

Investigating Motion History Images and Convolutional Neural Networks for Isolated Irish Sign Language Fingerspelling Recognition

Hafiz Muhammad Sarmad Khan^{*†} , Irene Murtagh^{*†} , Simon D McLoughlin[†] 

^{*}ADAPT: Centre for AI-Driven Digital Content Technology, [†]Technological University Dublin
Dublin, Ireland

sarmad.khan@adaptcentre.ie, irene.murtagh@adaptcentre.ie, simon.d.mcloughlin@tudublin.ie

Abstract

The limited global competency in sign language makes the objective of improving communication for the deaf and hard-of-hearing community through computational processing both vital and necessary. In an effort to address this problem, our research leverages the Irish Sign Language hand shape (ISL-HS) dataset and state-of-the-art deep learning architectures to recognize the Irish Sign Language alphabet. We streamline the feature extraction methodology and pave the way for the efficient use of Convolutional Neural Networks (CNNs) by using Motion History Images (MHIs) for monitoring the sign language motions. The effectiveness of numerous powerful CNN architectures in deciphering the intricate patterns of motion captured in MHIs is investigated in this research. The process includes generating MHIs from the ISL dataset and then using these images to train several CNN neural network models and evaluate their ability to recognize the Irish Sign Language alphabet. The results demonstrate the possibility of investigating MHIs with advanced CNNs to enhance sign language recognition, with a noteworthy accuracy percentage. By contributing to the development of language processing tools and technologies for Irish Sign Language, this research has the potential to address the lack of technological communicative accessibility and inclusion for the deaf and hard-of-hearing community in Ireland.

Keywords: Motion History Images, Irish Sign Language Recognition, Convolutional Neural Networks

1. Introduction

Sign Languages (SLs), expressed visually through gestures within a three-dimensional signing space, and without a written form serve as the principal mode of communication for numerous deaf and hard-of-hearing communities in their daily interactions. The fact that sign languages are often overlooked by current natural language processing and machine translation technologies exacerbates the existing communication challenges faced by the estimated 72 million deaf individuals worldwide (Murtagh et al., 2022; Murtagh, 2021). Irish Sign Language (ISL) maintains a unique place in the sign language landscape, serving as the principal means of communication for Ireland's deaf and hard-of-hearing community (Leeson and Saeed, 2012). Irish Sign Language (ISL) constitutes a gestural mode of communication devoid of written or spoken articulation. It serves as the primary means of interaction for approximately 5,000 Deaf individuals within Ireland. An additional 40,000 hearing individuals engage with ISL, exhibiting a spectrum of usage frequency from regular to occasional within the country (School of Linguistic, Speech and Communication Sciences, 2016) (Irish Deaf Society). Notwithstanding its cultural and linguistic significance, ISL like numerous other sign languages across the world faced with technologi-

cal constraints due to its visual and spatial properties. Human-Computer Interaction (HCI) is strongly linked to advancements in computer vision, with recognition being a focus for research. The failure to integrate sign languages into modern technologies has hindered the development of accessible information and services for the ISL community, compounded by the challenge of the limited availability of comprehensive datasets for training and evaluating AI models in computational processing. The purpose of this research is to utilize an Irish Sign Language dataset and explore the effectiveness of sophisticated neural network frameworks in recognizing ISL hand motions from motion history images. The efforts are ongoing in the development of a computational system that will automatically annotate sign language data, hence improving communication accessibility and inclusivity for the ISL community.

The paper's outline is structured as follows: Section 2 offers an overview of relevant research in sign language recognition. In Section 3, the proposed methodology is presented, covering the dataset description, data augmentation techniques, and experimental architectures. Section 4 elaborates on the experimental results, and Section 5 culminates in the conclusion.

2. Related Work

In recent years, the evolution of artificial intelligence (AI) and computer vision has fueled dramatic advances in sign language understanding, solving major issues for those with deaf and hard-of-hearing communities (LeCun et al., 2015). This section presents the progress of sign language recognition, with an emphasis on the critical role of deep learning approaches in improving accessibility and communication among the deaf and hard-of-hearing populations.

