Accessible Narratives: Storytelling Strategies for Hearing-Impaired Children

[1]Nithiksha Patel G S , MCA student, PES Institute of Technology and Management Studies, Shivamogga, Karnataka, India.

[2]Ms. Kavya H V, Assistant Professor, MCA, PES Institute of Technology and Management Studies, Shivamogga, Karnataka, India

Abstract

Deep learning has showed potential in helping assistive communication systems bridge the gap between sign language and spoken/written language. This work introduces a novel framework for Story-to-Sign creation, which uses deep learning algorithms to convert textual narratives into animations in sign language. The suggested system uses sequence-to-sequence architectures to translate text into gloss, which is a written representation of sign language, after using natural language processing (NLP) models to comprehend the input text semantically. A later generative model uses pose estimation or avatar-based animation to create comparable sign language gestures. Curated datasets that combine written narratives with matching sign language annotations are used to train and assess the model. The system's ability to maintain narrative coherence while producing precise and contextually relevant sign sequences is demonstrated by the results. This method lays the groundwork for further studies in multimodal language translation while promoting accessibility for the Deaf and hard-of-hearing groups.

Keywords: Hearing impairment, Natural language processing, Application, Storytelling, Sign Language , Child Development, Deaf ,Education.

1. Introduction

Barriers to communication between the hearing and Deaf cultures still pose serious problems for everyday interactions, media access, and education. For millions of Deaf people throughout the world, sign languages—visual-manual languages with their own grammar and syntax—are their main form of communication. Access to accurate and real-time translation of spoken or written content into sign language is still restricted, even with advancements in computer vision and natural language processing (NLP). This is especially true when it comes to storytelling and instructional materials.

For both children and adults, storytelling is essential to language development, cultural transmission, and learning. Stories in sign language improve inclusivity, engagement, and comprehension for Deaf and hard-of-hearing audiences. Human interpreters must manually translate text into sign language, which is expensive, time-consuming, and not always possible. As a result, deep learning-based process automation offers a viable and expandable option.

In order to translate textual stories into sign language animations, this study investigates a deep learning-based system for Story-to-Sign creation. The suggested system consists of many steps: first, the input text is understood in natural language, then it is translated into sign language gloss, and last, using pose synthesis or avatar animation, corresponding sign movements are generated. By utilizing developments in transformers, generative neural networks, and sequence-to-sequence models, this method aims to produce expressive and semantically correct sign language representations.

In addition to advancing the science of sign language translation, our initiative aims to make storytelling more inclusive and accessible for Deaf communities. This study shows how deep learning can be utilized to bridge linguistic barriers through multimodal translation technologies by outlining the system architecture, training technique, datasets used, and experimental findings.

2. Literature survey

Sign language recognition from videos using manually created features and traditional machine learning approaches was the main focus of early sign language translation research. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been used to identify indications from spatial-temporal data since the development of deep learning (Camgoz et al., 2017). Sequence modeling for the creation of sign gloss from spoken or written language was enhanced by the use of Long Short-Term Memory (LSTM) networks and later Transformer-based models (Vaswani et al., 2017).

Gloss is a textual version of sign language that is frequently used as a bridge. A neural sign translation system was presented by Camgoz et al. (2018). It converts spoken words into gloss sequences and subsequently into sign language films. The RWTH-PHOENIX-Weather corpus, which they introduced, has since become a standard dataset for these kinds of applications. An end-to-end transformer model for gloss prediction and translation was also presented by Saunders et al. (2020), who demonstrated notable advancements over RNN-based methods.

Avatar-based and skeletal pose-based sign creation were investigated in a number of studies to visually communicate signs. Stoll et al. (2020) suggested animating sign language straight from gloss or gloss-aligned sequences utilizing 2D and 3D human pose estimate models. In order to create realistic sign gestures from text or gloss inputs, Generative Adversarial Networks (GANs) and Variational Auto encoders (VAEs) have also been used.

Beyond translating individual sentences, story-to-sign production necessitates a deeper level of semantic comprehension and context preservation. Longer text sequences can now be contextually encoded thanks to NLP innovations like BERT (Devlin et al., 2018) and GPT-based models. While few recent works specifically target translation into sign language, several use these models into storytelling applications. Combining visual gesture production with semantic-aware natural language processing is necessary to close this gap.

The field of multimodal translation, which combines text and visual gesture representation, is still in its infancy. Translation pipeline projects like Text2Sign and Sign2Text are gaining traction. These systems create end-to-end pipelines for real-time sign generation by combining many modalities—text, gloss, pose, and video.

