Proceedings of the 7th International Workshop on Sign Language Translation and Avatar Technology (SLTAT 7), pages 79–84
Language Resources and Evaluation Conference (LREC 2022), Marseille, 20-25 June 2022

Greek Sign Language Recognition for the SL-ReDu Learning Platform

Katerina Papadimitriou [1], Gerasimos Potamianos [1], Galini Sapountzaki [2], Theodor Goulas [3], Eleni Efthimiou [3], Stavroula-Evita Fotinea [3], Petros Maragos [4]

[1] Department of Electrical & Computer Engineering, University of Thessaly, Volos, Greece
[2] Department of Special Education, University of Thessaly, Volos, Greece
[3] Institute for Language & Speech Processing, Athena Research & Innovation Center, Athens, Greece
[4] School of Electrical & Computer Engineering, National Technical University of Athens, Greece

aipapadimitriou@uth.gr, gpotam@ieee.org, gsapountz@sed.uth.gr,
{tgoulas, eleni_e, evita}@athenarc.gr, maragos@cs.ntua.gr

Abstract

There has been increasing interest lately in developing education tools for sign language (SL) learning that enable self-assessment and objective evaluation of learners' SL productions, assisting both students and their instructors. Crucially, such tools require the automatic recognition of SL videos, while operating in a signer-independent fashion and under realistic recording conditions. Here, we present an early version of a Greek Sign Language (GSL) recognizer that satisfies the above requirements, and integrate it within the SL-ReDu learning platform that constitutes a first in GSL with recognition functionality. We develop the recognition module incorporating state-of-the-art deep-learning based visual detection, feature extraction, and classification, designing it to accommodate a medium-size vocabulary of isolated signs and continuously fingerspelled letter sequences. We train the module on a specifically recorded GSL corpus of multiple signers by a web-cam in non-studio conditions, and conduct both multi-signer and signer-independent recognition experiments, reporting high accuracies. Finally, we let student users evaluate the learning platform during GSL production exercises, reporting very satisfactory objective and subjective assessments based on recognition performance and collected questionnaires, respectively.

Keywords: Greek Sign Language recognition, MediaPipe, MobileNet, ResNet, CNN, BiLSTM, sign language learning, user evaluation

1. Introduction

Sign languages (SLs) involve a complex non-vocal means of communication in the 3D visible space around the signer, with both manual and non-manual articulation carrying linguistic content of a set of glosses (Armstrong et al., 2002). Such complexity renders SL education a difficult and time-consuming process (Kemp, 1998) for both learners and their instructors, thus motivating recently the development of automatic SL assessment and tutoring tools (Aran et al., 2009; Zafrulla et al., 2011; Ebling et al., 2018; Joy et al., 2019; Mohammdi and Elbourhamy, 2021). A critical functionality in such applications is the ability to assess the validity of the learners' SL productions, necessitating automatic SL recognition (SLR) of the produced videos in a signer-independent fashion and under realistic, non-ideal recording conditions. Not surprisingly, this constitutes a challenging problem, due to the aforementioned complexity of SL production, coupled with the intricacies of robust video processing (detection, tracking, representation) and inherent inter-signer production variability.

Motivated by the above, in conjunction with the lack of learning tools in the under-resourced Greek Sign Language (GSL), we have recently initiated the "SL-ReDu" project (Potamianos et al., 2020). This aims to considerably advance the current state-of-the-art in automatic recognition of GSL from videos, while focusing on the use-case of standardized GSL teaching as a second language. For this purpose, in previous work we have already developed a suitable platform that allows "passive"-type GSL learning exercises (e.g., multiple-choice questions) and populated it with appropriate learning material (Sapountzaki et al., 2021; Efthimiou et al., 2021). However, we have not yet enabled "active", production-type assessment, which requires an appropriate SLR module.

In this paper, we proceed to enable such functionality, presenting our initial GSL recognition module that we integrate to the SL-ReDu platform. In particular, we focus on two recognition problems: (i) that of isolated GSL signs within a medium-size vocabulary, developing separate models for numerals and non-numerals, and (ii) that of continuous sequences of fingerspelled letters of the Greek alphabet. Note that the latter task plays a critical role in SLs, as it is regularly used for words that lack unique signs, such as names, technical phrases, and foreign words, among others (Armstrong et al., 2002).