One of the foundational contributions to this field was proposed by Mathieu De Coster et al. (De Coster et al., 2021), by presenting a novel approach to enhance the performance of the Video Transformer Network (VTN) for isolated sign recognition leveraging multi-modal inputs, including human pose key points and hand crops, extracted from RGB videos. Their adaptation addresses the challenge of limited labeled data available for sign language recognition by enriching the model's input with pre-processed information that captures essential features of sign language, such as hand shapes and body movements. The methodology demonstrated a significant improvement in sign recognition accuracy, achieving 92.92% on the AUTSL dataset, underscoring the potential of combining pose estimation and self-attention mechanisms in deep learning models for more accurate and interpretable sign language recognition. This research was conducted under the SignON project (Sig), funded by the European Union's Horizon 2020. Mathias Müller et al. (Müller et al., 2022), present the inaugural shared task for automatic translation between signed and spoken languages, specifically focusing on Swiss German Sign Language (DSGS) to German and vice versa. This pioneering effort marks a significant departure from the traditional text-to-text machine translation, necessitating the processing of visual information such as video frames or human pose estimation. The task attracted seven teams, all participating in the DSGS-to-German track, showcasing state-of-the-art techniques. Additionally, it generated the first publicly available dataset of system outputs paired with human evaluation scores for sign language translation, thereby setting a foundational benchmark for future research in this emergent field. Neha Deshpande et al. (Deshpande et al., 2022) investigate the use of convolutional neural networks (CNNs) for facial expression recognition in sign language videos, targeting Ekman's six basic expressions (fear, disgust, surprise, sadness, happiness, anger) plus a neutral category. They enhance the performance of pre-trained general facial expression models through fine-tuning, data

augmentation, class balancing, and image pre-processing. Their method, validated using K-fold cross-validation, significantly improves accuracy on sign language datasets, showcasing the effectiveness of CNNs in sign language facial expression recognition and contributing valuable insights to the field. In the pioneering work, Wong et al. (Wong et al., 2022) introduce a novel Hierarchical Sign I3D model (HS-I3D), significantly advancing the field of sign spotting in continuous sign language videos. By innovatively applying a hierarchical spatiotemporal network architecture to learn coarse-to-fine sign features, their approach adeptly captures signs at varying temporal levels, leading to more accurate sign localization. Evaluated on the ChaLearn 2022 Sign Spotting Challenge - MSSL track, the HS-I3D model notably achieved a state-of-the-art 0.607 F1 score, marking it as the competition's top-performing solution. This achievement not only demonstrates the model's effectiveness in identifying and localizing signs with high precision but also emphasizes the utility of incorporating random sampling techniques during model training. (Hsieh et al., 2010) introduced an adaptive approach for hand gesture recognition in human-machine interactions. The novel approach, which integrated an adaptive skin color algorithm with facial recognition algorithms, demonstrated outstanding accuracy even in low-light circumstances and complicated backdrops. This research conducted experiments in which five persons made 250 hand motions at different distances from the webcam. The proposed system demonstrated its practicality and usefulness in real-world applications, with an average accuracy of 94.1% and a processing time of 3.81 milliseconds per frame. This study lays the framework for future advances in sign language recognition algorithms. (Yalçinkaya et al., 2016) highlighted the importance of sign language recognition in improving communication for those with speech and hearing impairments. Their system, which used Motion History Images (MHI) and a nearest neighbor approach, obtained an excellent classification accuracy of 95%, demonstrating the capacity of machine learning to bridge communication gaps. This demonstrates AI's revolutionary influence on increasing accessibility for underserved populations. The implementation of convolutional neural networks (CNNs) has accelerated developments in the recognition of sign language. (Barbhuiya et al., 2021) used CNN frameworks to extract and categorize characteristics in sign language motions, resulting in excellent classification accuracy. Using pre-trained CNN models like "AlexNet" and "VGG-16," they demonstrated the usefulness of deep neural networks in practical applications of sign language recognition systems. Quantitative evaluations demonstrate the efficacy of the