3. Methodology

3.1 System Architecture

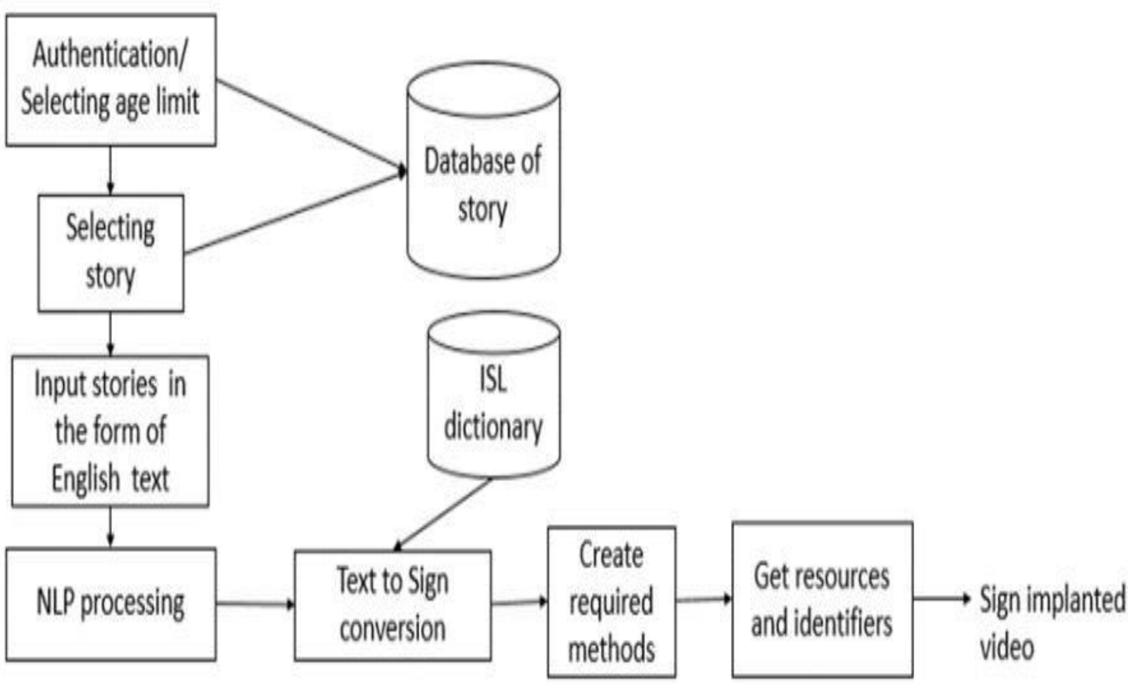

Fig 3.1.1 System Architecture

1. Authentication

User authentication and choosing a suitable age group are the first steps in the procedure. This guarantees that stories are selected according to the developmental stage of the child. It offers an extra degree of security and customization. Learning and engagement are improved by age-appropriate content.

2. Selecting Story

The user chooses a narrative from the system's library after authenticating. Based on the chosen age group, stories are suggested. This guarantees age-appropriate themes and relevance. It keeps the child's attention and understanding sustained.

3. Input Stories in the Form of English Text

The chosen narrative is either typed, uploaded, or retrieved as English text. The primary input for additional processing is this text. Translation accuracy is increased by input that is clear and grammatically correct. Language analysis of the narrative is now possible.

4. Database of Story

For convenience, every story is kept in a single database. Topic, age, and difficulty classifications are supported by the database. As new content is introduced, it permits expansion and changes. Because of this, narrative management is scalable and effective.

5. ISL Dictionary

English terms and their sign language glosses are mapped by the Indian Sign Language (ISL) dictionary. It guarantees that the signs of every word or phrase match. Standardized and meaningful translation depends on this. It serves as the language basis for the output of signs.

6. NLP Processing

The supplied text is structurally analyzed by Natural Language Processing (NLP). Tokenization, phrase segmentation, and syntax parsing are among the activities it involves. This stage aids in meaning extraction and gets the text ready for conversion to signs. Better gesture generation is ensured by accurate NLP.

7. Text to Sign Conversion

In this step, processed English text is converted into glosses in sign language. It matches words with their sign counterparts using the ISL dictionary. Context is thought to preserve the meaning of sentences. A sequence that is prepared for animation is the result.