We develop the corresponding SLR module incorporating state-of-the-art deep-learning techniques. Specifically, we utilize the MediaPipe library for detecting the signer and relevant landmarks from RGB video (Lugaresi et al., 2019), thus avoiding the use of special sensing equipment, such as hand gloves (Mehdi and Khan, 2002) or depth cameras (Ren et al., 2011). Fur-

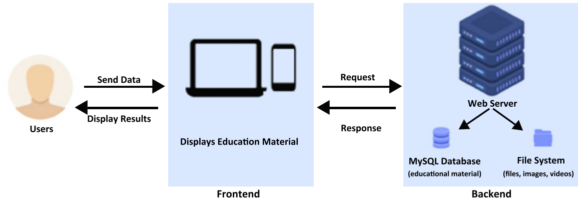


Figure 1: Illustration of the SL-ReDu prototype system web-based architecture.

ther, we employ convolutional neural networks (CNNs) for visual feature learning, namely a 3D CNN (Tran et al., 2018) and MobileNet (Howard et al., 2017). Finally, in the case of fingerspelling, for sequence learning we use a bidirectional long short-term memory (BiLSTM) encoder (Schuster and Paliwal, 1997) and connectionist temporal classification (CTC) based decoding (Graves et al., 2006).

We train and evaluate the recognition module on a suitable GSL corpus, collected as part of this work. The data contain multiple signers, recorded using a typical web-cam in non-studio conditions. We report both multi-signer and signer-independent recognition experiments on this corpus. Moreover, we evaluate the SL-ReDu platform and its recognition functionality with a small number of student users that conduct GSL production exercises, reporting both objective and subjective evaluation results.

The rest of the paper is structured as follows: Section 2 overviews the SL-ReDu platform; Section 3 describes the developed SLR module; Section 4 presents the SLR corpus and its evaluation; Section 5 discusses the user evaluation; and Section 6 concludes the paper.

2. The SL-ReDu Platform

The SL-ReDu platform attempts to handle the drawbacks of conventional practice and testing strategies in learning GSL as a second language by enabling self-monitoring of learning and objective learner evaluation. For the system's design all aspects of GSL linguistics are being considered, i.e. GSL semantics, as well as morpho-syntactic effects in both GSL recognition and GSL production. In particular, teaching techniques and content are integrated into the system design, including various SL practice assignments that cover GSL phenomena from sign formation to complicated syntactic and semantic utterance production. Ordinary multiple-choice questions that utilize images, videos, and text to elicit a response from the user, as well as user feedback by means of video recordings of GSL production, are examples of exercise types. With the integration of SLR technology, SL-ReDu enables the user to actively sign and be assessed for the capacity to appropriately generate signs.

The SL-ReDu prototype system is a web-based application that runs on a web server managing the end-user's interaction. Self-monitoring and objective assessment system modalities entail a variety of components, namely the system database, the front-end and back-end user interfaces, as well as image and video files. Further, the system involves a content management system (back-end) that is exploited by the instructor to create learners' assessment tests and track performance over time. Figure 1 depicts the adopted architecture.

To build the dynamic web platform, the PHP programming language in conjunction with HTML5, CSS3, and JavaScript is used. A MySQL open-source database is employed for the construction of the web application, including the storage of the content, as well as the results of the platform users. An Apache Web Server hosts the web application.

SLR represents a separate module of the system that runs as standalone on the learner's device (typically a higher-end laptop with an available camera). The technical details of the communication between the web server and the SLR engine are available in a Technical Report (Potamianos et al., 2021).

3. The GSL Recognition Module

We next detail the SLR module for the two GSL recognition tasks considered, namely that of isolated signs and continuous fingerspelling. The module also contains a pre-processing stage.

3.1. Pre-processing

This stage is employed to detect the signer, extract the region-of-interest (RoI), and provide feedback in case signer positioning is incorrect.