CNN-based method, with the model achieving high accuracy rates in sign language categorization of 99% when using random validation and 70% when utilizing leave-one-out validation. Simultaneously, (Wadhawan and Kumar, 2020) made significant advances in deep learning-based CNNs for sign language identification by representing static signs. Through testing and analysis, they were able to get exceptional training precision, outperforming earlier methods and creating new opportunities for identifying a wider variety of hand signals. The suggested approach achieved remarkable training accuracy of 99.90% and 99.72%, respectively. This demonstrates how AI-driven methods for sign language processing are evolving and improving. Bantupalli et al. (Bantupalli and Xie, 2018) proposed an innovative technique to address communication challenges encountered by individuals who have speech impairments. Their research focuses on the creation of a vision-based application for sign language translation into text. Using current advances in deep learning and computer vision, they extracted important temporal and spatial information from video sequences. They specifically used Inception to recognize spatial features and a Recurrent Neural Network to analyze temporal data. The experiment yielded good results, with an average accuracy of 90% with the softmax layer and 55% with the pooling layer. The study emphasizes the transformative potential of technology-driven solutions in overcoming societal difficulties, as well as the significance of interdisciplinary collaboration in fostering social innovation. R.S. Sabeenian et al. (Sabeenian et al., 2020) investigated the challenges linked to speech impairment affecting communication via speech and hearing. Despite the growing usage of sign language as an alternate communication tool, non-signers continue to face a hurdle in communicating effectively with signers. Making use of recent advances in computer vision and deep learning, the authors concentrated on creating a deep learning-based application for translating sign language into text. Their method used a proprietary Convolutional Neural Network (CNN) to recognize signs in video frames, with the MNIST dataset used for model training. The constructed model attained 93% accuracy, indicating its usefulness in sign language identification and translation. Dongxu Li et al. (Li et al., 2020) developed the Word-Level American Sign Language (WLASL) (Li, 2020) video collection, which includes over 2000 words performed by 100+ signers, to overcome the limitations of existing sign language datasets. Their research enabled testing with deep learning methods for word-level sign identification, contrasting holistic visual appearance-based and 2D human pose-based approaches. Furthermore, they suggested a novel Pose-based Temporal Graph

Convolution Networks (Pose-TGCN) method to improve pose-based recognition. Both approaches produced equivalent results, with up to 62.63% top-10 accuracy on 2000 words/glosses, demonstrating the dataset's importance in improving sign language recognition research.

In spite of these developments, there is still a significant shortfall of research using this approach for Irish Sign Language (ISL). By filling up this gap, future research will have the chance to improve accessibility and inclusion for members of the ISL community.

3. Proposed Methodology

This section covers the architectures used for sign language recognition, as well as the dataset utilized and data augmentation aspects.

3.1. Dataset

In this research, we utilized the Irish Sign Language hand-shape dataset (ISL-HS) (Oliveira et al., 2017), which consists of real hand images. The ISL-HS dataset consists of 23 static gestures representing English alphabet signs and three dynamic motions (J, X, and Z). The recording method was led by the dataset documentation, with six people (three men and three women) practicing fingerspelling of the ISL alphabet, with each action recorded three times. The videos were captured at 30 frames per second (fps) and 640x480 pixels, for a total of 468 recordings. From these films, 52,688 frames were retrieved for static forms and 5,426 frames for dynamic motions, for a total of 58,114. We used both static and dynamic form images in this investigation.

3.2. Preprocessing and Augmentation Strategies

3.2.1. Data PreProcessing

In our approach to the Irish Sign Language hand shape dataset, the first stage entailed converting the dataset into motion history images (MHIs). To accomplish this, we employed a tailored Python script that made use of computer vision libraries like OpenCV. This script systematically traversed through the sequence of images, computing the absolute variance between successive frames. Subsequently, it updates the Motion History Image by considering a predefined motion threshold and persistence parameter. The resultant MHI effectively encapsulates the temporal motion information, ultimately yielding the final MHI image. Following the generation of MHI, the dataset now comprises a total of 18 images for each letter of the alphabet, culminating in a robust set of 468 images. Figure 1

