaloglou et al., 2020] Adaloglou, N., Chatzis, T., Papastratis, I., Stergioulas, A., Papadopoulos, G. T., Zacharopoulou, V., Xydopoulos, G. J., Atzakas, K., Papazachariou, D., and Daras, P. (2020). A Comprehensive Study on Sign Language Recognition Methods. *arXiv:2007.12530 [cs].*

[Agarwal and Thakur, 2013] Agarwal, A. and Thakur, M. K. (2013). Sign language recognition using Microsoft Kinect. In *Proc. Int. Conf. on Contemporary Computing (IC3)*, pages 181–185.

[Albanie et al., 2020] Albanie, S., Varol, G., Momeni, L., Afouras, T., Chung, J. S., Fox, N., and Zisserman, A. (2020). BSL-1K: Scaling up co-articulated sign language recognition using mouthing cues. *arXiv:2007.12131 [cs].*

[Almeida et al., 2014] Almeida, S. G. M., Guimarães, F. G., and Ramírez, J. A. (2014). Feature extraction in Brazilian Sign Language Recognition based on phonological structure and using RGB-D sensors. *Expert Systems with Applications*, 41(16):7259–7271.

[Aran and Akarun, 2008] Aran, O. and Akarun, L. (2008). Multi-class classification strategies for Fisher scores of gesture and sign sequences. In *Proc. Int. Conf. on Pattern Recognition (ICPR)*, pages 1–4, Tampa, FL, USA.

[Aran and Akarun, 2010] Aran, O. and Akarun, L. (2010). A multi-class classification strategy for Fisher scores: Application to signer independent sign language recognition. *Pattern Recognition*, 43(5):1776–1788.

[Aran et al., 2009a] Aran, O., Ari, I., Akarun, L., Sankur, B., Benoit, A., Caplier, A., Campr, P., Carrillo, A. H., and Fanard, F. X. (2009a). SignTutor: An Interactive System for Sign Language Tutoring. *IEEE MultiMedia*, 16(1):81–93.

[Aran et al., 2009b] Aran, O., Burger, T., Caplier, A., and Akarun, L. (2009b). A belief-based sequential fusion approach for fusing manual signs and non-manual signals. *Pattern Recognition*, 42(5):812–822.

[Ariesta et al., 2018] Ariesta, M. C., Wiryana, F., and Zahra, a. A. (2018). Sentence Level Indonesian Sign Language Recognition Using 3D Convolutional Neural Network and Bidirectional Recurrent Neural Network. In *2018 Indonesian Association for Pattern Recognition International Conference (INAPR)*, pages 16–22.

[Assaleh et al., 2010] Assaleh, K., Shanableh, T., Fanaswala, M., Amin, F., and Bajaj, H. (2010). Continuous Arabic Sign Language Recognition in User Dependent Mode. *Journal of Intelligent Learning Systems and Applications*, 02:19.

[Assan and Grobel, 1997] Assan, M. and Grobel, K. (1997). Video-based sign language recognition using Hidden Markov Models. In *Gesture and Sign Language in Human-Computer Interaction*, Lecture Notes in Computer Science, pages 97–109.

[Athitsos et al., 2008] Athitsos, V., Neidle, C., Sclaroff, S., Nash, J., Stefan, A., Yuan, Q., and Thangali, A. (2008). The American Sign Language Lexicon Video Dataset. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1–8.

[Athitsos et al., 2010] Athitsos, V., Wang, H., and Stefan, A. (2010). A database-based framework for gesture recognition. *Personal and Ubiquitous Computing*, 14(6):511–526.

[Avola et al., 2019] Avola, D., Bernardi, M., Cinque, L., Foresti, G. L., and Massaroni, C. (2019). Exploiting Recurrent Neural Networks and Leap Motion Controller for the Recognition of Sign Language and Semaphoric Hand Gestures. *IEEE Transactions on Multimedia*, 21(1):234–245.

[Awad et al., 2009] Awad, G., Han, J., and Sutherland, A. (2009). Novel Boosting Framework for Subunit-Based Sign Language Recognition. In *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, pages 2729–2732.

[Barczak et al., 2011] Barczak, A. L. C., Reyes, N. H., Abastillas, M., Piccio, A., and Susnjak, T. (2011). A new 2D static hand gesture colour image dataset for asl gestures. *Research Letters in the Information and Mathematical Sciences*, 15:12–20.

[Bauer and Kraiss, 2002a] Bauer, B. and Kraiss, K.-F. (2002a). Towards an Automatic Sign Language Recognition System Using Subunits. In Wachsmuth, I. and Sowa, T., editors, *Gesture and Sign Language in Human-Computer Interaction*, volume 2298 of *Lecture Notes in Computer Science*, pages 123–173. Springer Berlin / Heidelberg.

[Bauer and Kraiss, 2002b] Bauer, B. and Kraiss, K. F. (2002b). Video-based sign recognition using self-organizing subunits. In *Proc. Int. Conf. on Pattern Recognition (ICPR)*, volume 2, pages 434–437, Quebec, Canada.

[Bauer et al., 1999] Bauer, B., Nießen, S., and Hienz, H. (1999). Towards an Automatic Sign Language Translation System. In *Proc. of the International Workshop on Physicality and Tangibility in Interaction: Towards New Paradigms for Interaction Beyond the Desktop*, Siena, Italy.

[Bilge et al., 2019] Bilge, Y. C., Ikizler-Cinbis, N., and Cinbis, R. G. (2019). Zero-Shot Sign Language Recognition: Can Textual Data Uncover Sign Languages? In *Proc. British Machine Vision Conference (BMVC)*, Cardiff, UK.

[Bowden et al., 2004] Bowden, R., Windridge, D., Kadir, T., Zisserman, A., and Brady, M. (2004). A Linguistic Feature Vector for the Visual Interpretation of Sign Language. In *European Conference on Computer Vision (ECCV)*, pages 390–401, Czech Republic, Prague.

[Brashear et al., 2006] Brashear, H., Henderson, V., Park, K.-H., Hamilton, H., Lee, S., and Starner, T. (2006). American sign language recognition in game development for deaf children. In *Proc. Int. ACM SIGACCESS Conf. on Computers and Accessibility (ASSETS)*, pages 79–86.

[Brashear et al., 2003] Brashear, H., Starner, T., Lukowicz, P., and Junker, H. (2003). Using multiple sensors for mobile sign language recognition. In *Proc. IEEE Int. Symposium on Wearable Computers*, pages 45–52, White Plains, NY, USA.

[Buehler et al., 2010] Buehler, P., Everingham, M., and Zisserman, A. (2010). Employing signed TV broadcasts for automated learning of British Sign Language. In *LREC Workshop on the Representation and Processing of Sign Languages*, pages 22–23, Valletta, Malta.

[Buehler et al., 2009] Buehler, P., Zisserman, A., and Everingham, M. (2009). Learning sign language by watching TV (using weakly aligned subtitles). In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 2961–2968, Miami, FL, USA.

[Camgoz et al., 2018] Camgoz, C., Hadfield, S., Koller, O., Ney, H., and Bowden, R. (2018). Neural Sign Language Translation. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 7784–7793, Salt Lake City, UT.

[Camgoz et al., 2017] Camgoz, N. C., Hadfield, S., Koller, O., and Bowden, R. (2017). SubUNets: End-to-end Hand Shape and Continuous Sign Language Recognition. In *Proc. IEEE Int. Conf. on Computer Vision (ICCV)*, pages 22–27, Venice, Italy.

[Camgöz et al., 2016] Camgöz, N. C., Kındıroğlu, A. A., and Akarun, L. (2016). Sign Language Recognition for Assisting the Deaf in Hospitals. In Chetouani, M., Cohn, J., and Salah, A. A., editors, *Human Behavior Understanding*, volume 9997, pages 89–101. Springer International Publishing, Cham.

[Camgoz et al., 2016] Camgoz, N. C., Kindiroglu, A. A., Karabuklu, S., Kelepir, M., Ozsoy, A. S., and Akarun, L. (2016). BosphorusSign: A Turkish Sign Language Recognition Corpus in Health and Finance Domains. In *Proc. Int. Conf. on Language Resources and Evaluation (LREC)*, pages 1383–1388.

[Camgoz et al., 2020] Camgoz, N. C., Koller, O., Hadfield, S., and Bowden, R. (2020). Sign Language Transformers: Joint End-to-end Sign Language Recognition and Translation. *arXiv:2003.13830 [cs]*.

[Caridakis et al., 2012] Caridakis, G., Karpouzis, K., Drosopoulos, A., and Kollias, S. (2012). Non parametric, self organizing, scalable modeling of spatiotemporal inputs: The sign language paradigm. *Neural Networks*, 36:157–166.

[Chai et al., 2013] Chai, X., Li, G., Lin, Y., Xu, Z., Tang, Y., Chen, X., and Zhou, M. (2013). Sign language recognition and translation with kinect. In *Proc. Int. Conf. on Automatic Face and Gesture Recognition (FG)*, volume 655, page demo.

[Chai et al., 2014] Chai, X., Wang, H., and Chen, X. (2014). The DEVISIGN Large Vocabulary of Chinese Sign Language Database and Baseline Evaluations. Technical report, Key Lab of Intelligent Information Processing of Chinese Academy of Sciences.

[Chai et al., 2015] Chai, X., Wang, H., Yin, F., and Chen, X. (2015). Communication tool for the hard of hearings: A large vocabulary sign language recognition system. In *Int. Conf. Affective Computing and Intelligent Interaction (ACII)*, pages 781–783, Xian, China.

[Charayaphan and Marble, 1992] Charayaphan, C. and Marble, A. E. (1992). Image processing system for interpreting motion in American Sign Language. *Journal of Biomedical Engineering*, 14(5):419–425.

[Cheng et al., 2015] Cheng, J., Chen, X., Liu, A., and Peng, H. (2015). A Novel Phonology- and Radical-Coded Chinese Sign Language Recognition Framework Using Accelerometer and Surface Electromyography Sensors. *Sensors*, 15(9):23303–23324.

[Cheng et al., 2020] Cheng, K. L., Yang, Z., Chen, Q., and Tai, Y.-W. (2020). Fully Convolutional Networks for Continuous Sign Language Recognition. *arXiv:2007.12402 [cs]*.

[Cooper and Bowden, 2007a] Cooper, H. and Bowden, R. (2007a). Large Lexicon Detection of Sign Language. In *Human–Computer Interaction*, Lecture Notes in Computer Science, pages 88–97. Springer, Berlin, Heidelberg.

[Cooper and Bowden, 2007b] Cooper, H. and Bowden, R. (2007b). Sign language recognition using boosted volumetric features. In *Proc. IAPR Conf. on Machine Vision Applications*, pages 359–362, Tokyo, Japan.

[Cooper and Bowden, 2009] Cooper, H. and Bowden, R. (2009). Learning signs from subtitles: A weakly supervised approach to sign language recognition. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 2568–2574, Miami, FL.

[Cooper and Bowden, 2010] Cooper, H. and Bowden, R. (2010). Sign Language Recognition using Linguistically Derived Sub-Units. In *LREC Workshop on the Representation and Processing of Sign Languages*, pages 57–61, Valletta, Malta.

[Cooper et al., 2012] Cooper, H., Ong, E.-J., Pugeault, N., and Bowden, R. (2012). Sign language recognition using sub-units. *The Journal of Machine Learning Research*, 13(1):2205–2231.

[Cooper et al., 2011] Cooper, H., Pugeault, N., and Bowden, R. (2011). Reading the signs: A video based sign dictionary. In *Proc. IEEE Int. Conf. on Computer Vision Workshops (ICCVW)*, pages 914–919.

[Cooper, 2010] Cooper, H. M. (2010). *Sign Language Recognition : Generalising to More Complex Corpora*. PhD thesis, University Of Surrey, Guildford, UK.

[Costa Filho et al., 2017] Costa Filho, C. F. F., de Souza, R. S., dos Santos, J. R., dos Santos, B. L., Costa, M. G. F., Costa Filho, C. F. F., de Souza, R. S., dos Santos, J. R., dos Santos, B. L., and Costa, M. G. F. (2017). A fully automatic method for recognizing hand configurations of Brazilian sign language. *Research on Biomedical Engineering*, 33(1):78–89.

[Cui et al., 2017] Cui, R., Liu, H., and Zhang, C. (2017). Recurrent Convolutional Neural Networks for Continuous Sign Language Recognition by Staged Optimization. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 7361–7369, Honolulu, HI, USA.

[Cui et al., 2019] Cui, R., Liu, H., and Zhang, C. (2019). A Deep Neural Framework for Continuous Sign Language Recognition by Iterative Training. *IEEE Transactions on Multimedia*, 0:1–1.

[Cui and Weng, 2000] Cui, Y. and Weng, J. (2000). Appearance-based hand sign recognition from intensity image sequences. *Computer Vision and Image Understanding (CVIU)*, 78(2):157–176.

[De Coster et al., 2020] De Coster, M., Van Herreweghe, M., and Dambre, J. (2020). Sign language recognition with transformer networks. In *12th International Conference on Language Resources and Evaluation (LREC 2020), Proceedings*, pages 6018–6024. European Language Resources Association (ELRA).

[Deng and Tsui, 2002] Deng, J. and Tsui, H. T. (2002). A Two-step Approach based on PaHMM for the Recognition of ASL. In *Asian Conf. on Computer Vision (ACCV)*, pages 126–131, Melbourne, Australia.

[Derpanis et al., 2008] Derpanis, K., Wildes, R., and Tsotsos, J. (2008). Definition and recovery of kinematic features for recognition of American sign language movements. *Image and Vision Computing*, 26(12):1650–1662.

[Dias et al., 2009] Dias, D., Madeo, R., Rocha, T., Biscaro, H., and Peres, S. (2009). Hand movement recognition for Brazilian Sign Language: A study using distance-based neural networks. In *Proc. Int. Joint Conf. on Neural Networks (IJCNN)*, pages 697–704.

[Ding and Martinez, 2009] Ding, L. and Martinez, A. M. (2009). Modelling and recognition of the linguistic components in American Sign Language. *Image and Vision Computing*, 27(12):1826–1844.

[Dreuw, 2008] Dreuw, P. (2008). Visual Modeling and Tracking Adaptation for Automatic Sign Language Recognition. In *International Computer Vision Summer School*, Sicily, Italy.

[Dreuw, 2012] Dreuw, P. (2012). *Probabilistic Sequence Models for Image Sequence Processing and Recognition*. PhD thesis, RWTH Aachen University, Computer Science Department, RWTH Aachen University, Aachen, Germany.

[Dreuw and Ney, 2008] Dreuw, P. and Ney, H. (2008). Visual Modeling and Feature Adaptation in Sign Language Recognition. In *ITG Conference on Speech Communication*, pages 1–4, Aachen, Germany.

[Dreuw et al., 2007] Dreuw, P., Rybach, D., Deselaers, T., Zahedi, M., and Ney, H. (2007). Speech Recognition Techniques for a Sign Language Recognition System. In *Proc. of the Ann. Conf. of the Int. Speech Commun. Assoc. (Interspeech)*, pages 2513–2516, Antwerp, Belgium.