Specifically, the recorded video frames are fed to the MediaPipe holistic tool (Lugaresi et al., 2019). This is a multi-stage pipeline that integrates separate models for pose, face, and hand components, extracting 543 whole-body landmarks from RGB data (33 pose, 468 face, and 21 hand landmarks per hand). Lack of detected landmarks of the two hands, face, and upper torso is assumed to imply incorrect user positioning with respect to the camera field of view. In such case, the signer is prompted by the system to reposition.

If user positioning is correct, the detected landmarks are utilized to extract the RoI for subsequent appearance feature generation. In the case of isolated signs, where multiple articulators may participate in signing, the entire upper body is cropped producing the RoI (see also Figure 2(a)). This is then normalized to the subsequent CNN input layer size (i.e., 256×256 pixels of ResNet2+1D). In the case of fingerspelling though, where typically one hand constitutes the sole articulator, the RoI consists of the signing hand only (see also Figure 2(b)), which is determined based on the motion of the landmarks (3D skeletal keypoints) of each hand in the video. The RoI is then normalized to the input layer size of the MobileNet CNN (i.e., 224×224 pixels). Note that we use the estimated landmarks exclusively for RoI extraction, thus minimizing the impact of occasional MediaPipe failures (Moryossef et al., 2021).

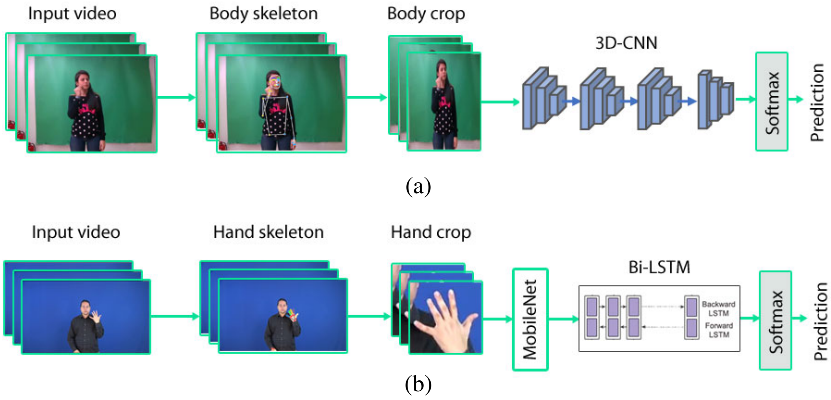


Figure 2: Schematics of GSL recognition modules for (a) isolated signs and (b) continuous fingerspelling.

3.2. Isolated Sign Recognition

A 3D CNN is employed for isolated sign recognition (see also Figure 2(a)). Specifically, the 18-layer ResNet2+1D model is used (Tran et al., 2018) that separates spatial and temporal convolutions of 3D CNNs. Note that two recognition subtasks are considered employing separate models, one for numeral signs with a vocabulary size of 18, and a second for non-numeral ones with a vocabulary size of 36.

Note that in all cases the CNN is pretrained on the Kinetics dataset (Carreira et al., 2018). Model training (finetuning) then proceeds via the Adam optimizer (Kingma and Ba, 2014) with initial learning rate set to 0.0001 and weight decay 0.0001. For sign prediction, the cross-entropy loss is used with label smoothing (Szegedy et al., 2016). The mini-batch size is fixed to 16.

3.3. Continuous Fingerspelling Recognition

A CNN-BiLSTM combination is employed for recognizing continuously fingerspelled sequences of the 24 Greek alphabet letters. In the adopted approach (see also Figure 2(b)), the CNN serves as visual feature learner of each video frame and the BiLSTM learns their temporal relations. Specifically, the CNN uses the MobileNet architecture (Howard et al., 2017), pretrained on the ImageNet corpus (Deng et al., 2009). Feature maps are generated by taking the output of the last fully-connected layer, yielding 1024-dimensional (dim) features. These are then fed to a linear projection layer for size reduction, resulting in 512-dim features. Subsequently, a two-layer BiLSTM encoder is employed with 512-dim hidden states (Schuster and Paliwal, 1997) followed by CTC decoding (Graves et al., 2006) for letter sequence prediction.