8. Create Required Methods

Here, techniques are created to specify the animation of the signage. Scripting face emotions and gesture movements is part of this. Transitions and timing are designed to flow naturally. It guarantees that the signing seems expressive and natural.

9. Get Resources and Identifiers

Avatar identifiers and other necessary visual resources, such as gesture files, are gathered. The created sign glosses correspond to these. The data for animation is put together by the system. In this step, visual elements and spoken output are connected.

10. Sign Implanted Video

Lastly, a video featuring the avatar speaking sign language is produced. The signals correspond to the plot of the novel. Children with hearing impairments can easily watch and enjoy this entertaining film. It completes the narrative experience in a format that is inclusive and visual.

3.2 Implementation

The suggested storytelling system for kids with hearing loss is implemented as a modular pipeline that uses deep learning to process natural language and translate it into representations in sign language. Along with necessary frameworks like PyTorch or TensorFlow for model training and tools like spaCy, HuggingFace, OpenPose, Blender, or Unity for natural language processing and animation, it starts with setting up the development environment using Python. Formatting datasets with aligned text, gloss annotations, and animation data is part of the data pre-processing phase. Standard NLP techniques like tokenization and normalization are then applied. To preserve semantic integrity, gloss annotations are meticulously matched with matching avatar posture sequences.

The text-to-gloss model and the gloss-to-pose model are the two main parts of model development. While the second uses models like RNNs or GNNs to transform gloss sequences into posture data, the first uses neural machine translation approaches like sequence-to-sequence or Transformer models to translate preprocessed text into sign glosses. In order to ensure coordinated gestures and facial expressions for organic storytelling, the resultant pose data is converted into animations using 2D or 3D avatars. A backend (created with Flask or FastAPI) that manages model inference and animation control communicates with an intuitive user interface that is constructed using frameworks like React. To improve sign accuracy, animation quality, and user experience, the entire system is regularly updated and integrated with feedback. It is also extensively tested and deployed on cloud platforms for scalability.

4. Results

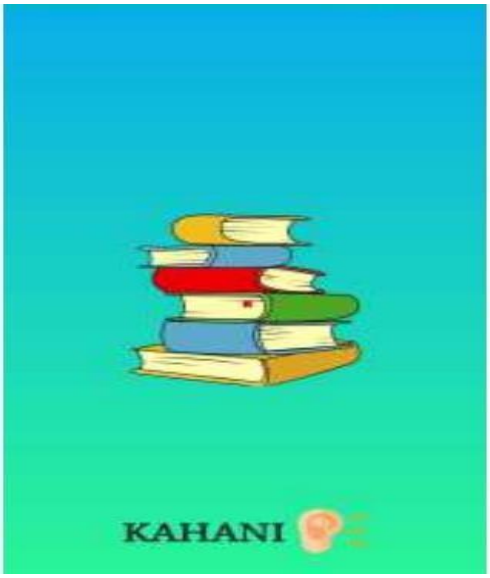

Figure 4.1 Splash screen

Figure 4.2 Age group selection page

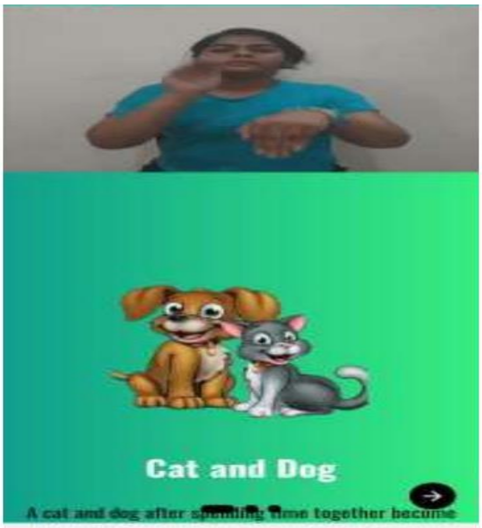

Figure 4.3 Story selection page

Figure 4.4 Read now page

Figure 4.5 Story page

Figure 4.6 Speed meter

5. Conclusion

The creation of a deep learning-based storytelling to sign language generation system meets a vital demand for inclusive communication by enabling the Deaf and Hard of Hearing (DHH) community to access narrative content. The system effectively converts spoken or written narratives into precise, animated sign language sequences by utilizing developments in computer vision, neural machine translation, and natural language processing.

Flexibility, scalability, and adaptation across several sign languages and content kinds are ensured by the modular method, which consists of input processing, gloss translation, sign creation, and avatar-based rendering. While Unity-based animation offers a dynamic and captivating user experience, deep learning models improve the accuracy and fluency of translation.