[Dreuw et al., 2009] Dreuw, P., Steingrube, P., Deselaers, T., and Ney, H. (2009). Smoothed Disparity Maps for Continuous American Sign Language Recognition. In *Proc. Iberian Conf. on Pattern Recognition and Image Analysis (IbPRIA)*, LNCS, pages 24–31, Póvoa de Varzim, Portugal.

[Ebling et al., 2018] Ebling, S., Camgoz, N., Braem, P., Tissi, K., Sidler-Miserez, S., Stoll, S., Hadfield, S., Haug, T., Bowden, R., Tornay, S., Razavi, M., and Magimai-Doss, M. (2018). SMILE Swiss German Sign Language Dataset. In *Proc. Int. Conf. on Language Resources and Evaluation (LREC)*, pages 4221–4229, Myazaki, Japan.

[Elons et al., 2013] Elons, A. S., Abull-ela, M., and Tolba, M. F. (2013). A proposed PCNN features quality optimization technique for pose-invariant 3D Arabic sign language recognition. *Applied Soft Computing*, 13(4):1646–1660.

[Fagiani et al., 2012] Fagiani, M., Principi, E., Squartini, S., and Piazza, F. (2012). A New Italian Sign Language Database. In Zhang, H., Hussain, A., Liu, D., and Wang, Z., editors, *Advances in Brain Inspired Cognitive Systems*, number 7366 in Lecture Notes in Computer Science, pages 164–173. Springer Berlin Heidelberg.

[Fagiani et al., 2015] Fagiani, M., Principi, E., Squartini, S., and Piazza, F. (2015). Signer independent isolated Italian sign recognition based on hidden Markov models. *Pattern Analysis and Applications*, 18(2):385–402.

[Fang et al., 2017] Fang, B., Co, J., and Zhang, M. (2017). DeepASL: Enabling Ubiquitous and Non-Intrusive Word and Sentence-Level Sign Language Translation. In *Proceedings of SenSys 2017*, page 13, Delft, Netherlands.

[Fang et al., 2001a] Fang, G., Gao, W., Chen, X., Wang, C., and Ma, J. (2001a). Signer-independent continuous sign language recognition based on SRN/HMM. In *International Gesture Workshop*, pages 76–85, London, UK.

[Fang et al., 2001b] Fang, G., Gao, W., and Ma, J. (2001b). Signer-independent sign language recognition based on SOFM/HMM. In *Proc. IEEE Int. Conf. on Computer Vision Workshops (ICCVW)*, pages 90–95.

[Fang et al., 2007] Fang, G., Gao, W., and Zhao, D. (2007). Large-Vocabulary Continuous Sign Language Recognition Based on Transition-Movement Models. *IEEE Transactions on Systems, Man, and Cybernetics*, 37(1):1–9.

[Farag and Brock, 2019] Farag, I. and Brock, H. (2019). Learning Motion Disfluencies for Automatic Sign Language Segmentation. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7360–7364.

[Farhadi and Forsyth, 2006] Farhadi, A. and Forsyth, D. (2006). Aligning ASL for statistical translation using a discriminative word model. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 1471–1476, New York, NY, USA.

[Fels and Hinton, 1993] Fels, S. S. and Hinton, G. E. (1993). Glove-Talk: A neural network interface between a data-glove and a speech synthesizer. *IEEE Transactions on Neural Networks*, 4(1):2–8.

[Forster, 2008] Forster, J. (2008). *An Integrated Tracking And Recognition Approach For Video*. PhD thesis, RWTH Aachen University, Aachen, Germany.

[Forster et al., 2013a] Forster, J., Koller, O., Oberdörfer, C., Gweth, Y., and Ney, H. (2013a). Improving Continuous Sign Language Recognition: Speech Recognition Techniques and System Design. In *Workshop on Speech and Language Processing for Assistive Technologies*, pages 41–46, Grenoble, France.

[Forster et al., 2013b] Forster, J., Oberdörfer, C., Koller, O., and Ney, H. (2013b). Modality Combination Techniques for Continuous Sign Language Recognition. In *Proc. Iberian Conf. on Pattern Recognition and Image Analysis (IbPRIA)*, Lecture Notes in Computer Science 7887, pages 89–99, Madeira, Portugal. Springer.

[Forster et al., 2012] Forster, J., Schmidt, C., Hoyoux, T., Koller, O., Zelle, U., Piater, J., and Ney, H. (2012). RWTH-PHOENIX-Weather: A Large Vocabulary Sign Language Recognition and Translation Corpus. In *Proc. Int. Conf. on Language Resources and Evaluation (LREC)*, pages 3785–3789, Istanbul, Turkey.

[Forster et al., 2014] Forster, J., Schmidt, C., Koller, O., Bellgardt, M., and Ney, H. (2014). Extensions of the Sign Language Recognition and Translation Corpus RWTH-PHOENIX-Weather. In *Proc. Int. Conf. on Language Resources and Evaluation (LREC)*, pages 1911–1916, Reykjavik, Island.

[Fujimura and Xia Liu, 2006] Fujimura, K. and Xia Liu (2006). Sign recognition using depth image streams. In *Proc. Int. Conf. on Automatic Face and Gesture Recognition (FG)*, pages 381–386.

[Gao et al., 2004a] Gao, W., Fang, G., Zhao, D., and Chen, Y. (2004a). A Chinese sign language recognition system based on SOFM/SRN/HMM. *Pattern Recognition*, 37(12):2389–2402.

[Gao et al., 2004b] Gao, W., Fang, G., Zhao, D., and Chen, Y. (2004b). Transition movement models for large vocabulary continuous sign language recognition. In *Proc. Int. Conf. on Automatic Face and Gesture Recognition (FG)*, pages 553–558.

[García-Bautista et al., 2017] García-Bautista, G., Trujillo-Romero, F., and Caballero-Morales, S. O. (2017). Mexican sign language recognition using kinect and data time warping algorithm. In *2017 International Conference on Electronics, Communications and Computers (CONIELECOMP)*, pages 1–5.

[Geng et al., 2014] Geng, L., Ma, X., Wang, H., Gu, J., and Li, Y. (2014). Chinese sign language recognition with 3D hand motion trajectories and depth images. In *Proceeding of the 11th World Congress on Intelligent Control and Automation*, pages 1457–1461.

[Grimes, 1983] Grimes, G. J. (1983). Digital data entry glove interface device.

[Gruber et al., 2018] Gruber, I., Ryumin, D., Hrúz, M., and Karpov, A. (2018). Sign Language Numeral Gestures Recognition Using Convolutional Neural Network. In Ronzhin, A., Rigoll, G., and Meshcheryakov, R., editors, *Interactive Collaborative Robotics*, Lecture Notes in Computer Science, pages 70–77, Leipzig, Germany.

[Gunawan et al., 2018] Gunawan, a. H., Thiracitta, N., and Nugroho, A. (2018). Sign Language Recognition Using Modified Convolutional Neural Network Model. In *2018 Indonesian Association for Pattern Recognition International Conference (INAPR)*, pages 1–5.

[Guo et al., 2019] Guo, D., Wang, S., Tian, Q., and Wang, M. (2019). Dense temporal convolution network for sign language translation. In *Proc. of the AAAI Conf. on Artificial Intelligence*, pages 744–750. AAAI Press.

[Guo et al., 2017] Guo, D., Zhou, W., Li, H., and Wang, M. (2017). Online Early-Late Fusion Based on Adaptive HMM for Sign Language Recognition. *ACM Trans. Multimedia Comput. Commun. Appl.*, 14(1):8:1–8:18.

[Guo et al., 2018] Guo, D., Zhou, W., Li, H., and Wang, M. (2018). Hierarchical LSTM for Sign Language Translation. In *Proc. of the AAAI Conf. on Artificial Intelligence*, pages 6845–6852, New Orleans, Louisiana, USA.

[Gweth et al., 2012] Gweth, Y., Plahl, C., and Ney, H. (2012). Enhanced Continuous Sign Language Recognition using PCA and Neural Network Features. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 55–60, Providence, Rhode Island, USA.

[Han et al., 2009] Han, J., Awad, G., and Sutherland, A. (2009). Modelling and segmenting subunits for sign language recognition based on hand motion analysis. *Pattern Recognition Letters*, 30(6):623–633.

[Han et al., 2013] Han, J., Awad, G., and Sutherland, A. (2013). Boosted subunits: A framework for recognising sign language from videos. *IET Image Processing*, 7(1):70–80.

[Handhika et al., 2018] Handhika, T., Zen, R. I. M., Murni, Lestari, D. P., and Sari, I. (2018). Gesture recognition for Indonesian Sign Language (BISINDO). *Journal of Physics: Conference Series*, 1028:012173.

[Hashim and Alizadeh, 2018] Hashim, A. D. and Alizadeh, F. (2018). Kurdish Sign Language Recognition System. *UKH Journal of Science and Engineering*, 2(1):1–6.

[Hassan et al., 2019] Hassan, M., Assaleh, K., and Shanableh, T. (2019). Multiple Proposals for Continuous Arabic Sign Language Recognition. *Sensing and Imaging*, 20(1):4.

[Hernandez-Rebollar et al., 2004] Hernandez-Rebollar, J. L., Kyriakopoulos, N., and Lindeman, R. W. (2004). A New Instrumented Approach for Translating American Sign Language into Sound and Text. In *Proc. Int. Conf. on Automatic Face and Gesture Recognition (FG)*, pages 547–552.

[Holden and Owens, 2000] Holden, E.-J. and Owens, R. (2000). Visual Sign Language Recognition. In Klette, R., Gimel'farb, G., and Huang, T., editors, *Multi-Image Analysis*, Lecture Notes in Computer Science, pages 270–287, Dagstuhl Castle, Germany.

[Hrúz et al., 2009] Hrúz, M., Campr, P., and Karpov, A. (2009). Input and output modalities used in a sign-language-enabled information kiosk. In *Proc. of SPECOM*, volume 1, pages 113–116.

[Hu et al., 2017] Hu, Y., Zhao, H.-F., and Wang, Z.-G. (2017). Sign Language Fingerspelling Recognition Using Depth Information and Deep Belief Networks. *International Journal of Pattern Recognition and Artificial Intelligence*, 32(06):1850018.

[Huang and Huang, 1998] Huang, C.-L. and Huang, W.-Y. (1998). Sign language recognition using model-based tracking and a 3D Hopfield neural network. *Machine Vision and Applications*, 10(5):292–307.

[Huang et al., 2015] Huang, J., Zhou, W., Li, H., and Li, W. (2015). Sign Language Recognition using 3D convolutional neural networks. In *IEEE Int. Conf. on Multimedia and Expo (ICME)*, pages 1–6.

[Huang et al., 2018a] Huang, J., Zhou, W., Li, H., and Li, W. (2018a). Attention based 3D-CNNs for Large-Vocabulary Sign Language Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, early access:1–1.

[Huang et al., 2018b] Huang, J., Zhou, W., Zhang, Q., Li, H., and Li, W. (2018b). Video-based sign language recognition without temporal segmentation. In *Proc. of the AAAI Conf. on Artificial Intelligence*, pages 2257–2264, New Orleans, Louisiana, USA.

[Huang et al., 2018c] Huang, S., Mao, C., Tao, J., and Ye, Z. (2018c). A Novel Chinese Sign Language Recognition Method Based on Keyframe-Centered Clips. *IEEE Signal Processing Letters*, 25(3):442–446.

[Igari and Fukumura, 2014] Igari, S. and Fukumura, N. (2014). Sign language word recognition using via-point information and correlation of they bimanual movements. In *Proc. Int. Conf. of Advanced Informatics: Concept, Theory and Application (ICAICTA)*, pages 75–80.

[Imagawa et al., 2000] Imagawa, I., Matsuo, H., Taniguchi, R., Arita, D., Shan Lu, and Igi, S. (2000). Recognition of local features for camera-based sign language recognition system. In *Proc. Int. Conf. on Pattern Recognition (ICPR)*, volume 4, pages 849–853 vol.4.

[Infantino et al., 2007] Infantino, I., Rizzo, R., and Gaglio, S. (2007). A Framework for Sign Language Sentence Recognition by Commonsense Context. *IEEE Transactions on Systems, Man, and Cybernetics*, 37(5):1034–1039.

[Izutov, 2020] Izutov, E. (2020). ASL Recognition with Metric-Learning based Lightweight Network. *arXiv:2004.05054 [cs]*.

[Ji et al., 2017] Ji, Y., Kim, S., and Lee, K. B. (2017). Sign Language Learning System with Image Sampling and Convolutional Neural Network. In *2017 First IEEE International Conference on Robotic Computing (IRC)*, pages 371–375.

[Jose and Julian, 2019] Jose, H. and Julian, A. (2019). Tamil Sign Language Translator—An Assistive System for Hearing- and Speech-Impaired People. In Satapathy, S. C. and Joshi, A., editors, *Information and Communication Technology for Intelligent Systems*, Smart Innovation, Systems and Technologies, pages 249–257. Springer Singapore.

[Kadir et al., 2004] Kadir, T., Bowden, R., Ong, E.-J., and Zisserman, A. (2004). Minimal Training, Large Lexicon, Unconstrained Sign Language Recognition. In *Proc. British Machine Vision Conference (BMVC)*, pages 1–10, London, UK.

[Kadous, 1996] Kadous, M. W. (1996). Machine recognition of Auslan signs using PowerGloves: Towards large-lexicon recognition of sign language. In *Proc. of the Workshop on the Integration of Gesture in Language and Speech*, pages 165–174, Wilmington, DE, USA.

[Kadous and Taylor, 1995] Kadous, W. and Taylor, S. A. (1995). *GRASP: Recognition of Australian Sign Language Using Instrumented Gloves*. PhD thesis, University of New South Wales, Sydney, Australia.

[Kapuscinski and Wysocki, 2005] Kapuscinski, T. and Wysocki, M. (2005). Recognition of Isolated Words of the Polish Sign Language. In *Computer Recognition Systems*, Advances in Soft Computing, pages 697–704. Springer, Berlin, Heidelberg.

[Kelly et al., 2009a] Kelly, D., McDonald, J., and Markham, C. (2009a). Recognizing Spatiotemporal Gestures and Movement Epenthesis in Sign Language. In *Proc. Int. Conf. on Machine Vision and Image Processing (IMVIP)*, pages 145–150. IEEE.

[Kelly et al., 2011] Kelly, D., McDonald, J., and Markham, C. (2011). Weakly Supervised Training of a Sign Language Recognition System Using Multiple Instance Learning Density Matrices. *IEEE Transactions on Systems, Man, and Cybernetics*, 41(2):526–541.

[Kelly et al., 2009b] Kelly, D., Reilly Delannoy, J., Mc Donald, J., and Markham, C. (2009b). A Framework for Continuous Multimodal Sign Language Recognition. In *Proc. Int. Conf. on Multimodal Interfaces (ICMI)*, pages 351–358, Cambridge, MA, USA.

[Kim et al., 2008] Kim, J., Wagner, J., Rehm, M., and Andre, E. (2008). Bi-channel sensor fusion for automatic sign language recognition. In *Proc. Int. Conf. on Automatic Face and Gesture Recognition (FG)*, pages 1–6.