The model's linear projection layer is jointly trained with the BiLSTM. Training is conducted using the Adam optimizer with initial learning rate equal to 0.001, decayed by a factor of 0.1 if the validation score remains consistent for 9 steps. In addition, a dropout rate of 0.1 is used, and the mini-batch size is fixed to 16. Finally, during inference, beam search decoding is adopted with beam width 3. Note also that no letter language model is employed.

4. GSL Data and Experiments

To support the development of the GSL recognizer, we have collected a suitable database. We describe it next, followed by the adopted experimental framework and our GSL recognition experiments on it.

4.1. The GSL Database

Signing data by multiple volunteer informants (both native and non-native in GSL) have been collected to allow isolated GSL recognition of numerals (18-sign vocabulary), isolated SLR of non-numerals (36-sign GSL vocabulary), and continuous recognition of fingerspelled sequences of the 24 Greek alphabet letters.[1]

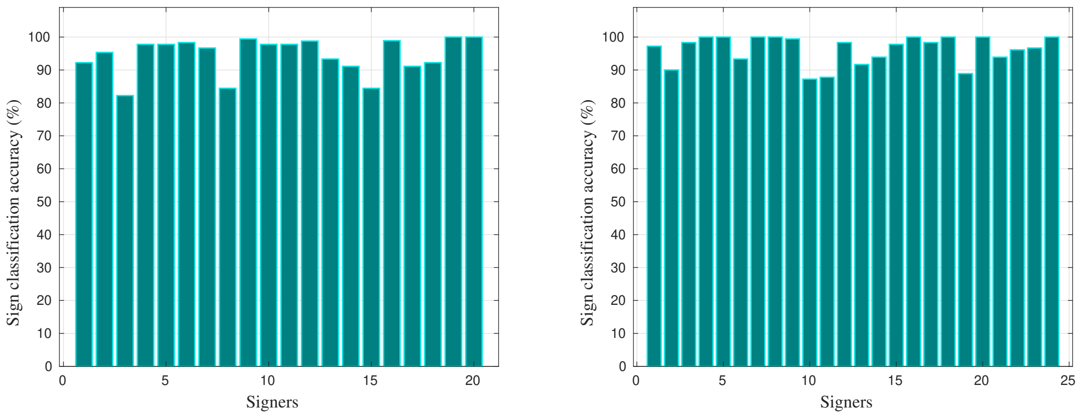
The data have been recorded indoors, under realistic, non-studio conditions with varying background and lighting, using a Logitech C615 web-camera at a frame rate of 30 Hz, YUV411 video format, and 640×480-pixel resolution.

In the case of numeral signs, data from 20 signers have been collected. Each signer articulated the 18 numerals 5 consecutive times, resulting in a total of 1,800 database videos.

In the case of non-numeral signs, data from 17 signers have been collected. Each informant articulated the 36 signs five times. In addition, these data have been supplemented with videos from the publicly available ITI GSL corpus (Adaloglou et al., 2022), resulting to 7 more informants signing the same 36 signs five times. Note that the latter have been recorded using an Intel RealSense D435 RGB-D camera under studio-quality conditions, but here only the RGB stream is utilized. Thus, the combined data contain 24 (17 + 7) signers and a total of 4,320 videos.

Finally, in the case of fingerspelling, data from 12 signers have been recorded. Each informant signed once the 24 Greek alphabet letters in isolation, as well as 50 fingerspelled words (unique to each signer) composed

[1] All informants have signed consent forms, and the data will become publicly available in the future, as part of a larger data release of SL-ReDu project resources.



(a) (b)

Figure 3: SI isolated GSL recognition accuracy (%) per signer for (a) numerals and (b) non-numeral signs.

of 4-5 letters. In addition, 7 informants performed 16 words (common to all) composed of 3-7 letters, and 3 signers expressed an extra 71 words of 4-5 letters. This process resulted in a total of 1,071 videos. Note that each informant has signed each letter at least 4 times.

