Enhancing Syllabic Component Classification in Japanese Sign Language by Pre-training on Non-Japanese Sign Language Data

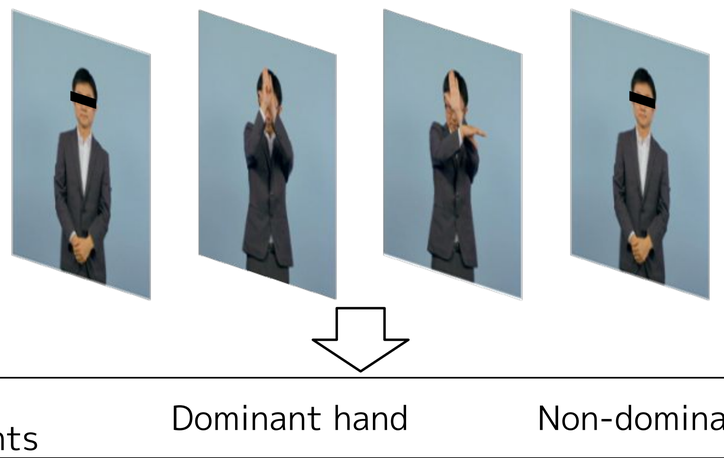
Jundai Inoue, Makoto Miwa, Yutaka Sasaki, and Daisuke Hara
Toyota Techonological Institute, Japan
{sd24410, makoto-miwa, yutaka.sasaki, daisuke}@toyota-ti.ac.jp

Introduction

Background

- Syllables of the sign language are combinations of syllabic components, and the composition rules for the syllables are still unclear. [1]
  - Locations, movements, and handshapes are the syllabic components

⇒ We want to decompose a number of syllables into syllabic components to analyze the rules, but manual decomposition is costly.



Objective

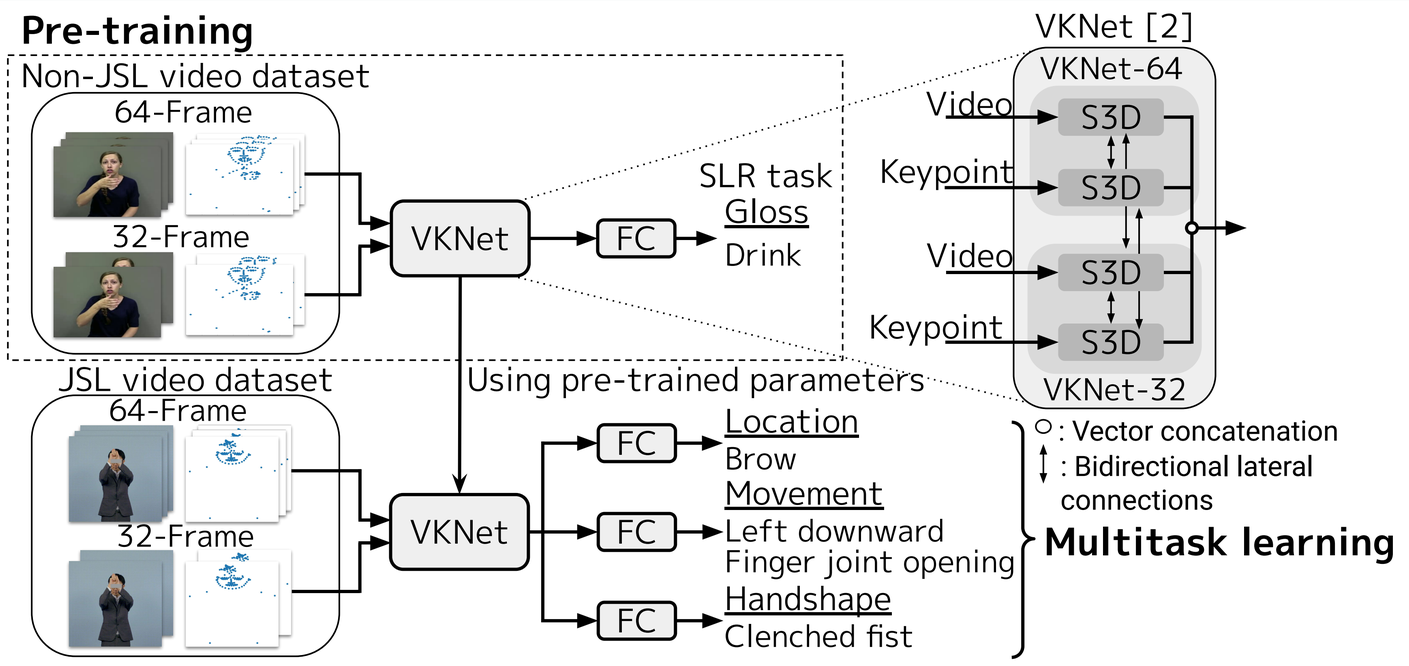
- This study aims to construct an automatic syllabic component classification system for Japanese Sign Language (JSL) .
  - As the first step toward this goal, this study focuses on the location, movements, and handshape of the dominant hand.
  - The number of JSL videos with labeled syllabic components is limited