By making it possible for real-time, automated transmission of sign language content, this method promotes digital accessibility. Its accuracy and usability will be further improved with further modeling, linguistic validation, and user input. The installation lays the groundwork for further study and use of assistive communication, media, and education technology.

6. Future Enhancement

By utilizing cutting-edge technologies, future improvements to storytelling for kids with hearing loss seek to make the experience more inclusive, accessible, and entertaining. The usage of AI-powered sign language avatars that instantly translate stories into expressive sign language will be one of the major innovations, facilitating children's engagement with the story. Characters and scenes will come to life in interactive visual storybooks that combine AR and VR, creating a captivating setting. Furthermore, the smooth conversion of spoken narrations into sign language movies will be made possible by voice-to-sign translation systems, which will enhance accessibility during live storytelling sessions.

Gamified learning modules, including drag-and-drop exercises, interactive tests, and visual puzzles, will be added to the narrative process to enhance it even more and promote engagement and understanding. Multisensory experiences, such as tactile tools and haptic feedback, will improve the story's emotional impact. Additionally, parents and teachers will be able to modify storytelling sessions according to the particular requirements of each kid thanks to customizable options like text formatting, sign speed, and subtitles, which will make the experience effective and individualized.

References

[1] H. Rangoonwala, V. Kaushik, P. Mohith and D. samiappan,"TEXT TO SPEECHCONVERSION MODULE," proceedings of the 2017 International Journal of pure and Applied Mathematics, Volume 115 No.6,389.

[2] Palavi Bhole, Samiksha Kale, Ajaykumar Gujja, Prof. Deepti Vijay Chandran, "Storytelling App for Children with Hearing Impairment Using Natural Language Processing " Proceedings of the 2020 International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 0

[3] Hidayah, N., Astuti, E. M., & Triantoro, W. (2020). Story Hub: A Mobile Application for Hearing Impairment Using Natural Language Processing. In Proceedings of the 2020 3rd International Conference on Computer Science and Artificial Intelligence (pp. 201-205).

[4] Zaidan, S., Al-Samarraie, H., & Zaidan, B. (2021). Design and Development of Storytelling Mobile Application for Hearing Impaired Children using Natural Language Processing. IEEE Access, 9, 135158-135177.

[5] Almaraashi, M., Alsolami, F., & Altalhi, H. (2020). Sign Language Translation System Using Natural Language Processing for Hearing-Impaired People. In Proceedings of the 2020 6th International Conference on Information Management (pp. 7-11).

[6] Lin, C. H., & Tsai, M. H. (2019). Design and Implementation of a Sign Language Storytelling Application for Deaf Children. In 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC) (pp. 3808-3813).

[7] Choi, J. Y., & Lee, J. (2019). Sign Language Storytelling Application for Deaf Children Using Symbolic Representation. In 2019 16th International Conference on Ubiquitous Robots (UR) (pp. 472-477).

[8] Cho, H., Cho, H., & Kwon, O. (2018). Development of a storytelling application for hearing-impaired children using tokenization.

[9] Adaloglou, N., Nikoloulopoulos, A., & Papavassiliou, V. (2021). Sign Language Transformers: Joint End-to-end Sign Language Recognition and Translation. *ICASSP 2021 - IEEE International Conference on Acoustics, Speech and Signal Processing*, 7503–7507.

[10] Albanie, S., Varol, G., Momeni, L., Afouras, T., & Zisserman, A. (2020). BSL-1K: Scaling up British Sign Language Recognition using Weakly Labelled YouTube Videos. *European Conference on Computer Vision (ECCV)*, 35–53.

[11] Stoll, S., Camgoz, N. C., Hadfield, S., & Bowden, R. (2018). Sign Language Production Using Neural Machine Translation and Generative Adversarial Networks. *British Machine Vision Conference (BMVC)*.

[12] Starner, T., Weaver, J., & Pentland, A. (1998). Real-Time American Sign Language Recognition Using Desk and Wearable Computer-Based Video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12), 1371–1375.

[13] Pfister, T., Charles, J., & Zisserman, A. (2013). Large-Scale Learning of Sign Language by Watching TV (Using Subtitles). *British Machine Vision Conference (BMVC)*.

[14] Bragg, D., Koller, O., Bellard, M., Berke, L., & Kushalnagar, R. (2019). Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. *The 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*.