4.2. Experimental Framework

We are interested primarily in signer-independent (SI) SLR, since learner users of the SL-ReDu platform are typically "unseen" during GSL model training. For comparison purposes, we also report multi-signer (MS) recognition results, where data from all signers are used for both training and test sets (with the sets remaining disjoint), being an easier learning scenario.

In the MS case, we use ten-fold cross-validation. In each fold, we allocate 80% of all videos to training (numerals: 1,440; non-numerals: 3,456; fingerspelling: 857), 10% to validation (numerals: 180; non-numerals: 432; fingerspelling: 107), and the remaining 10% to testing (same number of videos as in validation).

In the SI scenario, we employ 20-fold cross-validation in the numerals case, 24 folds for non-numerals, and 12 ones for fingerspelling. In all cases, each fold contains one test signer, while the model is trained on all others.

In addition to these paradigms, GSL models are also trained to be used by the SL-ReDu platform in its user-evaluation, as reported in Section 5. For this purpose, we allocate 90% of the available videos to training (numerals: 1,620; non-numerals: 3,888; fingerspelling: 964) and the remaining 10% to validation (numerals: 180; non-numerals: 432; fingerspelling: 107).

4.3. Recognition Results

In Table 1, we report the recognition performance of the isolated GSL and continuous fingerspelling tasks on the datasets of Section 4.1, under both MS and SI training/testing paradigms of Section 4.2. Results are reported in word accuracy (WAcc), %, and in the case of fingerspelling in letter accuracy (LAcc), %, as well.

In all cases, performance degrades in the SI case, compared to the MS scenario, which is not surprising. Nevertheless, WAcc remains satisfactory in both isolated SLR tasks (in the 95% WAcc range for SI), showing the potential of utilizing the module in learning platforms like SL-ReDu. Note also that performance varies among signers, as shown in Figure 3 for the isolated tasks in the SI case, remaining nevertheless well above 80% WAcc, even for the worse performing ones.

Concerning continuous fingerspelling, it is natural that performance suffers at the WAcc level, since letter recognition errors (including insertions and deletions) accumulate at the word level, especially for longer letter sequences. This effect is exacerbated due the lack of a language model in the recognizer, as well as the significantly smaller amount of collected data and number of signers compared to the isolated tasks. As expected, LAcc results are higher, but clearly further improvement is needed.

5. User Evaluation of SL-ReDu Platform

We have also conducted a user evaluation of the SL-ReDu platform, producing both objective results (focusing on GSL recognition performance), as well as a subjective assessment based on user responses to a questionnaire.

5.1. Volunteer Users

Two groups of students (with the Department of Special Education at University of Thessaly) and two professional volunteers participated in the preliminary SL-

GSL recog. task	Metric	MS	SI	Eval.
iso. numerals	WAcc	97.78	94.48	98.61
iso. non-numerals	WAcc	99.44	96.20	97.22
cont. fingerspelling	WAcc	75.22	65.30	90.28
	LAcc	86.12	77.66	91.03

Table 1: GSL recognition performance for the various tasks considered here under MS and SI training/testing on the GSL corpus of Section 4.1. Also shown, at the right-most column, is the recognition performance during user evaluation of the SL-ReDu platform (Section 5.2). Results are reported in word accuracy (WAcc, %) or letter accuracy (LAcc, %).

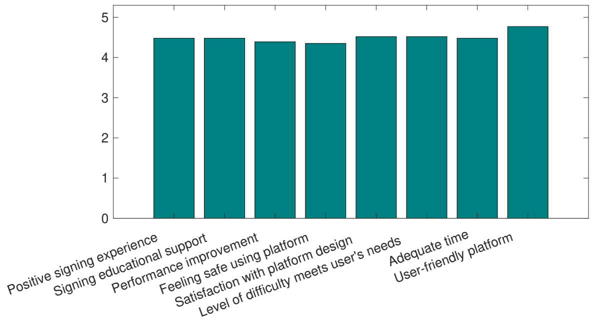


Figure 4: Mean values (on the 1-5 Likert scale) of the platform subjective user assessment along eight aspects.

