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

Hafiz Muhammad Sarmad Khan
SFI: ADAPT Centre for AI-Driven Digital Content Technology
sarmad.khan@adaptcentre.ie

Simon D McLoughlin
Technological University Dublin
Simon.D.McLoughlin@tudublin.ie

Irene Murtagh
SFI: ADAPT Centre for AI-Driven Digital Content Technology
Irene.murtagh@adaptcentre.ie

Introduction

- The research aims to improve communication for the deaf and hard-of-hearing community through computational processing.
- Using the ISL-HS dataset and state-of-the-art deep learning architectures, we focus on recognizing the Irish Sign Language (ISL) fingerspelling alphabet.
- Employed the Motion History Images (MHIs) for monitoring sign language motions and investigate the effectiveness of various CNN architectures in deciphering intricate motion patterns.

Problem Statement

- Enhance comprehension in Irish Sign Language communication by developing an efficient system for automatically annotating sign language.
- Leveraging advanced artificial intelligence technologies as our primary objective.
- Initial focus on Irish Sign Language recognition.

Proposed Methodology

Dataset & Pre-processing:

- We utilize the Irish Sign Language hand-shape dataset (ISL-HS) containing static gestures and dynamic motions representing the English alphabet.
- Pre-processing involves converting raw video frames into Motion History Images (MHIs) to highlight temporal differences and data augmentation.

Model Development:

- We utilize Convolutional Neural Networks (CNNs) as the core methodology.
- Experimented with various advanced CNN architectures such as Densenet, Xception, Resnet, and Inception models known for their image analysis capabilities.
- These architectures are chosen for their ability to discern the nuanced patterns of hand movements essential for ISL recognition.

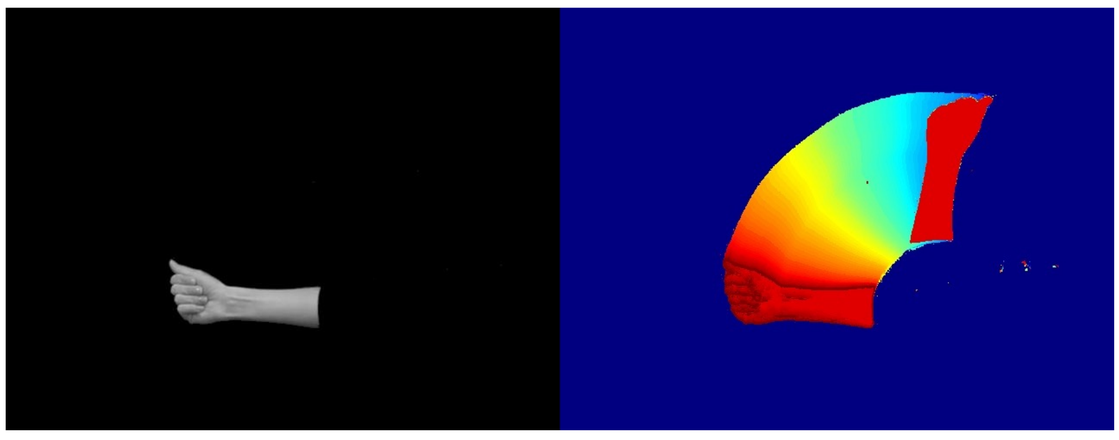

Figure 1: Original Handshape Image from ISH-HS (Left) and Generated MHI (Right)

Table 1: Dataset Overview: Original, processed, and augmented images distribution.

Description	Total Images	Images Per Class
Original Images	58112	2235
Post MHI Images	468	18
Post-Augmented Images	7020	270
Training Set	4420	170
Validation Set	1196	46
Testing Set	1404	54

Performance Evaluation

- Our evaluation of CNN architectures on the augmented ISL-HS dataset demonstrates promising results.
- Affirms the effectiveness of deep learning in sign language recognition.
- Densenet models, particularly Densenet 121, outperform other architectures.

