Searching for fingerspelled content in American Sign Language is a paper included in the proceedings of the 60th annual meeting of the Association for Computational Linguistics. It was written by Bowen Shi, Greg Shakhnarovich, and Karen Livescu, who are affiliated with the Toyota Technological Institute at Chicago, and Diane Brentari, who is affiliated with the University of Chicago.

Sign languages are a type of natural language that vary significantly from other natural languages because instead of being conveyed through speech and text, it is conveyed through hand gestures and signs and facial expressions to some degree. This raises unique challenges to natural language processing. A major aspect of sign language is fingerspelling, which is used in conjunction with other gestures to convey words that do not have an established gesture such as the name of a place or person. Despite how important fingerspelling is to tasks involving sign language, there has not been any research on the issue before this paper.

Previous works regarding natural language processing in sign language have all focused on the recognition and automatic transcription of lexical signs. The methods used are not effective for tasks involving fingerspelling. Compared to the identification of lexical signs, fingerspelling search is more difficult because of the requirement that it must be able to handle an open vocabulary and queries of arbitrary lengths. This requirement forces even more new ground to be broken in the sense that previous studies only relied on data from controlled environments with set vocabularies.

The solution this study introduces is Finger Spelling Search Network (FSS-Net). FSS-Net detects fingerspelling from videos of signing in natural contexts and matches clips to text

queries. It does this by taking raw video clips and written text sequences as input and outputting a score that indicated how well the inputs match each other. After this, it encodes the text input and generates fingerspelling segment proposals. The proposals are then encoded into a shared space with the encoded text where they will be paired together.

The authors of this study evaluated their model on how well it performed two tasks: Fingerspelled Word Search (FWS) and Fingerspelling-based video search (FVS). FWS is the task of finding which words are in a video clip while FVS is the task of finding all videos that have a word. The metric by which they judged these tasks on are mean average precision and mean F1 score which are standard for video-text retrieval.

On Google Scholar, Bowen Shi has received 465 citations. His work, which involves audio-visual speech processing, is important in developing machine learnings ability to communicate with users. Diane Brentari has 5802 citations and her work as a director of the Center for Gesture, Sign and Language and the Sign Language Linguistics Lab is important in understanding the modality of languages. Karen Livescu has 10572 citations. Her work in speech and language processing is important to the development of machine learning that can operate in different languages. Of these authors, Greg Shakhnarovich is the most highly cited at 12765 citations. His work, which has a focus on computer vision, is important because it contributes to the effectiveness of machine learning in visual tasks, which is very important in its integration to many tasks in the real world. These author's shared interest in the task of processing sign language makes their work important in making machine learning technology more accessible.

References

- Bowen Shi, Diane Brentari, Greg Shakhnarovich, and Karen Livescu. 2022. Searching for fingerspelled content in American Sign Language. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1699–1712, Dublin, Ireland. Association for Computational Linguistics.