ReDu evaluation. The first group (10 students) involved true beginners, i.e. university students who had had some contact with GSL for less than five months, the second group (11 students) was made up of students who had recently achieved the target A0-A1 level and had more than five months of experience, and the third group (2 experts) consisted of GSL experts who served as teachers to the student volunteer groups. The demographic characteristics of the student users were consistent with the demographics of the overall student population at the particular department, with ages between 19 and 22 years old and females outnumbering males.

5.2. Objective Evaluation of GSL Recognizer

This evaluation was carried out via "active"-type exercises that require SL production by the learner, captured by a camera and fed to the SL recognition module to provide learner binary feedback. For the isolated GSL recognition of numerals, we incorporated 3 assignments to the platform, each consisting of six GSL production questions of a numeral. For non-numerals we included 6 corresponding six-question production assignments. Finally, for continuous fingerspelling we used 6 six-question assignments that include letters as well as words that do not appear in the training set.

As already mentioned, the system also provides feedback to the user for correct positioning with respect to the camera. Note that participants are allowed to try twice each exercise in case of incorrect positioning feedback. Additionally, "active"-type exams designed by the instructor are automatically graded by the system, while limiting user interaction within prespecified time constraints.

For "active" GSL production and recognition evaluation, a subset of volunteers participated, namely 12 users, including 7 A0 level students, 4 A1 level students, and 1 expert, each performing 3 six-question assignments (one per task, totaling 18 questions).

The objective evaluation results in terms of WAcc (as well as LAcc for fingerspelling) are reported at the right-most column of Table 1. We observe that the results achieved are better than SI recognition performance of the isolated tasks on the collected GSL corpus of Section 4.1. This fact is probably due to the very careful signing and possible over-articulation by the volunteers. The difference is even larger in fingerspelling, due to the additional fact that the corresponding questions involved production of shorter words than those of the GSL corpus.

5.3. Subjective Assessment of the Platform

After signing the relevant consent forms and completing both self-monitoring and GSL production sessions of the SL-ReDu platform, participants were handed an anonymous subjective experience questionnaire that measures eight aspects concerning ease of use, usefulness, design, and user trust on the one-to-five Likert scale. The analysis of the filled subjective experience questionnaires provided valuable input, both in the form of numerical trends and via textual comments. In half (four out of eight) questions of the subjective evaluation the majority of the users provided the highest assessment ("very much"). More specifically, most of the users were completely satisfied with platform design, considered it to be a user-fiendly platform, felt that the level of difficulty meets their needs, and that signing educationally supports them (see also Figure 4 for the mean scores returned).

6. Conclusion

In this paper, we present a GSL recognizer capable of recognizing a medium-size vocabulary of isolated signs and continuously fingerspelled letter sequences, that is integrated in the SL-ReDu learning platform. The recognition module incorporates state-of-the-art deep-learning based visual detection, feature extraction, and classification, and is capable of operating in a signer-independent fashion in non-ideal visual environments. The designed module performs very well, as evidenced by recognition experiments on a suitable dataset collected for this purpose. Further, it yields very satisfactory objective and subjective user evaluation assessment of the SL-ReDu platform.

7. Acknowledgments

H.F.R.I. Hellenic Foundation for Research & Innovation

This research work was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "1st Call for H.F.R.I. Research Projects to support Faculty Members & Researchers and the procurement of high-cost research equipment grant" (Project "SL-ReDu", Project Number 2456).

8. Bibliographical References

Adaloglou, N., Chatzis, T., Papastratis, I., Stergioulas, A., Papadopoulos, G., Zacharopoulou, V., Xydopoulos, G. J., Atzakas, K., Papazachariou, D., and Daras, P. (2022). A comprehensive study on sign language recognition methods. *IEEE Transactions on Multimedia*, 24:1750–1762.

Aran, O., Ari, I., Akarun, L., Sankur, B., Benoit, A., Caplier, A., Campr, P., Carrillo, A. H., and Fanard, F.-X. (2009). SignTutor: An interactive system for sign language tutoring. *IEEE MultiMedia*, 16(1):81–93.