showcases the original dataset image and its corresponding MHI representation. The equation for calculating the Motion History Image (MHI) is given by:

$$MHI(x, y) = \begin{cases} \tau & \text{if } |I_t(x, y) - I_{t-1}(x, y)| > mv_thresh \\ MHI(x, y) - 1 & \text{otherwise} \end{cases} \quad (1)$$

where $MHI(x, y)$ represents the pixel value at location (x, y) in the MHI, $I_t(x, y)$ is the pixel value at location (x, y) in the current frame t , $I_{t-1}(x, y)$ is the pixel value at location (x, y) in the previous frame $t - 1$, τ is the persistence parameter, and mv_thresh is the motion threshold parameter.

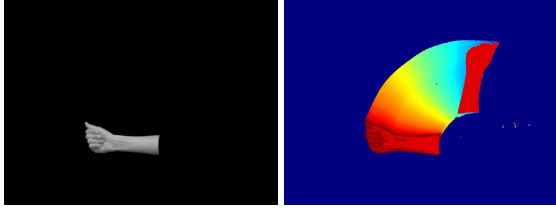


Figure 1: Left: Original image from the dataset. Right: Motion History Image (MHI) representation.

3.2.2. Data Augmentation

After developing the motion history images (MHIs), we employed data augmentation to expand the dataset and improve the robustness of frameworks. Due to the scarcity of the dataset, we chose data augmentation via synthesis (Keskin, 2023; Jo et al., 2017). This process involved developing new iterations of the images using various augmentations. The goal was to incorporate unpredictability into the dataset, which would improve the model’s ability to generalize to new data and increase its effectiveness in real-world applications. Specifically, augmentations such as additive Gaussian noise and multiplication were used to bring variation into the images. Additionally, affine transformations such as scaling and zooming were used to imitate changes in viewpoint and orientation. By applying various augmentation techniques to the 468 original images, we generated new iterations of these images, significantly expanding our dataset to a total of 7020 images. This increase has resulted in a substantial enhancement, providing 270 images for each class, thereby reinforcing the robustness of our frameworks. Of these images, 80% were designated for training and validation while the remaining 20% were reserved for testing in order to increase the resilience of the framework. These augmentations contributed to a more diversified and complete dataset, allowing for more efficient training of models and improved sign language recognition performance. Overall, data augmentation was critical in improving the quality and variety of our dataset,

Description	Total Images	Images per Class
Original Images	58112	2235
Post-MHI Images	468	18
Post-Augmentation Images	7020	270
Training Set	4420	170
Validation Set	1196	46
Test Set	1404	54

Table 1: Dataset Overview: Table presents key metrics at various processing stages, including total images and images per class.

resulting in enhanced model performance and generalization capabilities for detecting Irish Sign Language motions. Figure 2 illustrates examples of augmented images representing each letter in MHI format and the dataset composition, detailing total images, images per class, and the train-test split at different processing stages, is summarized in Table 1.

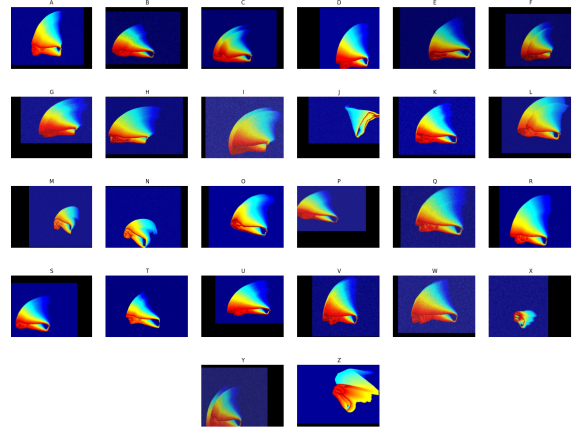


Figure 2: Sample Augmented Image for each Letter

3.3. Proposed Architectures

A multitude of deep learning algorithms exist, offering various capabilities and applications. However, among these algorithms, Convolutional Neural Networks stand out prominently in the field of computer vision. Researchers often choose CNNs for image classification tasks due to their ability to effectively analyze images as input and output probability values or class labels (Putzu et al., 2020). This capability makes CNN architectures highly suitable for addressing challenges in picture categorization, such as identifying objects, patterns, or gestures within images. In our research, we leverage the power of CNN architectures for sign language recognition