[Kim et al., 1996] Kim, J.-S., Jang, W., and Bien, Z. (1996). A dynamic gesture recognition system for the Korean sign language (KSL). *IEEE Transactions on Systems, Man, and Cybernetics*, 26(2):354–359.

[Kındıroğlu et al., 2019] Kındıroğlu, A. A., Özdemir, O., and Akarun, L. (2019). Temporal Accumulative Features for Sign Language Recognition. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 1288–1297.

[Kindiroglu et al., 2012] Kindiroglu, A. A., Yalcin, H., Aran, O., Hrúz, M., Campr, P., Akarun, L., and Karpov, A. (2012). Automatic recognition fingerspelling gestures in multiple languages for a communication interface for the disabled. *Pattern Recognition and Image Analysis*, 22(4):527–536.

[Kishore et al., 2018] Kishore, P. V. V., Kumar, D. A., Sastry, A. S. C. S., and Kumar, E. K. (2018). Motionlets Matching With Adaptive Kernels for 3-D Indian Sign Language Recognition. *IEEE Sensors Journal*, 18(8):3327–3337.

[Kishore and Kumar, 2012] Kishore, P. V. V. and Kumar, P. R. (2012). A video based Indian Sign Language Recognition System (INSLR) using wavelet transform and fuzzy logic. *Int. Journal of Engineering and Technology*, 4(5):537.

[Kobayashi and Haruyama, 1997] Kobayashi, T. and Haruyama, S. (1997). Partly-hidden Markov model and its application to gesture recognition. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, volume 4, pages 3081 –3084 vol.4.

[Koller et al., 2019] Koller, O., Camgoz, N. C., Ney, H., and Bowden, R. (2019). Weakly Supervised Learning with Multi-Stream CNN-LSTM-HMMs to Discover Sequential Parallelism in Sign Language Videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, accepted for publication:15.

[Koller et al., 2015] Koller, O., Forster, J., and Ney, H. (2015). Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. *Computer Vision and Image Understanding (CVIU)*, 141:108–125.

[Koller et al., 2016a] Koller, O., Ney, H., and Bowden, R. (2016a). Automatic Alignment of HamNoSys Subunits for Continuous Sign Language Recognition. In *LREC Workshop on the Representation and Processing of Sign Languages*, pages 121–128, Portorož, Slovenia.

[Koller et al., 2016b] Koller, O., Ney, H., and Bowden, R. (2016b). Deep Hand: How to Train a CNN on 1 Million Hand Images When Your Data Is Continuous and Weakly Labelled. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 3793–3802, Las Vegas, NV, USA.

[Koller et al., 2017] Koller, O., Zargaran, S., and Ney, H. (2017). Re-Sign: Re-Aligned End-To-End Sequence Modelling With Deep Recurrent CNN-HMMs. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 4297–4305, Honolulu, HI, USA.

[Koller et al., 2016c] Koller, O., Zargaran, S., Ney, H., and Bowden, R. (2016c). Deep Sign: Hybrid CNN-HMM for Continuous Sign Language Recognition. In *Proc. British Machine Vision Conference (BMVC)*, pages 1–12, York, UK.

[Koller et al., 2018] Koller, O., Zargaran, S., Ney, H., and Bowden, R. (2018). Deep Sign: Enabling Robust Statistical Continuous Sign Language Recognition via Hybrid CNN-HMMs. *International Journal of Computer Vision (IJCV)*, 126(12):1311–1325.

[Kong, 2011] Kong, W. W. (2011). *Towards Subject Independent Sign Language Recognition: A Segment-Based Probabilistic Approach*. PhD thesis, National University of Singapore, Singapore.

[Kong and Ranganath, 2010] Kong, W. W. and Ranganath, S. (2010). Sign Language Phoneme Transcription with Rule-based Hand Trajectory Segmentation. *Journal of Signal Processing Systems*, 59(2):211–222.

[Kong and Ranganath, 2014] Kong, W. W. and Ranganath, S. (2014). Towards Subject Independent Continuous Sign Language Recognition: A Segment and Merge Approach. *Pattern Recognition*, 47(3):1294–1308.

[Konstantinidis et al., 2018a] Konstantinidis, D., Dimitropoulos, K., and Daras, P. (2018a). A Deep Learning Approach for Analyzing Video and Skeletal Features in Sign Language Recognition. In *2018 IEEE International Conference on Imaging Systems and Techniques (IST)*, pages 1–6.

[Konstantinidis et al., 2018b] Konstantinidis, D., Dimitropoulos, K., and Daras, P. (2018b). Sign Language Recognition Based on Hand and Body Skeletal Data. In *2018 - 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, pages 1–4.

[Kosmidou et al., 2011] Kosmidou, V., Petrantonakis, P., and Hadjileontiadis, L. (2011). Enhanced Sign Language Recognition Using Weighted Intrinsic-Mode Entropy and Signer's Level of Deafness. *IEEE Transactions on Systems, Man, and Cybernetics*, 41(6):1531–1543.

[Kumar et al., 2019] Kumar, D. A., Sastry, A. S. C. S., Kishore, P. V. V., Kumar, E. K., and Kumar, M. T. K. (2019). S3DRGF: Spatial 3-D Relational Geometric Features for 3-D Sign Language Representation and Recognition. *IEEE Signal Processing Letters*, 26(1):169–173.

[Kumar et al., 2018a] Kumar, E. K., Kishore, P. V. V., Kumar, M. T. K., Kumar, D. A., and Sastry, A. S. C. S. (2018a). Three-Dimensional Sign Language Recognition With Angular Velocity Maps and Connived Feature ResNet. *IEEE Signal Processing Letters*, 25(12):1860–1864.

[Kumar et al., 2018b] Kumar, E. K., Kishore, P. V. V., Sastry, A. S. C. S., Kumar, M. T. K., and Kumar, D. A. (2018b). Training CNNs for 3-D Sign Language Recognition With Color Texture Coded Joint Angular Displacement Maps. *IEEE Signal Processing Letters*, 25(5):645–649.

[Kumar et al., 2018c] Kumar, P., Saini, R., Roy, P. P., and Dogra, D. P. (2018c). A position and rotation invariant framework for sign language recognition (SLR) using Kinect. *Multimedia Tools and Applications*, 77(7):8823–8846.

[Kumar et al., 2018d] Kumar, S. S., Wangyal, T., Saboo, V., and Srinath, R. (2018d). Time Series Neural Networks for Real Time Sign Language Translation. In *Proc. IEEE Int. Conf. on Machine Learning and Applications (ICMLA)*, pages 243–248.

[Kuznetsova et al., 2013] Kuznetsova, A., Leal-Taixe, L., and Rosenhahn, B. (2013). Real-Time Sign Language Recognition Using a Consumer Depth Camera. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 83–90, Sydney, Australia.

[Lang et al., 2012] Lang, S., Block, M., and Rojas, R. (2012). Sign Language Recognition Using Kinect. In *Artificial Intelligence and Soft Computing*, Lecture Notes in Computer Science, pages 394–402. Springer, Berlin, Heidelberg.

[Latif et al., 2019] Latif, G., Mohammad, N., Alghazo, J., AlKhalaf, R., and AlKhalaf, R. (2019). ArASL: Arabic Alphabets Sign Language Dataset. *Data in Brief*, 23:103777.

[Lee and Lee, 2018] Lee, B. G. and Lee, S. M. (2018). Smart Wearable Hand Device for Sign Language Interpretation System With Sensors Fusion. *IEEE Sensors Journal*, 18(3):1224–1232.

[Lee et al., 1997] Lee, C.-S., Bien, Z., Park, G.-T., Jang, W., Kim, J.-S., and Kim, S.-K. (1997). Real-time recognition system of Korean sign language based on elementary components. In *Proc. Int. Fuzzy Systems Conference*, volume 3, pages 1463–1468.

[Li et al., 2020a] Li, D., Opazo, C. R., Yu, X., and Li, H. (2020a). Word-level Deep Sign Language Recognition from Video: A New Large-scale Dataset and Methods Comparison. *arXiv:1910.11006 [cs]*.

[Li et al., 2020b] Li, D., Yu, X., Xu, C., Petersson, L., and Li, H. (2020b). Transferring Cross-Domain Knowledge for Video Sign Language Recognition. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 6205–6214.

[Li et al., 2016] Li, K., Zhou, Z., and Lee, C.-H. (2016). Sign Transition Modeling and a Scalable Solution to Continuous Sign Language Recognition for Real-World Applications. *ACM Transactions on Accessible Computing (TACCESS)*, 8(2):7:1–7:23.

[Li et al., 2017] Li, X., Mao, C., Huang, S., and Ye, Z. (2017). Chinese Sign Language Recognition Based on SHS Descriptor and Encoder-Decoder LSTM Model. In *Biometric Recognition*, Lecture Notes in Computer Science, pages 719–728.

[Liang and Ouhyoung, 1995] Liang, R.-H. and Ouhyoung, M. (1995). A Real-time Continuous Alphabetic Sign Language to Speech Conversion VR System. In *Computer Graphics Forum*, volume 14, pages 67–76. Wiley Online Library.

[Liang and Ouhyoung, 1998] Liang, R.-H. and Ouhyoung, M. (1998). A Real-Time Continuous Gesture Recognition System for Sign Language. In *Proc. Int. Conf. on Automatic Face and Gesture Recognition (FG)*, pages 558–567, Nara, Japan.

[Liao et al., 2019] Liao, Y., Xiong, P., Min, W., Min, W., and Lu, J. (2019). Dynamic Sign Language Recognition Based on Video Sequence With BLSTM-3D Residual Networks. *IEEE Access*, 7:38044–38054.

[Lichtenauer et al., 2008] Lichtenauer, J. F., Hendriks, E. A., and Reinders, M. J. T. (2008). Sign Language Recognition by Combining Statistical DTW and Independent Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(11):2040–2046.

[Lichtenauer et al., 2009] Lichtenauer, J. F., ten Holt, G. A., Reinders, M. J. T., and Hendriks, E. A. (2009). Person-Independent 3D Sign Language Recognition. In Sales Dias, M., Gibet, S., Wanderley, M. M., and Bastos, R., editors, *Gesture-Based Human-Computer Interaction and Simulation*, Lecture Notes in Computer Science, pages 69–80, Lisbon, Portugal.

[Lim et al., 2016] Lim, K. M., Tan, A. W. C., and Tan, S. C. (2016). A feature covariance matrix with serial particle filter for isolated sign language recognition. *Expert Systems with Applications*, 54:208–218.

[Liu et al., 2016] Liu, T., Zhou, W., and Li, H. (2016). Sign Language Recognition with Long Short-Term Memory. In *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, pages 2871–2875.

[Liu et al., 2018] Liu, Z., Qi, X., and Pang, L. (2018). Self-boosted Gesture Interactive System with ST-Net. In *Proceedings of the 26th ACM International Conference on Multimedia*, MM '18, pages 145–153, Seoul, Republic of Korea. Association for Computing Machinery.

[Liwicki and Everingham, 2009] Liwicki, S. and Everingham, M. (2009). Automatic recognition of fingerspelled words in British Sign Language. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 50–57.

[Ma et al., 2000] Ma, J., Gao, W., and Wang, R. (2000). A Parallel Multistream Model for Integration of Sign Language Recognition and Lip Motion. In Tan, T., Shi, Y., and Gao, W., editors, *Advances in Multimodal Interfaces — ICMI 2000*, number 1948 in Lecture Notes in Computer Science, pages 582–589. Springer Berlin Heidelberg.

[Ma et al., 2018] Ma, Y., Zhou, G., Wang, S., Zhao, H., and Jung, W. (2018). SignFi: Sign Language Recognition Using WiFi. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(1):23:1–23:21.

[Maebatake et al., 2008] Maebatake, M., Suzuki, I., Nishida, M., Horiuchi, Y., and Kuroiwa, S. (2008). Sign Language Recognition Based on Position and Movement Using Multi-Stream HMM. In *2008 Second International Symposium on Universal Communication*, pages 478–481, Osaka, Japan.

[Maraqa and Abu-Zaiter, 2008] Maraqa, M. and Abu-Zaiter, R. (2008). Recognition of Arabic Sign Language (ArSL) using recurrent neural networks. In *Proc. Int. Conf. on Applications of Digital Information and Web Technologies (ICADIWT)*, pages 478–481.

[Mathur and Sharma, 2018] Mathur, S. and Sharma, P. (2018). Sign Language Gesture Recognition using Zernike Moments and DTW. In *2018 5th International Conference on Signal Processing and Integrated Networks (SPIN)*, pages 586–591.

[Matsuo et al., 1997] Matsuo, H., Igi, S., Lu, S., Nagashima, Y., Takata, Y., and Teshima, T. (1997). The recognition algorithm with non-contact for Japanese sign language using morphological analysis. In Wachsmuth, I. and Fröhlich, M., editors, *Gesture and Sign Language in Human-Computer Interaction*, Lecture Notes in Computer Science, pages 273–284, Bielefeld, Germany.

[McGuire et al., 2004] McGuire, R. M., Hernandez-Rebollar, J., Starner, T., Henderson, V., Brashear, H., and Ross, D. S. (2004). Towards a one-way American sign language translator. In *Proc. Int. Conf. on Automatic Face and Gesture Recognition (FG)*, pages 620–625.

[Mekala et al., 2011] Mekala, P., Gao, Y., Fan, J., and Davari, A. (2011). Real-time sign language recognition based on neural network architecture. In *IEEE Southeastern Symposium on System Theory (SSST)*, pages 195–199, Auburn, AL, USA.

[Mittal et al., 2019] Mittal, A., Kumar, P., Roy, P. P., Balasubramanian, R., and Chaudhuri, B. B. (2019). A Modified-LSTM Model for Continuous Sign Language Recognition using Leap motion. *IEEE Sensors Journal*, early access:1–1.

[Mohandes et al., 2014] Mohandes, M., Aliyu, S., and Deriche, M. (2014). Arabic sign language recognition using the leap motion controller. In *Proc. IEEE Int. Symposium on Industrial Electronics (ISIE)*, pages 960–965.

[Mohandes and Deriche, 2013] Mohandes, M. and Deriche, M. (2013). Arabic sign language recognition by decisions fusion using Dempster-Shafer theory of evidence. In *Computing, Communications and IT Applications Conf. (ComComAp)*, pages 90–94.

[Mohandes et al., 2012] Mohandes, M., Deriche, M., Johar, U., and Ilyas, S. (2012). A signer-independent Arabic Sign Language recognition system using face detection, geometric features, and a Hidden Markov Model. *Computers and Electrical Engineering*, 38(2):422–433.

[Mohandes et al., 2007] Mohandes, M., Quadri, S. I., and Deriche, M. (2007). Arabic Sign Language Recognition an Image-Based Approach. In *Proc. Int. Conf. on Advanced Information Networking and Applications Workshops (AINAW)*, volume 1, pages 272–276.

[Mohandes, 2013] Mohandes, M. A. (2013). Recognition of Two-Handed Arabic Signs Using the CyberGlove. *Arabian Journal for Science and Engineering*, 38(3):669–677.