Syllabic components	Dominant hand	Non-dominant hand
Location	Brow	Brow
Contact	Non-dominant palm	Brow
Movements	Left downward	-
	Finger joint opening	-
Handshape	Clenched fist	Flat hand
hand orientation	Left	Forward

Method

We propose a method for classifying syllabic components in JSL videos using pre-training on a non-JSL dataset.

- Classifying syllabic components in JSL videos
  - Multiclass classification for the location and handshape components
  - Multilabel classification for the movement component
- Pre-training VKNet [2] on a large amount of non-JSL data to address the problem of **limited data in JSL**
- Multitask learning of syllabic components to share the information among syllabic components



Experiments & Discussion

Experimental settings

- Syllable database in JSL [3]
  - 1,072 syllable videos recorded with a single signer
  - 22, 55, and 69 categories for location, movement, and handshape components

Movement	#	Handshape	#	Location	#
Rightward movement of a hand	142		138	*	835
Forward movement of a hand	135		125	Temples	40
Wrist rotation: outward rotation of a wrist	120		57	Mouth	32
Downward movement of a hand	117		55	Chest	23

- Pre-training dataset: WLASL [4]
  - 2,000 glosses for Sign Language Recognition (SLR)
- Evaluation: micro-F score

	Train	Dev	Test
Syllable datasbase	750	161	161
WLASL	14,289	3,916	2,878

Results

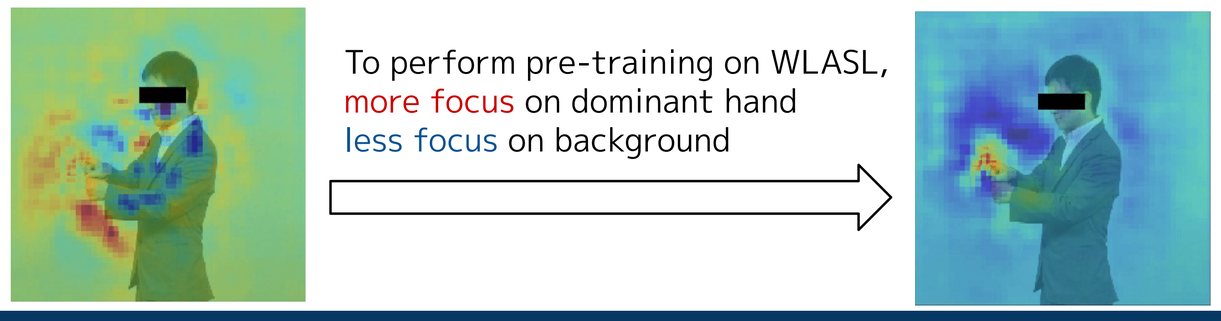
- The pre-training method improved the performance of movement and handshape components.
- Multitask learning was ineffective or harmful in classifying syllabic components of JSL.

Method	Location	Movement	Handshape
VKNet	80.33 (±1.06)	38.55 (±1.25)	35.20 (±2.55)
+ Pre-training	81.16 (±2.05)	**52.41** (±0.86)	**44.72** (±3.55)
+ Multitask learning	81.99 (±0.00)	45.76 (±0.82)	42.23 (±1.34)

Discussion

- To verify the influence of pre-training, we visualized the part of the video VKNet focused on, using AOSA [5].

⇒ VKNet focused on the dominant hand due to pre-training on ASL.



Conclusions & Future Work

Conclusions

- Objective: Constructing a syllabic component classification system based on JSL videos using deep learning.
- Method: Syllabic component classification for the dominant hand using pre-training on ASL data and multitask learning
- Results: Pre-training improved performance of movement and handshape components, but multitask learning was not effective

Future Work

- Classify syllabic components for both the dominant and non-dominant hands
- Examine models and training methods to improve classification performance

References

[1] Hara. An information-based approach to the syllable formation of Japanese Sign Language. 2016
[2] Zuo et al. Natural language-assisted sign language recognition. CVPR, 2023
[3] Hara. New Japanese Sign Language Coding Manual. (In Japanese). 2019
[4] Li et al. Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. WACV, 2020
[5] Uchiyama et al. Visually explaining 3d-cnn predictions for video classification with an adaptive occlusion sensitivity analysis. WACV, 2023