using motion history images. As illustrated in Figure 3, the MHI sign language recognition pipeline encompasses a sequence of stages from collective dataset acquisition, through preprocessing and data augmentation, to feature extraction state-of-the-art CNN networks, and finally to recognition and probability testing yielding the results. As shown in Table 2, we explore a variety of CNN architectures tailored to the unique properties of motion history images, aiming to enhance the accuracy and robustness of the Irish Sign Language recognition system.

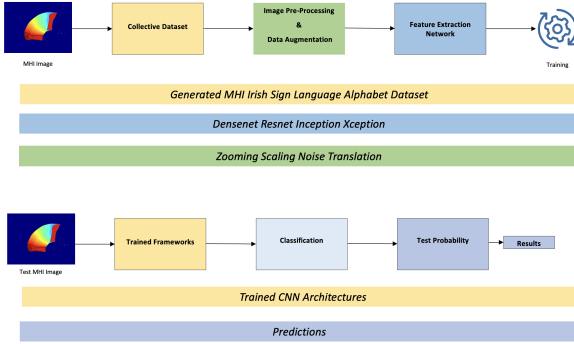


Figure 3: Flowchart depicting the training and testing pipeline for an Irish Sign Language recognition system using CNN architectures, with the top path illustrating the training phase on augmented MHI data, and the bottom path showing the testing phase leading to predictions.

Model	Architecture	Parameters
Resnet	50-V2	25.6 M
	101-V2	44.7 M
	152-V2	60.4 M
Xception	Xception	22.9 M
Densenet	121	8.1 M
	169	20.2 M
	201	14.3 M
Inception	V3	23.9 M
	Resnet-V2	55.9 M

Table 2: Frameworks utilized and Parameters

4. Experimental Results

4.1. Model Hyperparameters

TensorFlow and Keras libraries are employed to implement deep learning architectures. Additionally, image augmentation techniques are utilized, with the training epoch set to 20. Fine-tuning, a crucial step in model optimization, is also performed to adapt the pre-trained models. This process involves adjusting the parameters of the pre-trained

models to better suit the characteristics of the target dataset, thereby enhancing performance. A learning rate of 0.001 is utilized, along with a batch size of 16, and optimization is achieved using the ADAM optimizer with the "categorical cross-entropy" loss function. The softmax activation function is applied to the models. To expedite processing, the resolution is standardized to 160x120. Python scripts are executed on Google Colab, leveraging the Tesla K80 GPU for enhanced computational efficiency.

4.2. Evaluation Metrics

Accuracy, precision, recall, and F1-Score were used in the appraisal of the framework's performance.

Recall is the measure of how the model and algorithm predict True positives:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

where TP stands for True Positive and FN stands for False Negative.

Precision is determined by the ratio of properly identified true negative samples to the total number of outcomes, which includes both true negative and false positive results:

$$\text{Precision} = \frac{TN}{TN + FP} \quad (3)$$

where TN stands for True Negative and FP stands for False Positive.

The accuracy of the model is determined by the proportion of its predictions that are confirmed by testing:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

where TP, TN, FP, and FN have the same meanings as above.

F1-Score is a technique that is used to combine the accuracy and recall of the model and is also the harmonic combination of the model's recall and precision:

$$\text{F1 Score} = \frac{2TP}{2TP + FP + FN} \quad (5)$$

where TP, FP, and FN have the same meanings as above.