[Mukushev et al., 2020] Mukushev, M., Sabyrov, A., Imashev, A., Koishybay, K., Kimmelman, V., and Sandygulova, A. (2020). Evaluation of Manual and Non-manual Components for Sign Language Recognition. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 6073–6078, Marseille, France. European Language Resources Association.

[Murakami and Taguchi, 1991] Murakami, K. and Taguchi, H. (1991). Gesture recognition using recurrent neural networks. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pages 237–242. ACM.

[Nagendraswamy et al., 2015] Nagendraswamy, H. S., Guru, D. S., and Naresh, Y. G. (2015). Symbolic Representation of Sign Language at Sentence Level. *International Journal of Image, Graphics and Signal Processing(IJIGSP)*, 7(9):49–60.

[Nagendraswamy and Kumara, 2016] Nagendraswamy, H. S. and Kumara, B. M. (2016). LBPV for Recognition of Sign Language at Sentence Level: An Approach Based on Symbolic Representation. *Journal of Intelligent Systems*, 26(2):371–385.

[Nayak et al., 2005] Nayak, S., Sarkar, S., and Loeding, B. (2005). Unsupervised modeling of signs embedded in continuous sentences. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 81–81.

[Neidle and Vogler, 2012] Neidle, C. and Vogler, C. (2012). A new web interface to facilitate access to corpora: Development of the ASLLRP data access interface. In *Proc. Int. Conf. on Language Resources and Evaluation (LREC)*, pages 137–142, Istanbul, Turkey. ELRA.

[Neto et al., 2015] Neto, F. M. d. P., Cambuim, L. F., Macieira, R. M., Ludermir, T. B., Zanchettin, C., and Barros, E. N. (2015). Extreme Learning Machine for Real Time Recognition of Brazilian Sign Language. In *Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC)*, pages 1464–1469.

[Oberdörfer et al., 2012] Oberdörfer, C., Oberdörfer, C., Leibe, B., and Forster, J. (2012). Investigations on Decoding of Parallel Input Streams in Sign Language. Master's thesis, RWTH Aachen University, Aachen, Germany.

[O'Connor et al., 2017] O'Connor, T. F., Fach, M. E., Miller, R., Root, S. E., Mercier, P. P., and Lipomi, D. J. (2017). The Language of Glove: Wireless gesture decoder with low-power and stretchable hybrid electronics. *PLOS ONE*, 12(7):e0179766.

[Ong et al., 2012] Ong, E.-J., Cooper, H., Pugeault, N., and Bowden, R. (2012). Sign Language Recognition using Sequential Pattern Trees. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 2200–2207.

[Ong et al., 2014] Ong, E.-J., Koller, O., Pugeault, N., and Bowden, R. (2014). Sign Spotting using Hierarchical Sequential Patterns with Temporal Intervals. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 1931–1938, Columbus, OH, USA.

[Oszust and Wysocki, 2010] Oszust, M. and Wysocki, M. (2010). Determining Subunits for Sign Language Recognition by Evolutionary Cluster-Based Segmentation of Time Series. In Rutkowski, L., Scherer, R., Tadeusiewicz, R., Zadeh, L., and Zurada, J., editors, *Artifical Intelligence and Soft Computing*, volume 6114 of *Lecture Notes in Computer Science*, pages 189–196. Springer Berlin Heidelberg.

[Oszust and Wysocki, 2013a] Oszust, M. and Wysocki, M. (2013a). Polish sign language words recognition with Kinect. In *Proc. Int. Conf. on Human System Interactions (HSI)*, pages 219–226.

[Oszust and Wysocki, 2013b] Oszust, M. and Wysocki, M. (2013b). Recognition of signed expressions observed by Kinect Sensor. In *Proc. IEEE Int. Conf. on Advanced Video and Signal Based Surveillance*, pages 220–225.

[Ouhyoung and Liang, 1996] Ouhyoung, M. and Liang, R. H. (1996). A sign language recognition system using hidden markov model and context sensitive search. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, pages 59–66.

[Oz and Leu, 2005] Oz, C. and Leu, M. C. (2005). Linguistic Properties Based on American Sign Language Recognition with Artificial Neural Networks Using a Sensory Glove and Motion Tracker. In *Computational Intelligence and Bioinspired Systems*, Lecture Notes in Computer Science, pages 1197–1205. Springer, Berlin, Heidelberg.

[Özdemir et al., 2020] Özdemir, O., Kındıroğlu, A. A., Camgöz, N. C., and Akarun, L. (2020). BosphorusSign22k Sign Language Recognition Dataset. *arXiv:2004.01283 [cs]*.

[Papastratis et al., 2020] Papastratis, I., Dimitropoulos, K., Konstantinidis, D., and Daras, P. (2020). Continuous Sign Language Recognition Through Cross-Modal Alignment of Video and Text Embeddings in a Joint-Latent Space. *IEEE Access*, 8:91170–91180.

[Paudyal et al., 2019] Paudyal, P., Lee, J., Kamzin, A., Soudki, M., Banerjee, A., and Gupta, S. (2019). Learn2Sign: Explainable AI for sign language learning. In *CEUR Workshop Proceedings*, volume 2327.

[Paulraj et al., 2010] Paulraj, M. P., Yaacob, S., Azalan, B. Z., Shuhanaz, M., and Palaniappan, R. (2010). A phoneme based sign language recognition system using skin color segmentation. In *Int. Colloquium on Signal Processing and Its Applications (CSPA)*, pages 1–5, Malacca City, Malaysia.

[Paulraj et al., 2008] Paulraj, M. P., Yaacob, S., Desa, H., Hema, C. R., Ridzuan, W. M., and Majid, W. A. (2008). Extraction of head and hand gesture features for recognition of sign language. In *Proc. Int. Conf. on Electronic Design*, pages 1–6.

[Pei et al., 2019] Pei, X., Guo, D., and Zhao, Y. (2019). Continuous Sign Language Recognition Based on Pseudo-supervised Learning. In *Proceedings of the 2Nd Workshop on Multimedia for Accessible Human Computer Interfaces*, MAHCI '19, pages 33–39, New York, NY, USA. ACM.

[Pezzuoli et al., 2017] Pezzuoli, F., Corona, D., Corradini, M. L., and Cristofaro, A. (2017). Development of a Wearable Device for Sign Language Translation. In Ficuciello, F., Ruggiero, F., and Finzi, A., editors, *Human Friendly Robotics*, Springer Proceedings in Advanced Robotics, pages 115–126.

[Pfau and Quer, 2010] Pfau, R. and Quer, J. (2010). Nonmanuals: Their grammatical and prosodic roles. In Brentari, D., editor, *Sign Languages*, pages 381–402. Cambridge University Press, Cambrigde, UK.

[Pigou et al., 2016] Pigou, L., Van Herreweghe, M., and Dambre, J. (2016). Sign classification in sign language Corpora with deep neural networks. In *International Conference on Language Resources and Evaluation (LREC), Workshop, Proceedings*, pages 175–178.

[Pigou et al., 2017] Pigou, L., Van Herreweghe, M., and Dambre, J. (2017). Gesture and Sign Language Recognition With Temporal Residual Networks. In *Proc. IEEE Int. Conf. on Computer Vision Workshops (ICCVW)*, pages 3086–3093, Venice, Italy.

[Pitsikalis et al., 2010] Pitsikalis, V., Theodorakis, S., and Maragos, P. (2010). Data-Driven Sub-Units and Modeling Structure for Continuous Sign Language Recognition with Multiple Cues. In *LREC Workshop on the Representation and Processing of Sign Languages*, pages 196–203.

[Pitsikalis et al., 2011] Pitsikalis, V., Theodorakis, S., Vogler, C., and Maragos, P. (2011). Advances in phonetics-based sub-unit modeling for transcription alignment and sign language recognition. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1–6.

[Pu et al., 2016a] Pu, J., Zhou, W., and Li, H. (2016a). Sign Language Recognition with Multi-modal Features. In *Proc. Advances in Multimedia Information Processing*, pages 252–261.

[Pu et al., 2018] Pu, J., Zhou, W., and Li, H. (2018). Dilated Convolutional Network with Iterative Optimization for Continuous Sign Language Recognition. In *IJCAI*, pages 885–891, Stockholm, Sweden.

[Pu et al., 2019] Pu, J., Zhou, W., and Li, H. (2019). Iterative Alignment Network for Continuous Sign Language Recognition. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, page 10.

[Pu et al., 2016b] Pu, J., Zhou, W., Zhang, J., and Li, H. (2016b). Sign Language Recognition Based on Trajectory Modeling with HMMs. In *Int. Conf. on Multimedia Modeling*, Lecture Notes in Computer Science, pages 686–697.

[Pugeault and Bowden, 2011] Pugeault, N. and Bowden, R. (2011). Spelling It Out: Real–Time ASL Fingerspelling Recognition. In *Proc. IEEE Int. Conf. on Computer Vision Workshops (ICCVW)*, pages 1114–1119, Barcelona, Spain.

[Quiroga et al., 2017] Quiroga, F., Antonio, R., Ronchetti, F., Lanzarini, L. C., and Rosete, A. (2017). A Study of Convolutional Architectures for Handshape Recognition applied to Sign Language. In *XXIII Congreso Argentino de Ciencias de La Computación*, pages 13–22, La Plata, Argentinia.

[Rakowski and Wandzik, 2018] Rakowski, A. and Wandzik, L. (2018). Hand Shape Recognition Using Very Deep Convolutional Neural Networks. In *Proceedings of the 2018 International Conference on Control and Computer Vision*, ICCCV '18, pages 8–12, Singapore, Singapore. ACM.

[Rao and Kishore, 2018] Rao, G. A. and Kishore, P. V. V. (2018). Selfie video based continuous Indian sign language recognition system. *Ain Shams Engineering Journal*, 9(4):1929–1939.

[Rao et al., 2018] Rao, G. A., Syamala, K., Kishore, P. V. V., and Sastry, A. S. C. S. (2018). Deep convolutional neural networks for sign language recognition. In *Proc. Conf. on Signal Processing And Communication Engineering Systems (SPACES)*, pages 194–197.

[Rekha et al., 2011] Rekha, J., Bhattacharya, J., and Majumder, S. (2011). Hand Gesture Recognition for Sign Language: A New Hybrid Approach. In *Int. Conf. Image Processing, Computer Vision, and Pattern Recognition*, pages 80–86, Las Vegas, NV, USA.

[Ronchetti et al., 2016] Ronchetti, F., Quiroga, F., Lanzarini, L., and Estrebou, C. (2016). Handshape Recognition for Argentinian Sign Language using ProbSom. *Journal of Computer Science and Technology*, 16(1):1–5.

[Roussos et al., 2013] Roussos, A., Theodorakis, S., Pitsikalis, V., and Maragos, P. (2013). Dynamic affine-invariant shape-appearance handshape features and classification in sign language videos. *The Journal of Machine Learning Research*, 14(1):1627–1663.

[Rybach, 2006] Rybach, D. (2006). *Appearance-Based Features for Automatic Continuous Sign Language Recognition*. PhD thesis, Human Language Technology and Pattern Recognition Group, RWTH Aachen University, Aachen, Germany.

[Sabyrov et al., 2019] Sabyrov, A., Mukushev, M., and Kimmelman, V. (2019). Towards Real-time Sign Language Interpreting Robot: Evaluation of Non-manual Components on Recognition Accuracy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 75–82.

[Sagawa and Takeuchi, 2000] Sagawa, H. and Takeuchi, M. (2000). A method for recognizing a sequence of sign language words represented in a Japanese sign language sentence. In *Proc. Int. Conf. on Automatic Face and Gesture Recognition (FG)*, pages 434–439, Grenoble, France.

[Santemiz et al., 2009] Santemiz, P., Aran, O., Saraclar, M., and Akarun, L. (2009). Automatic sign segmentation from continuous signing via multiple sequence alignment. In *Proc. IEEE Int. Conf. on Computer Vision Workshops (ICCVW)*, pages 2001–2008. IEEE.

[Sarkar et al., 2011] Sarkar, S., Loeding, B., Yang, R., Nayak, S., and Parashar, A. (2011). Segmentation-robust representations, matching, and modeling for sign language. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 13–19.

[Shanableh and Assaleh, 2011] Shanableh, T. and Assaleh, K. (2011). User-independent recognition of Arabic sign language for facilitating communication with the deaf community. *Digital Signal Processing*, 21(4):535–542.

[Shanableh et al., 2007] Shanableh, T., Assaleh, K., and Al-Rousan, M. (2007). Spatio-Temporal Feature-Extraction Techniques for Isolated Gesture Recognition in Arabic Sign Language. *IEEE Transactions on Systems, Man, and Cybernetics*, 37(3):641–650.

[Shenoy et al., 2018] Shenoy, K., Dastane, T., Rao, V., and Vyavaharkar, D. (2018). Real-time Indian Sign Language (ISL) Recognition. In *Proc. Int. Conf. on Computing, Communication and Networking Technologies (ICCCNT)*, pages 1–9.

[Song et al., 2019] Song, P., Guo, D., Xin, H., and Wang, M. (2019). Parallel Temporal Encoder For Sign Language Translation. In *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, pages 1915–1919.

[Starner and Pentland, 1995] Starner, T. and Pentland, A. (1995). Real-time american sign language recognition from video using hidden markov models. In *International Symposium on Computer Vision*, pages 265–270, Coral Gables, Florida, USA.

[Starner et al., 1998] Starner, T., Weaver, J., and Pentland, A. (1998). Real-time american sign language recognition using desk and wearable computer based video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1371–1375.

[Stein et al., 2007] Stein, D., Dreuw, P., Ney, H., Morrissey, S., and Way, A. (2007). Hand in Hand: Automatic Sign Language to Speech Translation. In *Proc. Conf. on Theoretical and Methodological Issues in Machine Translation*, pages 214–220, Sk"ovde, Sweden.

[Su et al., 2001] Su, M.-C., Zhao, Y.-X., and Chen, H.-F. (2001). A fuzzy rule-based approach to recognizing 3-D arm movements. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 9(2):191–201.

[Sun et al., 2013a] Sun, C., Zhang, T., Bao, B., Xu, C., and Mei, T. (2013a). Discriminative Exemplar Coding for Sign Language Recognition With Kinect. *IEEE Transactions on Cybernetics*, 43(5):1418–1428.

[Sun et al., 2013b] Sun, C., Zhang, T., Bao, B.-K., and Xu, C. (2013b). Latent support vector machine for sign language recognition with Kinect. In *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, pages 4190–4194.

[Tamer and Saraçlar, 2020] Tamer, N. C. and Saraçlar, M. (2020). Keyword Search for Sign Language. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8184–8188.

[Tamura and Kawasaki, 1988] Tamura, S. and Kawasaki, S. (1988). Recognition of sign language motion images. *Pattern Recognition*, 21(4):343–353.

[Tanibata et al., 2002] Tanibata, N., Shimada, N., and Shirai, Y. (2002). Extraction of Hand Features for Recognition of Sign Language Words. In *Proc. Int. Conf. on Vision Interface*, pages 391–398, Calgary, Canada.