Armstrong, D. F., Karchmer, M. A., and VanCleve, J. V. (2002). *The Study of Signed Languages: Essays in Honor of William C. Stokoe*. Gallaudet University Press, Washington, DC.

Carreira, J., Noland, E., Banki-Horvath, A., Hillier, C., and Zisserman, A. (2018). A short note about Kinetics-600. *CoRR*, arXiv:1808.01340.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255.

Ebling, S., Camgöz, N. C., Braem, P. B., 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. Conference on Language Resources and Evaluation (LREC)*, pages 4221–4229.

Efthimiou, E., Fotinea, S.-E., Flouda, C., Goulas, T., Ametoglou, G., Sapountzaki, G., Papadimitriou, K., and Potamianos, G. (2021). The SL-ReDu environment for self-monitoring and objective learner assessment in Greek Sign Language. In *Proc. Conference on Universal Access in Human-Computer Interaction. Access to Media, Learning and Assistive Environments (HCII)*, volume LNCS-12769, pages 72–81.

Graves, A., Fernández, S., Gomez, F., and Schmidhuber, J. (2006). Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In *Proc. Int. Conference on Machine Learning (ICML)*, pages 369–376.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *CoRR*, arXiv:1704.04861.

Joy, J., Balakrishnan, K., and Sreeraj, M. (2019). SignQuiz: A quiz based tool for learning fingerspelled signs in Indian Sign Language using ASLR. *IEEE Access*, 7:28363–28371.

Kemp, M. (1998). Why is learning American Sign Language a challenge? *American Annals of the Deaf*, 143(3):255–259.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *CoRR*, arXiv:1412.6980.

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M., Lee, J., Chang, W.-T., Hua, W., Georg, M., and Grundmann, M. (2019). MediaPipe: A framework for perceiving and processing reality. In *Proc. Workshop on Computer Vision for AR/VR at IEEE CVPR*.

Mehdi, S. A. and Khan, Y. N. (2002). Sign language recognition using sensor gloves. In *Proc. Int. Conference on Neural Information Processing (ICONIP)*, pages 2204–2206.

Mohammdi, H. M. and Elbourhamy, D. M. (2021). An intelligent system to help deaf students learn Arabic Sign Language. *Interactive Learning Environments*.

Moryossef, A., Tsochantaridis, I., Dinn, J., Camgöz, N. C., Bowden, R., Jiang, T., Rios, A., Müller, M., and Ebling, S. (2021). Evaluating the immediate applicability of pose estimation for sign language recognition. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3429–3435.

Potamianos, G., Papadimitriou, K., Efthimiou, E., Fotinea, S. E., Sapountzaki, G., and Maragos, P. (2020). SL-ReDu: Greek sign language recognition for educational applications. Project description and early results. In *Proc. PErvasive Technologies Related to Assistive Environments Conference (PETRA)*.

Potamianos, G., Papadimitriou, K., Efthimiou, E., Fotinea, S.-E., Goulas, T., Flouda, C., and Ametoglou, G. (2021). SL-ReDu Deliverable D5.2: Technical specifications and system architecture definition. Technical report. [Online:] https://sl-redu.ece.uth.gr/deliverable/Deliverable_D5.2.pdf.

Ren, Z., Yuan, J., and Zhang, Z. (2011). Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera. In *Proc. ACM Int. Conference on Multimedia (MM)*, pages 1093–1096.

Sapountzaki, G., Efthimiou, E., Fotinea, S. E., Papadimitriou, K., and Potamianos, G. (2021). Educational material organization in a platform for Greek Sign Language self monitoring and assessment. In *Proc. Int. Conference on Education and New Learning Technologies (EDULEARN)*, pages 3322–3331.

Schuster, M. and Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the Inception architecture for computer vision. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826.

Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y., and Paluri, M. (2018). A closer look at spatiotemporal convolutions for action recognition. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6450–6459.

Zafrulla, Z., Brashear, H., Presti, P., Hamilton, H., and Starner, T. (2011). CopyCat: An American Sign Language game for deaf children. In *Proc. IEEE Int. Conference on Automatic Face and Gesture Recognition (FG)*.