4.3. Results

The findings of our research show that different deep learning models for motion history image (MHI)-based sign language recognition have differing accuracy levels. Table 3 shows that

Densenet 121 attains the greatest accuracy of 90.38%, Densenet 201 closely behind with 90.10%, and Densenet 169 with 89.60%. Of all the examined frameworks, Densenet continually showed the optimal accuracy rates. The ability of Densenet frameworks to reliably identify Irish Sign Language alphabets from MHIs is demonstrated by this. Although Densenet architectures showed remarkable performance, other models also yielded encouraging outcomes. With ResNet variations (ResNet 101 V2, ResNet 50 V2, and ResNet 152 V2) varied from 82.24% to 85.62%, Xception attained an accuracy of 80.56%. Furthermore, 77.64% and 75.76% accuracy were attained using Inception ResNet V2 and Inception V3, respectively. These results imply that different deep learning architectures could potentially be applied successfully to sign language challenges. Notably, we maintained the same amount of hyperparameters for each model, including epoch, batch size, loss function, and optimizer. However, the framework’s particular needs and limitations should be taken into account while selecting a model architecture. Densenet frameworks, for example, have the maximum accuracy, but alternative models could offer a fair trade-off between accuracy and computing economy, which would make them more appropriate for some deployment scenarios. As demonstrated in Figure 4, the confusion matrix showcases the performance of the Densenet 201 framework across all classes, with accuracy observed along the diagonal where predicted values coincide with actual values.

Model	Accuracy
Densenet 121	90.38
Densenet 169	89.60
Densenet 201	90.10
Xception	80.56
Resnet 101 V2	85.62
Resnet 50 V2	82.84
Resnet 152 V2	82.24
Inception Resnet V2	77.64
Inception V3	75.76

Table 3: Evaluation Metrics

5. Conclusion

In conclusion, our research offers valuable insights into the application of deep learning methodologies for sign language recognition, particularly leveraging motion history images (MHIs). Despite the challenges posed by limited datasets, our research highlights the effectiveness of data augmentation techniques in enhancing model performance. By evaluating various state-of-the-art architectures on an

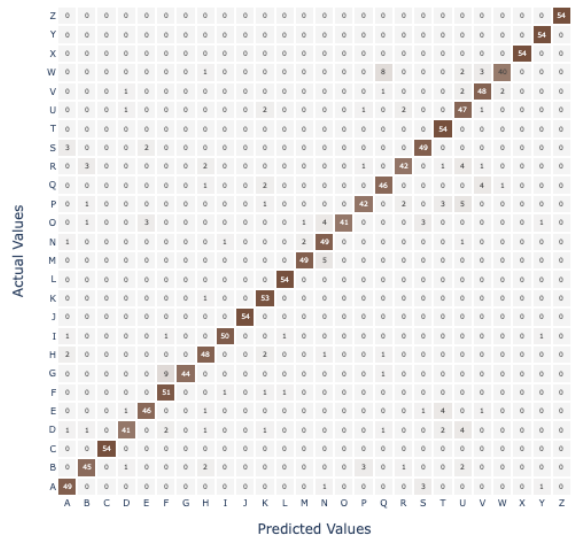


Figure 4: Confusion matrix illustrating the performance metrics on the ISL-HS test dataset

Irish Sign Language motion history images dataset, we have identified promising avenues for improving accessibility and inclusivity for the deaf and hard-of-hearing community. Looking ahead, we envision delving into more intricate challenges, such as sign language annotation and the development of automated annotation pipelines. Through further investigation into the efficiency of deep learning frameworks in computer vision, we aim to narrow the technological gap in sign language recognition, thus contributing to advancements in accessible communication technologies.

6. Acknowledgments

We would like to sincerely thank all the organizations who helped make this research project a reality. We are grateful for the financial assistance from the Science Foundation Ireland Research Centre for AI-Driven Digital Content Technology, which allowed us to carry out this research. Finally, we would like to express our gratitude to the anonymous reviewers whose insightful critiques and recommendations have greatly raised the caliber of this work.

7. Bibliographical References

Signon project. <https://signon-project.eu/>.

Kshitij Bantupalli and Ying Xie. 2018. American sign language recognition using deep learning and computer vision. pages 4896–4899.