[Thang et al., 2017] Thang, P. Q., Dung, N. D., and Thuy, N. T. (2017). A Comparison of SimpSVM and RVM for Sign Language Recognition. In *Proc. Int. Conf. on Machine Learning and Soft Computing*, ICMLSC '17, pages 98–104, Ho Chi Minh City, Vietnam.

[Thangali et al., 2011] Thangali, A., Nash, J. P., Sclaroff, S., and Neidle, C. (2011). Exploiting phonological constraints for handshape inference in ASL video. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 521–528, Colorado Springs, Colorado, USA.

[Theodorakis et al., 2009] Theodorakis, S., Katsamanis, A., and Maragos, P. (2009). Product-HMMs for automatic sign language recognition. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1601–1604, Taipei, Taiwan.

[Theodorakis et al., 2010] Theodorakis, S., Pitsikalis, V., and Maragos, P. (2010). Model-level data-driven sub-units for signs in videos of continuous sign language. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2262–2265, Dallas, Texas, USA.

[Theodorakis et al., 2014] Theodorakis, S., Pitsikalis, V., and Maragos, P. (2014). Dynamic-static unsupervised sequentiality, statistical subunits and lexicon for sign language recognition. *Image and Vision Computing*, 32(8):533–549.

[Tolba et al., 2010] Tolba, M. F., Abdellwahab, M. S., Aboul-Ela, M., and Samir, A. (2010). Image signature improving by PCNN for Arabic sign language recognition. *Can. J. Artif. Intell. Mach. Learn. Pattern Recognit*, 1(1):1–6.

[Tornay et al., 2019] Tornay, S., Razavi, M., Camgoz, N. C., Bowden, R., and Magimai. -Doss, M. (2019). HMM-based Approaches to Model Multichannel Information in Sign Language inspired from Articulatory Features-based Speech Processing. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2817–2821, Brighton, UK.

[Tran et al., 2018] Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y., and Paluri, M. (2018). A Closer Look at Spatiotemporal Convolutions for Action Recognition. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6450–6459, Salt Lake City, UT. IEEE.

[Trmal et al., 2008] Trmal, J., Hrúz, M., Zelinka, J., Campr, P., and Müller, L. (2008). Feature space transforms for czech sign-language recognition. In *Proc. of the Ann. Conf. of the Int. Speech Commun. Assoc. (Interspeech)*, pages 2036–2039, Brisbane, Australia.

[Tubaiz et al., 2015] Tubaiz, N., Shanableh, T., and Assaleh, K. (2015). Glove-based continuous Arabic sign language recognition in user-dependent mode. *IEEE Transactions on Human-Machine Systems*, 45(4):526–533.

[Uebersax et al., 2011] Uebersax, D., Gall, J., den Bergh, M. V., and Gool, L. V. (2011). Real-time sign language letter and word recognition from depth data. In *Proc. IEEE Int. Conf. on Computer Vision Workshops (ICCVW)*, pages 383–390.

[Vaezi Joze and Koller, 2019] Vaezi Joze, H. and Koller, O. (2019). MS-ASL: A large-scale data set and benchmark for understanding american sign language. In *Proc. British Machine Vision Conference (BMVC)*, Cardiff, UK.

[Vamplew and Adams, 1996] Vamplew, P. and Adams, A. (1996). Recognition of sign language gestures using neural networks. In *European Conf. on Disabilities, Virtual Reality and Associated Technologies*, pages 27–33, Maidenhead, UK.

[Vasudevan et al., 2020] Vasudevan, A., Negri, P., Linares-Barranco, B., and Serrano-Gotarredona, T. (2020). Introduction and Analysis of an Event-Based Sign Language Dataset. In *Proc. Int. Conf. on Automatic Face and Gesture Recognition (FG)*, pages 441–448, Buenos Aires, Argentina.

[Vogler and Metaxas, 1997] Vogler, C. and Metaxas, D. (1997). Adapting hidden Markov models for ASL recognition by using three-dimensional computer vision methods. In *Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC)*, pages 156–161, Orlando, USA.

[Vogler and Metaxas, 1999a] Vogler, C. and Metaxas, D. (1999a). Parallel hidden markov models for american sign language recognition. In *Proc. IEEE Int. Conf. on Computer Vision (ICCV)*, volume 1, pages 116–122, Corfu, Greece.

[Vogler and Metaxas, 1999b] Vogler, C. and Metaxas, D. (1999b). Toward Scalability in ASL Recognition: Breaking Down Signs into Phonemes. In *Gesture-Based Communication in Human-Computer Interaction*, Lecture Notes in Computer Science, pages 211–224. Springer, Berlin, Heidelberg.

[Vogler and Metaxas, 2001] Vogler, C. and Metaxas, D. (2001). A framework for recognizing the simultaneous aspects of american sign language. *Computer Vision and Image Understanding (CVIU)*, 81(3):358–384.

[Vogler and Metaxas, 2004] Vogler, C. and Metaxas, D. (2004). Handshapes and Movements: Multiple-Channel American Sign Language Recognition. In Camurri, A. and Volpe, G., editors, *Gesture-Based Communication in Human-Computer Interaction*, volume 2915 of *Lecture Notes in Computer Science*, pages 247–258. Springer, Berlin, Heidelberg.

[von Agris et al., 2008a] von Agris, U., Knorr, M., and Kraiss, K.-F. (2008a). The significance of facial features for automatic sign language recognition. In *Proc. Int. Conf. on Automatic Face and Gesture Recognition (FG)*, pages 1–6, Amsterdam, The Netherlands.

[von Agris and Kraiss, 2007] von Agris, U. and Kraiss, K.-F. (2007). Towards a Video Corpus for Signer-Independent Continuous Sign Language Recognition. In Dias, S. and Jota, editors, *GW 2007 The 7th International Workshop on Gesture in Human-Computer Interaction and Simulation*, pages 10–11, Lisbon, Portugal.

[von Agris et al., 2006] von Agris, U., Schneider, D., Zieren, J., and Kraiss, K. (2006). Rapid Signer Adaptation for Isolated Sign Language Recognition. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 159–159.

[von Agris et al., 2008b] von Agris, U., Zieren, J., Canzler, U., Bauer, B., and Kraiss, K. F. (2008b). Recent developments in visual sign language recognition. *Universal Access in the Information Society*, 6(4):323–362.

[Waldron and Kim, 1995] Waldron, M. B. and Kim, S. (1995). Isolated ASL sign recognition system for deaf persons. *IEEE Transactions on Rehabilitation Engineering*, 3(3):261–271.

[Wang et al., 2006a] Wang, C., Chen, X., and Gao, W. (2006a). Expanding training set for chinese sign language recognition. In *Proc. Int. Conf. on Automatic Face and Gesture Recognition (FG)*, pages 323–328, Southampton, UK.

[Wang et al., 2001] Wang, C., Gao, W., and Ma, J. (2001). A Real-Time Large Vocabulary Recognition System for Chinese Sign Language. In *Gesture and Sign Language in Human-Computer Interaction*, Lecture Notes in Computer Science, pages 86–95.

[Wang et al., 2002] Wang, C., Gao, W., and Shan, S. (2002). An approach based on phonemes to large vocabulary Chinese sign language recognition. In *Proc. Int. Conf. on Automatic Face and Gesture Recognition (FG)*, pages 411–416.

[Wang et al., 2019] Wang, F., Sun, S., and Liu, Y. (2019). A Bi-directional Interactive System of Sign Language and Visual Speech Based on Portable Devices. In *Proc. IEEE Int. Conf. on Robotics and Biomimetics (ROBIO)*, pages 1071–1076.

[Wang et al., 2016] Wang, H., Chai, X., and Chen, X. (2016). Sparse Observation (SO) Alignment for Sign Language Recognition. *Neurocomputing*, 175(Part A):674–685.

[Wang et al., 2015] Wang, H., Chai, X., Zhou, Y., and Chen, X. (2015). Fast sign language recognition benefited from low rank approximation. In *Proc. Int. Conf. on Automatic Face and Gesture Recognition (FG)*, volume 1, pages 1–6, Ljubljana, Slovenia. IEEE.

[Wang et al., 2006b] Wang, H., Leu, M. C., and Oz, C. (2006b). American Sign Language Recognition Using Multi-dimensional Hidden Markov Models. *Journal of Information Science and Engineering*, 22(5):1109–1123.

[Wang et al., 2010] Wang, H., Stefan, A., Moradi, S., Athitsos, V., Neidle, C., and Kamangar, F. (2010). A System for Large Vocabulary Sign Search. In *European Conference on Computer Vision Workshops (ECCVW)*, pages 342–353, Crete, Greece.

[Wang and Gao, 2000] Wang, J. and Gao, W. (2000). A Fast Sign Word Recognition Method for Chinese Sign Language. In Tan, T., Shi, Y., and Gao, W., editors, *Advances in Multimodal Interfaces — ICMI 2000*, Lecture Notes in Computer Science, pages 599–606, Beijing, China.

[Wang et al., 2007] Wang, Q., Chen, X., Zhang, L.-G., Wang, C., and Gao, W. (2007). Viewpoint invariant sign language recognition. *Computer Vision and Image Understanding (CVIU)*, 108(1–2):87–97.

[Wang et al., 2018] Wang, S., Guo, D., Zhou, W.-g., Zha, Z.-J., and Wang, M. (2018). Connectionist Temporal Fusion for Sign Language Translation. In *Proc. ACM Int. Conf. on Multimedia*, MM '18, pages 1483–1491, Seoul, Republic of Korea.

[Wei et al., 2019] Wei, C., Zhou, W., Pu, J., and Li, H. (2019). Deep Grammatical Multi-classifier for Continuous Sign Language Recognition. In *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, pages 435–442.

[Wei et al., 2016] Wei, S., Chen, X., Yang, X., Cao, S., and Zhang, X. (2016). A Component-Based Vocabulary-Extensible Sign Language Gesture Recognition Framework. *Sensors*, 16(4):556.

[Windridge and Bowden, 2004] Windridge, D. and Bowden, R. (2004). Induced Decision Fusion in Automated Sign Language Interpretation: Using ICA to Isolate the Underlying Components of Sign. In Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., Roli, F., Kittler, J., and Windeatt, T., editors, *Multiple Classifier Systems*, volume 3077, pages 303–313. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Wu and Gao, 2001] Wu, J. and Gao, W. (2001). The Recognition of Finger-Spelling for Chinese Sign Language. In Wachsmuth, I. and Sowa, T., editors, *Gesture and Sign Language in Human-Computer Interaction*, Lecture Notes in Computer Science, pages 96–100, London, UK.

[Xie et al., 2018] Xie, B., He, X., and Li, Y. (2018). RGB-D static gesture recognition based on convolutional neural network. *The Journal of Engineering*, 2018(16):1515–1520.

[Yang and Lee, 2011] Yang, H.-D. and Lee, S.-W. (2011). Combination of manual and non-manual features for sign language recognition based on conditional random field and active appearance model. In *Int. Conf. on Machine Learning and Cybernetics (ICMLC)*, volume 4, pages 1726–1731.

[Yang and Lee, 2013] Yang, H.-D. and Lee, S.-W. (2013). Robust sign language recognition by combining manual and non-manual features based on conditional random field and support vector machine. *Pattern Recognition Letters*, 34(16):2051–2056.

[Yang et al., 2009] Yang, H.-D., Sclaroff, S., and Lee, S.-W. (2009). Sign Language Spotting with a Threshold Model Based on Conditional Random Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(7):1264–1277.

[Yang et al., 2002] Yang, M.-H., Ahuja, N., and Tabb, M. (2002). Extraction of 2D motion trajectories and its application to hand gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1061–1074.

[Yang et al., 2007] Yang, R., Sarkar, S., and Loeding, B. (2007). Enhanced Level Building Algorithm for the Movement Epenthesis Problem in Sign Language Recognition. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE.

[Yang et al., 2010] Yang, R., Sarkar, S., and Loeding, B. (2010). Handling Movement Epenthesis and Hand Segmentation Ambiguities in Continuous Sign Language Recognition Using Nested Dynamic Programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3):462–477.

[Yang et al., 2016] Yang, W., Tao, J., and Ye, Z. (2016). Continuous sign language recognition using level building based on fast hidden Markov model. *Pattern Recognition Letters*, 78:28–35.

[Yang et al., 2019] Yang, Z., Shi, Z., Shen, X., and Tai, Y.-W. (2019). SF-Net: Structured Feature Network for Continuous Sign Language Recognition. *arXiv:1908.01341 [cs]*.

[Ye et al., 2018] Ye, Y., Tian, Y., Huenerfauth, M., Liu, J., Ruiz, N., Chong, E., Rehg, J. M., Palsson, S., Agustsson, E., and Timofte, R. (2018). Recognizing American Sign Language Gestures From Within Continuous Videos. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2064–2073.

[Yin et al., 2016] Yin, F., Chai, X., and Chen, X. (2016). Iterative Reference Driven Metric Learning for Signer Independent Isolated Sign Language Recognition. In *European Conference on Computer Vision (ECCV)*, pages 434–450, Amsterdam, The Netherlands.

[Yin et al., 2015a] Yin, F., Chai, X., Zhou, Y., and Chen, X. (2015a). Semantics Constrained Dictionary Learning for Signer-Independent Sign Language Recognition. In *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, pages 3310–3314, Quebec, Canada.

[Yin et al., 2015b] Yin, F., Chai, X., Zhou, Y., and Chen, X. (2015b). Weakly Supervised Metric Learning towards Signer Adaptation for Sign Language Recognition. In *Proc. British Machine Vision Conference (BMVC)*, pages 35.1–35.12, Swansea, UK.

[Yin et al., 2009] Yin, P., Starner, T., Hamilton, H., Essa, I., and Rehg, J. (2009). Learning the basic units in American Sign Language using discriminative segmental feature selection. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4757–4760, Taipei, Taiwan.

[Yuan et al., 2002] Yuan, Q., Geo, W., Yao, H., and Wang, C. (2002). Recognition of strong and weak connection models in continuous sign language. In *Proc. Int. Conf. on Pattern Recognition (ICPR)*, volume 1, pages 75–78 vol.1, Quebec, Canada.

[Yugopuspito et al., 2018] Yugopuspito, P., Murwantara, I. M., and Sean, J. (2018). Mobile Sign Language Recognition for Bahasa Indonesia Using Convolutional Neural Network. In *Proc. Int. Conf. on Advances in Mobile Computing and Multimedia*, MoMM2018, pages 84–91, New York, NY, USA.

[Zadghorban and Nahvi, 2018] Zadghorban, M. and Nahvi, M. (2018). An Algorithm on Sign Words Extraction and Recognition of Continuous Persian Sign Language Based on Motion and Shape Features of Hands. *Pattern Anal. Appl.*, 21(2):323–335.

[Zafrulla et al., 2010] Zafrulla, Z., Brashear, H., Hamilton, H., and Starner, T. (2010). A novel approach to American Sign Language (ASL) phrase verification using reversed signing. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 48–55.

[Zafrulla et al., 2011] Zafrulla, Z., Brashear, H., Starner, T., Hamilton, H., and Presti, P. (2011). American Sign Language Recognition with the Kinect. In *Proc. Int. Conf. on Multimodal Interfaces (ICMI)*, pages 279–286, Alicante, Spain.