Table 3: Confusion Matrix for Densenet 121

Confusion Matrix (True Positive Labels)									
Class	**TP**	**Class**	**TP**	**Class**	**TP**	**Class**	**TP**	**Class**	**TP**
A	49	**G**	49	**M**	52	**S**	32	**Y**	51
B	49	**H**	37	**N**	29	**T**	38		
C	50	**I**	52	**O**	53	**U**	43		
D	36	**J**	54	**P**	42	**V**	42	**Z**	54
E	51	**K**	42	**Q**	43	**W**	38		
F	44	**L**	49	**R**	34	**X**	54		

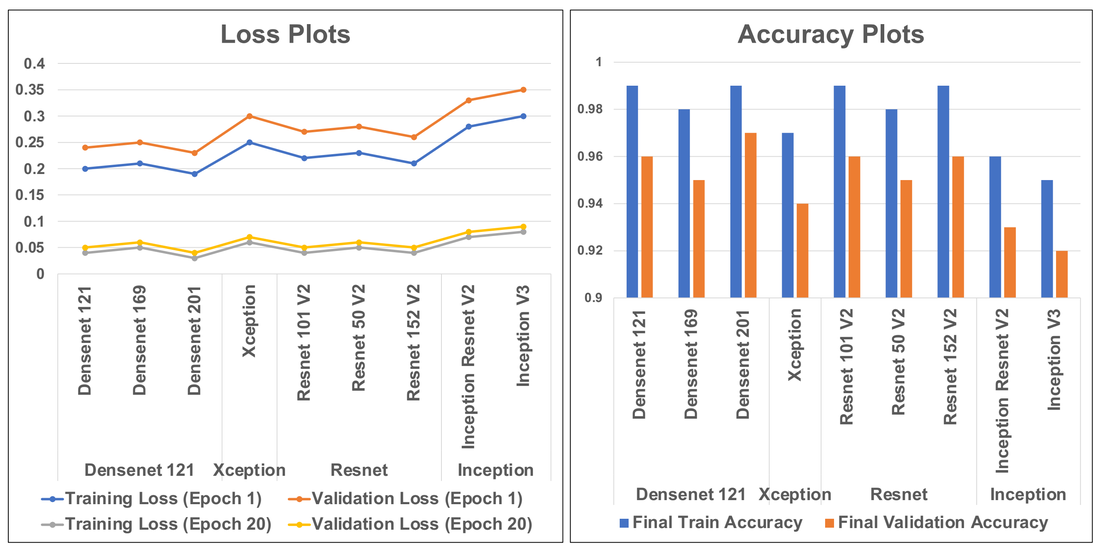

Figure 2: Performance evaluation showing Loss Plots (Left) and Accuracy Plots (Right).

Table 3: Evaluation Metrics

Model	Architecture	Accuracy
Densenet	121	90.38
	169	89.60
	201	90.10
Resnet	50 V2	81.84
	101 V2	81.62
	152 V2	82.34
Xception	Xception	80.56
Inception	V3	75.76
	Resnet V2	77.64

Conclusion

- The research aims to improve communication for the deaf and hard-of-hearing community through computational processing.
- Using the ISL-HS dataset and state-of-the-art deep learning architectures, we focus on recognizing the Irish Sign Language fingerspelling alphabets.
- Employed the Motion History Images (MHIs) for monitoring sign language motions and investigate the effectiveness of various CNN architectures in deciphering intricate motion patterns.

References

- L. Leeson and J. I. Saeed, Irish Sign Language: A Cognitive Linguistic Approach. Edinburgh University Press, 2012, pp. 256. [Online]. Available: https://www.jstor.org/stable/10.3366/j.ctt3fgrr5.
- M. Oliveira et al., "A Dataset for Irish Sign Language Recognition," Dublin City University, Ireland; Trinity College Dublin, Ireland; Arab Academy for Science & Technology, Egypt. [Online]. Available: https://core.ac.uk/download/pdf/95876247.pdf.

This research was conducted with the financial support of Science Foundation Ireland at ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology at TU Dublin [13/RC/2106_P2].