- Abul Abbas Barbhuiya, Ram Kumar Karsh, and Rahul Jain. 2021. Cnn based feature extraction and classification for sign language. *Multimedia Tools and Applications*, 80:3051–3069.
- Mathieu De Coster, Mieke Van Herreweghe, and Joni Dambre. 2021. Isolated sign recognition from rgb video using pose flow and self-attention. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3441–3450.
- Neha Deshpande, Fabrizio Nunnari, and Eleftherios Avramidis. 2022. [Fine-tuning of convolutional neural networks for the recognition of facial expressions in sign language video samples](#). In *Proceedings of the 7th International Workshop on Sign Language Translation and Avatar Technology: The Junction of the Visual and the Textual: Challenges and Perspectives*, pages 29–38, Marseille, France. European Language Resources Association.
- Chen Chiung Hsieh, Dung Hua Liou, and David Lee. 2010. A real time hand gesture recognition system using motion history image. In *2010 2nd International Conference on Signal Processing Systems (ICSPS)*, volume 2. IEEE.
- Irish Deaf Society. Irish sign language. <https://www.irishdeafociety.ie/irish-sign-language/>.
- Hyunjun Jo, Yong Ho Na, and Jae Bok Song. 2017. Data augmentation using synthesized images for object detection. In *International Conference on Control, Automation and Systems*, pages 1035–1038.
- Doğan Keskin. 2023. [Synthetic data and data augmentation | by doğan keskin | medium](#).
- Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. [Deep learning](#). *Nature*, 521:436–444.
- Lorraine Leeson and John I. Saeed. 2012. *Irish Sign Language: A Cognitive Linguistic Approach*. Publisher Name.
- Dongxu Li. 2020. [Wlasl: A dataset for word-level american sign language](#).
- Dongxu Li, Cristian Rodriguez, Xin Yu, and Hongdong Li. 2020. Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison.
- Mathias Müller, Sarah Ebling, Eleftherios Avramidis, Alessia Battisti, Michèle Berger, Richard Bowden, Annelies Braffort, Necati Cihan Camgöz, Cristina España-bonet, Roman Grundkiewicz, Zifan Jiang, Oscar Koller, Amit Moryossef, Regula Perrollaz, Sabine Reinhard, Annette Rios, Dimitar Shterionov, Sandra Sidler-miserez, and Katja Tissi. 2022. [Findings of the first WMT shared task on sign language translation \(WMT-SLT22\)](#). In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 744–772, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.
- Irene Murtagh. 2021. The nature of verbs in sign languages: A role and reference grammar account of irish sign language verbs.
- Irene Murtagh, Víctor Ubieta Nogales, and Josep Blat. 2022. Sign language machine translation and the sign language lexicon: A linguistically informed approach. In *Proceedings of the 15th biennial conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, pages 240–251.
- Marlon Oliveira, Houssein Chatbri, Ylva Ferstl, Mohamed Farouk, Suzanne Little, Noel E O'Connor, and Alistair S Sutherland. 2017. [A dataset for irish sign language recognition](#).
- Lorenzo Putzu, Luca Piras, and Giorgio Giacinto. 2020. Convolutional neural networks for relevance feedback in content based image retrieval: A content based image retrieval system that exploits convolutional neural networks both for feature extraction and for relevance feedback. *Multimedia Tools and Applications*, 79:26995–27021.
- RS Sabeenian, S Sai Bharathwaj, and M Mohamed Aadhil. 2020. Sign language recognition using deep learning and computer vision. *J Adv Res Dyn Control Syst*, 12(5 Special Issue):964–968.
- School of Linguistic, Speech and Communication Sciences. 2016. What is irish sign language and who uses it? <https://www.tcd.ie/slscs/faqs/irish-sign-language/>.
- Ankita Wadhawan and Parteek Kumar. 2020. Deep learning-based sign language recognition system for static signs. *Neural Computing and Applications*, 32:7957–7968.
- Ryan Wong, Necati Cihan Camgöz, and Richard Bowden. 2022. Hierarchical i3d for sign spotting. In *European Conference on Computer Vision*, pages 243–255. Springer.
- Özge Yalçinkaya, Anil Atvar, and Pinar Duygulu. 2016. Hareket geçmişi görüntüsü yöntemi ile türkçe işaret dilini tanıma uygulaması. In *2016 24th Signal Processing and Communication Application Conference (SIU)*, pages 801–804. IEEE.