[Zahedi, 2007] Zahedi, M. (2007). *Robust Appearance-Based Sign Language Recognition*. PhD thesis, RWTH Aachen University, Computer Science Department, RWTH Aachen University, Aachen, Germany.

[Zahedi et al., 2008] Zahedi, M., Dreuw, P., Deselaers, T., and Ney, H. (2008). Using Different Aspects of the Signings for Appearance-based Sign Language Recognition. *International Journal of Computational Intelligence*, 4(4):300–307.

[Zahedi et al., 2006] Zahedi, M., Dreuw, P., Rybach, D., Deselaers, T., and Ney, H. (2006). Using Geometric Features to Improve Continuous Appearance-based Sign Language Recognition. In *Proc. British Machine Vision Conference (BMVC)*, volume 3, pages 1019–1028, Edinburgh, UK.

[Zahedi et al., 2005a] Zahedi, M., Keysers, D., and Ney, H. (2005a). Appearance-Based Recognition of Words in American Sign Language. In *Proc. Iberian Conf. on Pattern Recognition and Image Analysis (IbPRIA)*, pages 513–520, Estoril, Portugal.

[Zahedi et al., 2005b] Zahedi, M., Keysers, D., and Ney, H. (2005b). Pronunciation Clustering and Modeling of Variability for Appearance-Based Sign Language Recognition. In *International Workshop on Gesture in Human-Computer Interaction and Simulation*, volume 3881, pages 68–79, Ile-de-Berder, France.

[Zaki and Shaheen, 2011] Zaki, M. M. and Shaheen, S. I. (2011). Sign language recognition using a combination of new vision based features. *Pattern Recognition Letters*, 32(4):572–577.

[Zhang et al., 2016a] Zhang, C., Tian, Y., and Huenerfauth, M. (2016a). Multi-modality American Sign Language recognition. In *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, pages 2881–2885.

[Zhang et al., 2014] Zhang, J., Zhou, W., and Li, H. (2014). A Threshold-based HMM-DTW Approach for Continuous Sign Language Recognition. In *Proc. of Int. Conf. on Internet Multimedia Computing and Service*, pages 237:237–237:240, Xiamen, China.

[Zhang et al., 2015] Zhang, J., Zhou, W., and Li, H. (2015). A new system for Chinese sign language recognition. In *Proc. IEEE China Summit and Int. Conf. on Signal and Information Processing (ChinaSIP)*, pages 534–538.

[Zhang et al., 2016b] Zhang, J., Zhou, W., Xie, C., Pu, J., and Li, H. (2016b). Chinese sign language recognition with adaptive HMM. In *IEEE Int. Conf. on Multimedia and Expo (ICME)*, pages 1–6.

[Zhang et al., 2019] Zhang, Z., Pu, J., Zhuang, L., Zhou, W., and Li, H. (2019). Continuous Sign Language Recognition via Reinforcement Learning. In *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, pages 285–289.

[Zheng and Liang, 2016] Zheng, L. and Liang, B. (2016). Sign language recognition using depth images. In *14th Int. Conf. on Control, Automation, Robotics and Vision (ICARCV)*, pages 1–6, Phuket, Thailand.

[Zhou et al., 2019] Zhou, H., Zhou, W., and Li, H. (2019). Dynamic Pseudo Label Decoding for Continuous Sign Language Recognition. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1282–1287.

[Zhou et al., 2020a] Zhou, H., Zhou, W., Zhou, Y., and Li, H. (2020a). Spatial-Temporal Multi-Cue Network for Continuous Sign Language Recognition. *arXiv:2002.03187 [cs]*.

[Zhou et al., 2020b] Zhou, M., Ng, M., Cai, Z., and Cheung, K. C. (2020b). Self-Attention-based Fully-Inception Networks for Continuous Sign Language Recognition. *Santiago de Compostela*, page 8.

[Zhou et al., 2010] Zhou, Y., Chen, X., Zhao, D., Yao, H., and Gao, W. (2010). Adaptive Sign Language Recognition With Exemplar Extraction and MAP/IVFS. *IEEE Signal Processing Letters*, 17(3):297–300.

[Zhou et al., 2015] Zhou, Y., Yang, X., Zhang, Y., Xu, X., Wang, Y., Chai, X., and Lin, W. (2015). Unsupervised adaptive sign language recognition based on hypothesis comparison guided cross validation and linguistic prior filtering. *Neurocomputing*, 149:1604–1612.

Supplemental Material

Table S1: The table covers most published works up until early 2020 in the field of sign language recognition that report results on a sign language recognition task or introduce a related corpus. Each work is presented with a number of key properties. ‘Vocabulary’ refers to the unique signs in a data set. ‘Signer’ represents the joint number of signers in training and test sets. ‘Isolated’ and ‘Continuous’ refer to the nature of recorded sign language. ‘Input Data’ specifies the employed capturing method for the input (‘RGB’ are colored images, ‘Depth’ is depth information, ‘Colored’ or ‘Electronic Gloves’ represent intrusives gloves that ease the extraction of hand related features, ‘Mocap’ means motion capture to accurately track the body parts). The three columns ‘Manuals’, ‘Non-Manuals’ and ‘Fullframe’ characterize the employed features used for modeling. ‘Benchmark Dataset’ specifies if the employed dataset is a known benchmark dataset, providing comparable tasks.

Reference	Year	Vocabulary	Signer	Isolated	Continuous	Input Data					Manuals				Non-Manuals						Fullframe				Benchmark Dataset	Language
						RGB	Depth	Colored Glove	Electronic Glove	Mocap	Location	Movement	Shape	Orientation	Head	Mouth	Eyes	Eyeblink	Eyebrows	Eyegaze	Bodyjoints	RGB	Depth	Motion		
[Grimes, 1983]	1983	26		x					x				x	x												ASL
[Tamura and Kawasaki, 1988]	1988	10		x		x					x	x	x													JSL
[Murakami and Taguchi, 1991]	1991	10	1	x					x		x		x	x												JSL
[Charayaphan and Marble, 1992]	1992	31	1	x		x					x	x														ASL
[Fels and Hinton, 1993]	1993	203		x					x				x	x												ASL
[Kadous and Taylor, 1995]	1995	95	5	x					x		x	x	x	x												Auslan
[Liang and Ouhyoung, 1995]	1995	26		x					x	x		x	x	x												ASL
[Waldron and Kim, 1995]	1995	14		x					x	x	x	x	x	x												ASL
[Starner and Pentland, 1995]	1995	40			x	x							x													ASL
[Ouhyoung and Liang, 1996]	1996	71			x				x	x	x		x	x												TaiwanSL
[Kim et al., 1996]	1996	25		x					x		x		x	x												KSL
[Kadous, 1996]	1996	95	5	x					x		x		x	x												Auslan
[Vamplew and Adams, 1996]	1996	52	7	x					x			x	x	x												Auslan
[Assan and Grobel, 1997]	1997	26	1		x			x			x		x	x												NGT
[Assan and Grobel, 1997]	1997	262	2	x				x			x		x	x												NGT
[Matsuo et al., 1997]	1997	38	1	x		x					x	x														JSL
[Lee et al., 1997]	1997	131			x				x	x		x	x	x												KSL
[Vogler and Metaxas, 1997]	1997	53			x	x				x	x	x		x												ASL
[Kobayashi and Haruyama, 1997]	1997	6	20	x		x					x		x		x											JSL
[Huang and Huang, 1998]	1998	15		x		x					x		x	x												TaiwanSL
[Liang and Ouhyoung, 1998]	1998	250			x				x			x	x	x												TaiwanSL
[Starner et al., 1998]	1998	40			x	x					x	x	x													ASL
[Vogler and Metaxas, 1999a]	1999	22			x					x	x	x														ASL
[Vogler and Metaxas, 1999b]	1999	22			x					x	x	x														ASL
[Bauer et al., 1999]	1999	100	1		x			x			x		x													DGS
[Imagawa et al., 2000]	2000	33	6	x		x							x													JSL
[Ma et al., 2000]	2000	5177		x		x			x		x			x												CSL
[Wang and Gao, 2000]	2000	274	1	x					x				x													CSL

Continued on next page

Table S1 – continued from previous page

Reference	Year	Vocabulary	Signer	Isolated	Continuous	Input Data					Manuals				Non-Manuals						Fullframe				Benchmark Dataset	Language
						RGB	Depth	Colored Glove	Electronic Glove	Mocap	Location	Movement	Shape	Orientation	Head	Mouth	Eyes	Eyeblink	Eyebrows	Eyegaze	Bodyjoints	RGB	Depth	Motion		
[Holden and Owens, 2000]	2000	22	1	x				x					x													Auslan
[Cui and Weng, 2000]	2000	28		x		x					x	x	x													ASL
[Sagawa and Takeuchi, 2000]	2000	17			x				x			x	x	x												JSL
[Su et al., 2001]	2001	90	2	x					x	x		x														TaiwanSL
[Wang et al., 2001]	2001	5100	1	x					x	x	x	x	x	x												CSL
[Wu and Gao, 2001]	2001	30		x					x	x	x	x	x	x												CSL
[Fang et al., 2001a]	2001	208	7		x				x	x	x	x		x												CSL
[Fang et al., 2001b]	2001	208	7	x					x	x	x	x	x	x												CSL
[Vogler and Metaxas, 2001]	2001	22			x					x	x	x														ASL
[Bauer and Kraiss, 2002a]	2002	12			x					x	x		x	x												
[Tanibata et al., 2002]	2002	65	1	x		x					x	x	x													JSL
[Deng and Tsui, 2002]	2002	192	2	x				x			x	x	x	x												ASL
[Bauer and Kraiss, 2002b]	2002	100	1	x				x			x		x													DGS
[Bauer and Kraiss, 2002b]	2002	50	1	x				x			x		x													DGS
[Yang et al., 2002]	2002	40		x		x						x	x													ASL
[Yuan et al., 2002]	2002	40		x					x	x	x	x	x	x												CSL
[Wang et al., 2002]	2002	5119	1		x				x	x	x		x	x												CSL
[Wang et al., 2002]	2002	5119	1	x					x	x	x		x	x												CSL
[Brashear et al., 2003]	2003	5	1		x	x				x	x		x													ASL
[Gao et al., 2004b]	2004	5113	6		x				x	x	x		x	x												CSL
[Windridge and Bowden, 2004]	2004	115	1	x		x					x	x	x													BSL
[Kadir et al., 2004]	2004	164	1	x		x					x	x	x	x												BSL
[Hernandez-Rebollar et al., 2004]	2004	176	17	x					x	x	x	x	x	x												ASL
[Bowden et al., 2004]	2004	43	1	x		x					x	x	x	x												BSL
[Vogler and Metaxas, 2004]	2004	22			x				x	x		x	x													ASL
[McGuire et al., 2004]	2004	141	1		x				x		x		x													ASL
[Gao et al., 2004a]	2004	5113	6		x				x	x	x		x	x												CSL
[Gao et al., 2004a]	2004	5113	6	x					x	x	x		x	x												CSL
[Nayak et al., 2005]	2005	18	1		x	x						x														ASL
[Zahedi et al., 2005a]	2005	10	3	x		x					x															ASL
[Oz and Leu, 2005]	2005	60	6	x					x	x	x	x	x	x												ASL
[Zahedi et al., 2005b]	2005	50	3	x		x					x															ASL
[Kapuscinski and Wysocki, 2005]	2005	101	2		x	x					x		x													PolishSL
[Zahedi et al., 2006]	2006	103	3		x	x					x		x												BU-104	ASL
[Rybach, 2006]	2006	103	3		x	x						x	x		x										BU-104	ASL
[Wang et al., 2006a]	2006	2435	1	x					x		x	x	x	x												CSL
[Farhadi and Forsyth, 2006]	2006	21			x	x					x	x	x		x											ASL
[Brashear et al., 2006]	2006	22	5		x			x	x		x	x	x													ASL
[von Agris et al., 2006]	2006	153	4	x		x					x	x	x													BSL
[Fujimura and Xia Liu, 2006]	2006	100		x		x	x				x		x													JSL
[Wang et al., 2006b]	2006	26		x					x	x	x	x	x	x												ASL
[Fang et al., 2007]	2007	5113	2		x				x	x	x	x	x	x												CSL
[Zahedi, 2007]	2007	102	3		x	x					x	x	x												BU-104	ASL
[Zahedi, 2007]	2007	10	2	x		x					x	x	x													ASL

Continued on next page

Table S1 – continued from previous page

Reference	Year	Vocabulary	Signer			Input Data					Manuals				Non-Manuals						Fullframe				Benchmark Dataset	Language
				Isolated	Continuous	RGB	Depth	Colored Glove	Electronic Glove	Mocap	Location	Movement	Shape	Orientation	Head	Mouth	Eyes	Eyeblink	Eyebrows	Eyegaze	Bodyjoints	RGB	Depth	Motion		
[Zahedi, 2007]	2007	50	3	x		x					x	x	x												BU-50	ASL
[Stein et al., 2007]	2007	103	3		x	x																x			BU-104	ASL
[von Agris and Kraiss, 2007]	2007	450	5		x	x							x												Signum	DGS
[Dreuw et al., 2007]	2007	103	3		x	x					x	x														ASL
[Mohandes et al., 2007]	2007	300	1	x				x			x		x													ArSL
[Infantino et al., 2007]	2007	40			x	x					x	x				x										LIS
[Infantino et al., 2007]	2007	40		x		x					x	x				x										LIS
[Shanableh et al., 2007]	2007	23	3	x		x					x															ArSL
[Cooper and Bowden, 2007b]	2007	5	9	x		x						x														BSL
[Wang et al., 2007]	2007	100	1	x				x			x		x													CSL
[Cooper and Bowden, 2007a]	2007	164	1	x		x						x	x	x												BSL
[Yang et al., 2007]	2007	39			x	x							x													ASL
[von Agris et al., 2008b]	2008	152			x			x			x	x	x		x	x										DGS
[von Agris et al., 2008b]	2008	229	4	x		x					x	x	x		x	x										BSL
[Forster, 2008]	2008	103	3		x	x							x												BU-104	ASL
[von Agris et al., 2008a]	2008	450	25		x	x					x	x	x		x	x									Signum	DGS
[Maebatake et al., 2008]	2008	183	4		x					x	x	x														JSL
[Paulraj et al., 2008]	2008	32		x		x							x													MalaySL
[Kim et al., 2008]	2008	7	8	x					x			x	x													DGS
[Maraqa and Abu-Zaiter, 2008]	2008	30	2	x				x					x													ArSL
[Aran and Akarun, 2008]	2008	19	8	x				x			x	x	x		x											TSL
[Trmal et al., 2008]	2008	25	20	x		x					x		x													CzSL
[Lichtenauer et al., 2008]	2008	120	75	x		x						x														NGT
[Derpanis et al., 2008]	2008	148	3	x		x						x														ASL
[Zahedi et al., 2008]	2008	102	3		x	x					x	x	x													ASL
[Athitsos et al., 2008]	2008	108	2	x		x					x															ASL
[Dreuw and Ney, 2008]	2008	104	3		x	x					x		x													ASL
[Dreuw, 2008]	2008	103	3		x	x					x	x	x													ASL
[Kelly et al., 2009b]	2009	8	1	x				x			x	x														IrishSL
[Kelly et al., 2009a]	2009	8	1	x				x			x	x														IrishSL
[Yin et al., 2009]	2009	141	1		x				x		x	x														ASL
[Theodorakis et al., 2009]	2009	93	1	x		x						x	x													GSL
[Lichtenauer et al., 2009]	2009	120	75	x		x					x	x														NGT
[Hrúz et al., 2009]	2009	50	1	x		x					x	x	x													CzSL
[Ding and Martinez, 2009]	2009	38	10	x		x					x	x	x													ASL
[Dreuw et al., 2009]	2009	103	3		x	x					x	x	x													ASL
[Dias et al., 2009]	2009	15	4	x		x						x														Libras
[Buehler et al., 2009]	2009	210	3		x	x					x		x	x												BSL
[Liwicki and Everingham, 2009]	2009	100		x		x							x													BSL
[Yang et al., 2009]	2009	48	1	x		x					x	x	x													ASL
[Cooper and Bowden, 2009]	2009	23	1		x	x					x															BSL
[Han et al., 2009]	2009	10	1	x		x						x	x													IrishSL
[Han et al., 2009]	2009	20	1	x				x				x	x													BSL

Continued on next page

Table S1 – continued from previous page

Reference	Year	Vocabulary	Signer			Input Data					Manuals				Non-Manuals						Fullframe				Benchmark Dataset	Language
				Isolated	Continuous	RGB	Depth	Colored Glove	Electronic Glove	Mocap	Location	Movement	Shape	Orientation	Head	Mouth	Eyes	Eyeblink	Eyebrows	Eyegaze	Bodyjoints	RGB	Depth	Motion		
[Awad et al., 2009]	2009	20	1	x				x				x	x													BSL
[Aran et al., 2009a]	2009	19	8	x				x			x	x	x		x											TSL
[Santemiz et al., 2009]	2009	40	1		x	x					x		x													TSL
[Aran et al., 2009b]	2009	19	8	x				x			x	x	x		x											TSL
[Oszust and Wysocki, 2010]	2010	10			x	x					x	x														PolishSL
[Paulraj et al., 2010]	2010	9		x		x							x													MalaySL
[Tolba et al., 2010]	2010	28		x		x							x													ArSL
[Aran and Akarun, 2010]	2010	19	8	x				x			x	x	x													TSL
[Buehler et al., 2010]	2010	210	3		x	x					x		x													BSL
[Zafrulla et al., 2010]	2010	19	5		x			x	x		x	x	x													ASL
[Kong and Ranganath, 2010]	2010	33	1		x					x	x	x														ASL
[Cooper, 2010]	2010	164	1	x		x					x	x	x													BSL
[Cooper, 2010]	2010	5	9	x		x					x	x	x													BSL
[Cooper and Bowden, 2010]	2010	164	1	x		x					x	x														BSL
[Assaleh et al., 2010]	2010	80	1		x	x					x															ArSL
[Assaleh et al., 2010]	2010	23	3	x		x					x															ArSL
[Zhou et al., 2010]	2010	256	6	x					x	x	x	x	x	x												CSL
[Wang et al., 2010]	2010	1113	2	x		x					x	x	x													ASL
[Athitsos et al., 2010]	2010	921	3	x		x						x														ASL
[Theodorakis et al., 2010]	2010	20	1		x	x					x	x														ASL
[Yang et al., 2010]	2010	39	1		x	x					x	x														ASL
[Yang et al., 2010]	2010	40	1		x	x					x	x														ASL
[Yang et al., 2010]	2010	99	3		x	x					x	x														ASL
[Pitsikalis et al., 2010]	2010	50	1		x	x					x	x														ASL
[Kong, 2011]	2011	107	8		x				x	x	x	x	x													ASL
[Yang and Lee, 2011]	2011	24	1		x			x			x	x			x	x	x		x							ASL
[Cooper et al., 2011]	2011	984	1	x		x					x	x	x													GSL
[Zafrulla et al., 2011]	2011	19	7		x	x	x				x	x	x													ASL
[Sarkar et al., 2011]	2011	65			x	x					x	x			x											ASL
[Sarkar et al., 2011]	2011	147	10	x		x					x	x			x											ASL
[Mekala et al., 2011]	2011	26		x		x						x	x													ASL
[Kosmidou et al., 2011]	2011	61	9	x					x			x	x													GSL
[Kelly et al., 2011]	2011	8	2	x				x			x	x														IrishSL
[Uebersax et al., 2011]	2011	26	7	x		x							x	x												ASL
[Rekha et al., 2011]	2011	10		x		x							x													ASL
[Pugeault and Bowden, 2011]	2011	24	4	x		x	x						x													ASL
[Barczak et al., 2011]	2011	36	5	x		x							x													ASL
[Thangali et al., 2011]	2011	82	2	x		x							x													ASL
[Zaki and Shaheen, 2011]	2011	30	3	x		x					x	x	x	x												ASL
[Shanableh and Assaleh, 2011]	2011	23	3	x				x			x	x														ArSL
[Pitsikalis et al., 2011]	2011	961	1	x		x					x	x														GSL
[Ong et al., 2012]	2012	40	14	x		x	x				x	x	x								x					DGS
[Ong et al., 2012]	2012	982	1	x		x					x	x	x								x					GSL
[Cooper et al., 2012]	2012	164	1	x		x					x	x	x								x					BSL

Continued on next page

Table S1 – continued from previous page

Reference	Year	Vocabulary	Signer	Isolated	Continuous	Input Data					Manuals				Non-Manuals						Fullframe				Benchmark Dataset	Language
						RGB	Depth	Colored Glove	Electronic Glove	Mocap	Location	Movement	Shape	Orientation	Head	Mouth	Eyes	Eyeblink	Eyebrows	Eyegaze	Bodyjoints	RGB	Depth	Motion		
[Cooper et al., 2012]	2012	20	6	x		x	x				x	x	x								x					GSL
[Cooper et al., 2012]	2012	40	15	x		x					x	x	x								x					DGS
[Cooper et al., 2012]	2012	984	1	x		x					x	x	x								x					GSL
[Forster et al., 2012]	2012	266	1		x	x																				DGS
[Forster et al., 2012]	2012	911	7		x	x																				DGS
[Oberdörfer et al., 2012]	2012	455	1		x	x					x		x		x										Signum	DGS
[Oberdörfer et al., 2012]	2012	455	25		x	x					x		x		x										Signum	DGS
[Gweth et al., 2012]	2012	455	1		x	x					x		x												Signum	DGS
[Kishore and Kumar, 2012]	2012	80	10	x		x							x													IndianSL
[Mohandes et al., 2012]	2012	300	3	x				x			x		x													ArSL
[Lang et al., 2012]	2012	25		x		x	x				x	x														DGS
[Kindiroglu et al., 2012]	2012	88	11	x		x							x													CzSL Rus- sianSL TSL
[Caridakis et al., 2012]	2012	118	3	x		x					x	x	x													GSL
[Dreuw, 2012]	2012	103	3		x	x					x	x	x		x											ASL
[Fagiani et al., 2012]	2012	147	10	x		x																				LIS
[Sun et al., 2013b]	2013	73	9	x		x	x				x	x	x								x					ASL
[Sun et al., 2013a]	2013	73	9	x		x	x				x	x	x								x					ASL
[Oszust and Wysocki, 2013b]	2013	30	1	x		x	x				x		x								x					PolishSL
[Oszust and Wysocki, 2013a]	2013	30	1	x		x	x				x	x	x								x					PolishSL
[Forster et al., 2013b]	2013	266	1		x	x						x	x		x							x				DGS
[Forster et al., 2013b]	2013	455	1		x	x						x	x		x							x			Signum	DGS
[Forster et al., 2013b]	2013	455	25		x	x						x	x		x							x			Signum	DGS
[Forster et al., 2013a]	2013	266	1		x	x						x	x		x							x				DGS
[Forster et al., 2013a]	2013	455	1		x	x						x	x		x							x			Signum	DGS
[Mohandes and Deriche, 2013]	2013	100	1	x					x		x	x	x	x												ArSL
[Chai et al., 2013]	2013	239		x		x	x					x														CSL
[Agarwal and Thakur, 2013]	2013	10		x		x	x					x	x													CSL
[Mohandes, 2013]	2013	100	2	x					x	x	x	x	x	x												ArSL
[Kuznetsova et al., 2013]	2013	24	3	x		x	x						x													ASL
[Elons et al., 2013]	2013	50		x		x	x						x													ArSL
[Yang and Lee, 2013]	2013	24			x	x					x	x			x											ASL
[Roussos et al., 2013]	2013	100	2	x		x					x	x	x													GSL
[Han et al., 2013]	2013	20	2	x		x						x	x													BSL
[Zhang et al., 2014]	2014	34	3		x	x	x				x										x					CSL
[Zhang et al., 2014]	2014	34	5	x		x	x				x										x					CSL
[Ong et al., 2014]	2014	48	3		x	x						x	x													ASL
[Ong et al., 2014]	2014	40	14	x		x	x				x	x	x								x					DGS
[Ong et al., 2014]	2014	981	1	x		x					x	x	x								x					GSL
[Chai et al., 2014]	2014	2000	8	x		x	x				x		x								x				Devisign-L	CSL
[Chai et al., 2014]	2014	36	8	x		x	x				x		x								x				Devisign-G	CSL
[Chai et al., 2014]	2014	500	8	x		x	x				x		x								x				Devisign-D	CSL
[Kong and Ranganath, 2014]	2014	107	8		x				x	x	x	x	x	x												ASL

Continued on next page

Table S1 – continued from previous page

Reference	Year	Vocabulary	Signer	Isolated	Continuous	Input Data					Manuals				Non-Manuals						Fullframe				Benchmark Dataset	Language
						RGB	Depth	Colored Glove	Electronic Glove	Mocap	Location	Movement	Shape	Orientation	Head	Mouth	Eyes	Eyeblink	Eyebrows	Eyegaze	Bodyjoints	RGB	Depth	Motion		
[Forster et al., 2014]	2014	1558	9		x	x																				DGS
[Forster et al., 2014]	2014	911	7		x	x																				DGS
[Mohandes et al., 2014]	2014	38	10	x			x						x	x												ArSL
[Igari and Fukumura, 2014]	2014	80	3	x						x		x														JSL
[Theodorakis et al., 2014]	2014	1046	2	x		x					x	x	x													GSL
[Theodorakis et al., 2014]	2014	94	1	x		x					x	x	x													ASL
[Theodorakis et al., 2014]	2014	97	2	x		x					x	x	x													ASL
[Almeida et al., 2014]	2014	34	1	x		x	x				x	x	x	x												Libras
[Geng et al., 2014]	2014	8	8	x		x	x				x	x	x													CSL
[Huang et al., 2015]	2015	25	9	x		x	x				x	x	x	x							x					CSL
[Yin et al., 2015b]	2015	1000	7	x		x	x				x	x	x								x					CSL
[Koller et al., 2015]	2015	1080	9		x	x					x	x	x		x	x	x		x						Phoenix14	DGS
[Zhang et al., 2015]	2015	30	5		x	x	x					x									x					CSL
[Zhang et al., 2015]	2015	30	5	x		x	x					x									x					CSL
[Yin et al., 2015a]	2015	1000	7	x		x	x					x	x								x					CSL
[Yin et al., 2015a]	2015	370	1	x		x	x					x	x								x					CSL
[Wang et al., 2015]	2015	1000	1	x		x	x				x		x								x					CSL
[Wang et al., 2015]	2015	1000	7	x		x	x				x		x								x					CSL
[Wang et al., 2015]	2015	370	1	x		x	x				x		x								x					CSL
[Neto et al., 2015]	2015	18		x		x							x													Libras
[Chai et al., 2015]	2015	1000	1	x		x	x				x	x	x													CSL
[Chai et al., 2015]	2015	1000	7	x		x	x				x	x	x													CSL
[Nagendraswamy et al., 2015]	2015	15	4	x		x					x															IndianSL
[Fagiani et al., 2015]	2015	147	10	x		x					x		x	x	x											LIS
[Tubaiz et al., 2015]	2015	80	1		x				x	x	x	x														ArSL
[Cheng et al., 2015]	2015	223	5	x						x			x	x												CSL
[Zhou et al., 2015]	2015	20	7	x		x	x						x													CSL
[Koller et al., 2016c]	2016	1080	9		x	x							x												Phoenix14	DGS
[Koller et al., 2016c]	2016	455	1		x	x							x												Signum	DGS
[Koller et al., 2016a]	2016	1080	9		x	x					x	x	x	x	x	x	x		x						Phoenix14	DGS
[Koller et al., 2016b]	2016	1080	9		x	x					x	x	x		x	x	x		x						Phoenix14	DGS
[Koller et al., 2016b]	2016	455	1		x	x					x	x	x		x	x	x		x						Signum	DGS
[Zheng and Liang, 2016]	2016	36	8	x		x	x																x		Devisign-G	CSL
[Yin et al., 2016]	2016	1000	1	x		x	x					x	x													CSL
[Yin et al., 2016]	2016	1000	7	x		x	x					x	x													CSL
[Yin et al., 2016]	2016	2000	8	x		x	x					x	x												Devisign-L	CSL
[Zhang et al., 2016b]	2016	100	1	x		x	x				x		x								x					CSL
[Zhang et al., 2016b]	2016	500	1	x		x	x				x		x								x					CSL
[Zhang et al., 2016a]	2016	99	5		x	x	x				x	x	x		x	x	x				x					ASL
[Pu et al., 2016a]	2016	500	50	x		x	x				x	x	x								x					CSL
[Liu et al., 2016]	2016	100	50	x		x	x														x					CSL
[Liu et al., 2016]	2016	500	50	x		x	x														x					CSL
[Lim et al., 2016]	2016	50	3	x		x							x												BU-50	ASL
[Nagendraswamy and Kumara, 2016]	2016	26	4		x	x																x				IndianSL
[Pigou et al., 2016]	2016	100	53	x		x																x				FlemishSL
[Pigou et al., 2016]	2016	100	78	x		x																x				NGT
[Camgöz et al., 2016]	2016	33	6	x		x	x				x	x	x													TSL

Continued on next page

Table S1 – continued from previous page

						Input Data					Manuals				Non-Manuals						Fullframe					
Reference	Year	Vocabulary	Signer	Isolated	Continuous	RGB	Depth	Colored Glove	Electronic Glove	Mocap	Location	Movement	Shape	Orientation	Head	Mouth	Eyes	Eyeblink	Eyebrows	Eyegaze	Bodyjoints	RGB	Depth	Motion	Benchmark Dataset	Language
[Yang et al., 2016]	2016	21	2		x	x	x														x					CSL
[Yang et al., 2016]	2016	21	8	x		x	x														x					CSL
[Ronchetti et al., 2016]	2016	16		x				x					x												LSA16	ArgentSL
[Pu et al., 2016b]	2016	100	14	x		x	x				x	x														CSL
[Wei et al., 2016]	2016	110	5	x						x		x	x	x												CSL
[Li et al., 2016]	2016	510	5		x				x		x	x	x	x												CSL
[Wang et al., 2016]	2016	1000	1	x		x	x					x	x													CSL
[Wang et al., 2016]	2016	370	1	x		x	x					x	x													CSL
[Camgoz et al., 2016]	2016	855	10	x		x	x																			TSL
[Cui et al., 2017]	2017	1080	9		x	x							x												Phoenix14	DGS
[Camgoz et al., 2017]	2017	1080	9		x	x																x			Phoenix14	DGS
[Koller et al., 2017]	2017	1080	9		x	x																x			Phoenix14	DGS
[Koller et al., 2017]	2017	455	1		x	x																x			Signum	DGS
[Li et al., 2017]	2017	80	10	x		x	x				x	x	x								x					CSL
[Guo et al., 2017]	2017	370	5	x		x	x				x		x								x					CSL
[García-Bautista et al., 2017]	2017	20	35	x		x	x				x										x					MexicanSL
[Pigou et al., 2017]	2017	10	53	x		x																x				FlemishSL
[Pigou et al., 2017]	2017	10	78	x		x																x				NGT
[Pigou et al., 2017]	2017	100	53	x		x																x				FlemishSL
[Pigou et al., 2017]	2017	100	78	x		x																x				NGT
[Pigou et al., 2017]	2017	20	53	x		x																x				FlemishSL
[Pigou et al., 2017]	2017	20	78	x		x																x				NGT
[Pigou et al., 2017]	2017	50	53	x		x																x				FlemishSL
[Pigou et al., 2017]	2017	50	78	x		x																x				NGT
[Hu et al., 2017]	2017	24	5	x		x	x						x													ASL
[Hu et al., 2017]	2017	24	9	x		x	x						x													ASL
[Quiroga et al., 2017]	2017	16		x				x					x												LSA16	ArgentSL
[Thang et al., 2017]	2017	45	6	x		x					x		x	x												ASL
[Thang et al., 2017]	2017	95	5	x					x	x	x		x	x												Auslan
[Pezzuoli et al., 2017]	2017	40	1	x					x				x	x												LIS
[O'Connor et al., 2017]	2017	26		x					x			x	x	x												ASL
[Ji et al., 2017]	2017	6		x		x					x		x													KSL
[Costa Filho et al., 2017]	2017	61	10	x		x	x						x													Libras
[Fang et al., 2017]	2017	16	11		x		x				x	x	x													ASL
[Fang et al., 2017]	2017	56	11	x			x				x	x	x													ASL
[Huang et al., 2018b]	2018	1080	9		x	x							x									x			Phoenix14	DGS
[Huang et al., 2018b]	2018	178	50		x	x							x									x			CSL	CSL
[Pu et al., 2018]	2018	1080	9		x	x																x			Phoenix14	DGS
[Pu et al., 2018]	2018	178	50		x	x																x			CSL	CSL
[Wang et al., 2018]	2018	1080	9		x	x																x			Phoenix14	DGS
[Wang et al., 2018]	2018	178	50		x	x																x			CSL	CSL
[Koller et al., 2018]	2018	1080	9		x	x							x												Phoenix14	DGS
[Koller et al., 2018]	2018	455	1		x	x							x												Signum	DGS
[Konstantinidis et al., 2018a]	2018	50	1	x		x															x	x		x		DGS
[Konstantinidis et al., 2018a]	2018	64	10	x				x													x	x		x	LSA64	ArgentSL

Continued on next page

Table S1 – continued from previous page

Reference	Year	Vocabulary	Signer	Isolated	Continuous	Input Data: RGB	Input Data: Depth	Input Data: Colored Glove	Input Data: Electronic Glove	Input Data: Mocap	Manuals: Location	Manuals: Movement	Manuals: Shape	Manuals: Orientation	Non-Manuals: Head	Non-Manuals: Mouth	Non-Manuals: Eyes	Non-Manuals: Eyeblink	Non-Manuals: Eyebrows	Non-Manuals: Eyegaze	Fullframe: Bodyjoints	Fullframe: RGB	Fullframe: Depth	Fullframe: Motion	Benchmark Dataset	Language
[Konstantinidis et al., 2018b]	2018	64	10	x				x			x	x	x	x							x				LSA64	ArgentSL
[Kishore et al., 2018]	2018	500	5	x						x	x	x	x	x	x						x					IndianSL
[Kumar et al., 2018c]	2018	30	10	x		x	x					x	x								x					IndianSL
[Huang et al., 2018a]	2018	500	50	x		x	x				x	x	x								x					CSL
[Liu et al., 2018]	2018	227		x		x	x				x		x			x						x				HKSL
[Rao and Kishore, 2018]	2018	18	10		x	x																x				IndianSL
[Kumar et al., 2018b]	2018	200	10	x						x											x					IndianSL
[Kumar et al., 2018a]	2018	500	10	x						x											x					IndianSL
[Camgoz et al., 2018]	2018	1066	9		x	x																x			Phoenix14T	DGS
[Ebling et al., 2018]	2018	100	30	x		x	x																			DSGS
[Huang et al., 2018c]	2018	310	14	x		x	x						x								x			x		CSL
[Guo et al., 2018]	2018	178	50		x	x					x	x	x												CSL	CSL
[Gunawan et al., 2018]	2018	10	10	x				x			x	x	x												LSA64	ArgentSL
[Yugopuspito et al., 2018]	2018	23		x		x							x													IndoSL
[Gruber et al., 2018]	2018	10	18	x		x	x						x													CzSL
[Kumar et al., 2018d]	2018	51		x		x					x	x	x													ASL
[Rao et al., 2018]	2018	200	5	x		x					x		x													IndianSL
[Shenoy et al., 2018]	2018	12		x		x							x													IndianSL
[Rakowski and Wandzik, 2018]	2018	24	5	x		x	x						x													ASL
[Mathur and Sharma, 2018]	2018	32		x		x					x	x	x	x												ASL
[Lee and Lee, 2018]	2018	28		x					x				x	x												ASL
[Hashim and Alizadeh, 2018]	2018	12		x		x							x													KurdishSL
[Handhika et al., 2018]	2018	25	2	x		x	x				x															IndoSL
[Ariesta et al., 2018]	2018	30	10	x		x					x	x	x													IndoSL
[Zadghorban and Nahvi, 2018]	2018	46	3	x		x						x	x													PersianSL
[Ye et al., 2018]	2018	27	14		x	x	x				x	x	x	x												ASL
[Xie et al., 2018]	2018	24	5	x		x	x						x													ASL
[Ma et al., 2018]	2018	276	5	x							x		x													ASL
[Pu et al., 2019]	2019	1080	9		x	x																x			Phoenix14	DGS
[Pu et al., 2019]	2019	178	50		x	x																			CSL	CSL
[Liao et al., 2019]	2019	500	8	x		x							x												Devisign-D	CSL
[Guo et al., 2019]	2019	1080	9		x	x																x			Phoenix14	DGS
[Guo et al., 2019]	2019	178	50		x	x																x			CSL	CSL
[Pei et al., 2019]	2019	1080	9		x	x																x			Phoenix14	DGS
[Cui et al., 2019]	2019	1080	9		x	x																x			Phoenix14	DGS
[Cui et al., 2019]	2019	1080	9		x	x																x		x	Phoenix14	DGS
[Cui et al., 2019]	2019	1080	9		x	x																		x	Phoenix14	DGS
[Cui et al., 2019]	2019	1080	9		x	x							x												Phoenix14	DGS
[Cui et al., 2019]	2019	455	1		x	x																x		x	Signum	DGS
[Song et al., 2019]	2019	1080	9		x	x																x			Phoenix14	DGS
[Zhou et al., 2019]	2019	1080	9		x	x																x			Phoenix14	DGS
[Zhou et al., 2019]	2019	178	50		x	x																x			CSL	CSL
[Zhang et al., 2019]	2019	1080	9		x	x																x			Phoenix14	DGS
[Yang et al., 2019]	2019	1080	9		x	x																x			Phoenix14	DGS
[Yang et al., 2019]	2019	178	50		x	x																x			CSL	CSL

Continued on next page

Table S1 – continued from previous page

Reference	Year	Vocabulary	Signer			Input Data					Manuals				Non-Manuals						Fullframe				Benchmark Dataset	Language
				Isolated	Continuous	RGB	Depth	Colored Glove	Electronic Glove	Mocap	Location	Movement	Shape	Orientation	Head	Mouth	Eyes	Eyeblink	Eyebrows	Eyegaze	Bodyjoints	RGB	Depth	Motion		
[Koller et al., 2019]	2019	1066	9		x	x							x			x						x			Phoenix14T	DGS
[Koller et al., 2019]	2019	1080	9		x	x							x			x						x			Phoenix14	DGS
[Bilge et al., 2019]	2019	50	1	x		x																x				ASL
[Bilge et al., 2019]	2019	50	1	x		x							x									x				ASL
[Bilge et al., 2019]	2019	50	1	x		x							x													ASL
[Vaezi Joze and Koller, 2019]	2019	100	189	x		x															x				MS-ASL	ASL
[Vaezi Joze and Koller, 2019]	2019	100	189	x		x																x			MS-ASL	ASL
[Vaezi Joze and Koller, 2019]	2019	1000	222	x		x															x				MS-ASL	ASL
[Vaezi Joze and Koller, 2019]	2019	1000	222	x		x																x			MS-ASL	ASL
[Vaezi Joze and Koller, 2019]	2019	200	196	x		x															x				MS-ASL	ASL
[Vaezi Joze and Koller, 2019]	2019	200	196	x		x																x			MS-ASL	ASL
[Vaezi Joze and Koller, 2019]	2019	500	222	x		x															x				MS-ASL	ASL
[Vaezi Joze and Koller, 2019]	2019	500	222	x		x																x			MS-ASL	ASL
[Sabyrov et al., 2019]	2019	20	3	x		x									x	x	x		x		x					K-RSL
[Kındıroğlu et al., 2019]	2019	174	4	x		x	x				x		x								x					TSL
[Wang et al., 2019]	2019	138	70	x					x	x	x	x	x	x												CSL
[Kumar et al., 2019]	2019	700	10	x						x											x					IndianSL
[Wei et al., 2019]	2019	178	50		x	x																x			CSL	CSL
[Tornay et al., 2019]	2019	94	30	x		x	x					x	x													DSGS
[Farag and Brock, 2019]	2019	12			x					x	x	x														JSL
[Paudyal et al., 2019]	2019	25	100	x		x					x	x	x													ASL
[Jose and Julian, 2019]	2019	31		x		x							x													TamisSL
[Latif et al., 2019]	2019	32	40	x		x																				ArSL
[Avola et al., 2019]	2019	30	20	x			x				x	x	x	x												ASL
[Mittal et al., 2019]	2019	35			x		x				x	x	x	x												ISL
[Hassan et al., 2019]	2019	80	1		x				x	x	x	x														ArSL
[Hassan et al., 2019]	2019	80	1		x	x					x	x														ArSL
[Hassan et al., 2019]	2019	80	2		x					x	x	x														ArSL
[Albanie et al., 2020]	2020	1000	222	x		x																x			MS-ASL	ASL
[Albanie et al., 2020]	2020	1064	40	x		x															x					BSL
[Albanie et al., 2020]	2020	1064	40	x		x									x							x				BSL
[Albanie et al., 2020]	2020	2000	119	x		x																x			WSASL	ASL
[Adaloglou et al., 2020]	2020	1080	9		x	x																x			Phoenix14	DGS
[Adaloglou et al., 2020]	2020	178	50		x	x																x			CSL	CSL
[Adaloglou et al., 2020]	2020	310	7		x	x																x				GSL
[Adaloglou et al., 2020]	2020	310	7	x		x																x				GSL
[Papastratis et al., 2020]	2020	1066	9		x	x																x			Phoenix14T	DGS
[Papastratis et al., 2020]	2020	1080	9		x	x																x			Phoenix14	DGS
[Papastratis et al., 2020]	2020	178	50		x	x																x			CSL	CSL
[Zhou et al., 2020b]	2020	1080	9		x	x																x			Phoenix14	DGS
[Cheng et al., 2020]	2020	1066	9		x	x																x			Phoenix14T	DGS
[Cheng et al., 2020]	2020	1080	9		x	x																x			Phoenix14	DGS

Continued on next page

Table S1 – continued from previous page

						Input Data					Manuals				Non-Manuals						Fullframe					
Reference	Year	Vocabulary	Signer	Isolated	Continuous	RGB	Depth	Colored Glove	Electronic Glove	Mocap	Location	Movement	Shape	Orientation	Head	Mouth	Eyes	Eyeblink	Eyebrows	Eyegaze	Bodyjoints	RGB	Depth	Motion	Benchmark Dataset	Language
[Cheng et al., 2020]	2020	178	50		x	x																x			CSL	CSL
[Zhou et al., 2020a]	2020	1066	9		x	x							x		x						x	x			Phoenix14T	DGS
[Zhou et al., 2020a]	2020	1080	9		x	x							x		x						x	x			Phoenix14	DGS
[Zhou et al., 2020a]	2020	178	50		x	x							x		x						x	x			CSL	CSL
[Mukushev et al., 2020]	2020	20	5	x		x									x	x	x		x		x					K-RSL
[Vasudevan et al., 2020]	2020	19	58	x		x																x				LSE
[De Coster et al., 2020]	2020	100	67	x		x															x					FlemishSL
[De Coster et al., 2020]	2020	100	67	x		x																x				FlemishSL
[Camgoz et al., 2020]	2020	1066	9		x	x																x			Phoenix14T	DGS
[Tamer and Saraçlar, 2020]	2020	1066	9		x	x															x				Phoenix14T	DGS
[Li et al., 2020a]	2020	100	97	x		x															x				WSASL	ASL
[Li et al., 2020a]	2020	100	97	x		x																x			WSASL	ASL
[Li et al., 2020a]	2020	1000	116	x		x															x				WSASL	ASL
[Li et al., 2020a]	2020	1000	116	x		x																x			WSASL	ASL
[Li et al., 2020a]	2020	2000	119	x		x															x				WSASL	ASL
[Li et al., 2020a]	2020	2000	119	x		x																x			WSASL	ASL
[Li et al., 2020a]	2020	300	109	x		x															x				WSASL	ASL
[Li et al., 2020a]	2020	300	109	x		x																x			WSASL	ASL
[Li et al., 2020b]	2020	100	189	x		x																x			MS-ASL	ASL
[Li et al., 2020b]	2020	100	97	x		x																x			WSASL	ASL
[Li et al., 2020b]	2020	200	196	x		x																x			MS-ASL	ASL
[Li et al., 2020b]	2020	300	109	x		x																x			WSASL	ASL
[Özdemir et al., 2020]	2020	744	6	x		x	x															x		x		TSL
[Izutov, 2020]	2020	100	189	x		x																x			MS-ASL	ASL
[Izutov, 2020]	2020	1000	222	x		x																x			MS-ASL	ASL
[Izutov, 2020]	2020	200	196	x		x																x			MS-ASL	ASL
[Izutov, 2020]	2020	500	222	x		x																x			MS-ASL	ASL