International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Reverse Sign Language Recognition

Sanika Gaikwad[1], Mukesh Mahindrakar[2], Anurag Shevale[3], Neha Thombare[4], Dr. Shraddha Khonde[5]

[1] *Department of Computer Engineering, MES's Wadia College of Engineering, Pune*
[2] *Department of Computer Engineering, MES's Wadia College of Engineering, Pune*
[3] *Department of Computer Engineering, MES's Wadia College of Engineering, Pune*
[4] *Department of Computer Engineering, MES's Wadia College of Engineering, Pune*
[5] *Department of Computer Engineering, MES's Wadia College of Engineering, Pune*

---***---

Abstract - Sign language is an important mode of communication for deaf, mute, and disabled people. Sign language is a language in which people communicate with the help of gestures in situations where they cannot speak or hear. Gestures are an effective method of human interaction and are often used by deaf people to communicate. This article describes various methods and techniques for recognizing text signatures in images. This review compares different methods and algorithms with the help of a pie chart.

Key Words: *communication, sign recognition, hand gestures, sign language*

1. INTRODUCTION

Communication plays an important role in our daily lives. The ability to communicate with each other is essential. Deaf and mute people make use of sign language to express their thoughts and ideas to their communities through gestures and body language. Language is widely used around the world to facilitate communication between people of different abilities is on everyone's radar.

Sign language is the only way of communication for the deaf and mute community. Sign language is a well-structured language where every gesture has some meaning assigned to it.

Text and image recognition are tasks that aim to generate accurate images from text descriptions. Reverse sign recognition involves recognizing the written text to provide the corresponding gestures.

The aim of reverse sign language recognition is to recognize words or characters and convert them into sign language.

Hence, any advancements in the existing systems will only enhance the technology.

2. LITERATURE SURVEY

Mourad Bahani, Aziza El Ouaazizi, Khalil Maalmi [1]. GAN also known as Generative Adversial Network it is generally used for text to photo processing and creation tasks. The paper presents a novel approach to generating realistic images from Arabic text using DF-GAN and AraBERT. They translate English descriptions to Arabic, use AraBERT embeddings, and train the system on CUB and Oxford-102 datasets. Evaluation shows promising results with FID scores of 55.96 and 59.45, and SI scores of 3.51 and 3.06 for CUB and Oxford-102 datasets respectively. The framework addresses the complexity of generating coherent Arabic text and achieves significant success in this area.

M. Alsulaiman, M. Faisal, M. Mekhtiche, M. Bencherif, T. Alrayes, G. Muhammad, H. Mathkour,W. Abdul, Y. Alohali, M. Alqahtani, H. Al-Habib,H. Alhalafi, M. Algabri, M. Al-hammadi, H.Altaheri, and T. Alfakih [2] The paper underscores the importance of facilitating communication with deaf individuals, introducing a comprehensive survey of Arabic sign language datasets. It details the creation of the largest Saudi Sign Language (SSL) database, containing 145,035 samples across various fields. A Convolutional Graph Neural Network (CGCN) architecture is proposed for sign language recognition. This research aims to develop a bidirectional communication system for Saudi deaf and mute individuals using Avatar technology. The SSL database includes both static and dynamic signs, providing a valuable resource for accurate sign language recognition models.

K. Kudrinko, E. Flavin, X. Zhu, and Q. Li, [3] The review surveys 72 studies from 1991 to 2019, emphasizing electronic tools for sign language recognition. Sign language utilizes signals, facial expressions, and body movements, with over 300 variations worldwide. Dominant hand usage and finger-written letters are key features. Facial expressions complement gestures, conveying nuances in meaning. Sign language is less reliant on grammar compared to spoken or written language. Challenges in written communication for the deaf underscore the need for wearable sensor-based systems.

F. Shah, M. S. Shah, W. Akram, A. Manzoor, R. O.Mahmoud, and D. S. Abdelminaam's [4] The IEEE paper focuses on Pakistan Sign Language (PSL) and user-friendly learning tools. It utilizes image-based features like HOG, EOH, LBP, and SURF, employing SVM classifiers with MKL. The proposed architecture achieves 91.93% accuracy in PSL recognition, with graphs illustrating precision, recall, and F-score. SURF feature efficiency is noted at 15%, highlighting areas for improvement like using MSER features. Despite limitations, the study recommends further exploration of feature enhancements for wider application.

S. M. Kamal, Y. Chen, S. Li, X. Shi, and J. Zheng [5] The overview presents Chinese language research, introducing the Chinese Translation Recognition System (CSLR) and related databases. Researchers utilize Kinect and data gloves to collect clinical data, including discrete and continuous samples. Deep learning models like CNNs and LSTMs show promise in sign language recognition, achieving 86% and 63% accuracy in data classification. The system outperforms HMM in speed and accuracy, indicating advancements in sign language processing.

H. D. Yang, S. Sclaroff, and S. W. Lee's [6] The paper presents a language recognition method using default patterns and a CRF model, achieving 87.0% and 93.5% recognition rates. It reviews studies employing DTW and HMM methods with rates of 90.8% and 99% respectively, suggesting potential for accuracy improvement in language acquisition tasks

C. M. Sharma, K. Tomar, R. K. Mishra, and V. M. Chariar [7] The PDF research paper focuses on Hindi language learning using deep learning models, emphasizing its importance for translation for the deaf. It manually generates a dataset of over 150,000 images of 26 letters, trained using VGG16, achieving 94.52% efficiency. However, automatic sign language recognition (ASLR) faces challenges such as tracking hands and recognizing patterns, with improved feature extraction reaching 98.125% accuracy. The paper also reviews other research, including MSVM achieving 63% accuracy with a small dataset, and image processing methods achieving 98.125% accuracy with 320 images..

S. C. J and L. A [8] Signet is a deep learning-based Indian Sign Language Recognition System (ISLRS) achieving a performance rate of 98.64%, surpassing existing methods. It employs a CNN architecture for static alphabet recognition from binary hand region silhouettes. Background elimination is crucial, with glove-based and vision-based approaches used for hand gesture recognition. Vision-based methods, relying on computer vision techniques, overcome challenges without requiring specialized gloves or sensors. Challenges include variations in distance and viewing angles, as well as hidden hand parts obstructing recognition. Vision-based approaches can be 2-D or 3-D, with 3-D techniques leveraging depth information for improved accuracy.

B. M. V and S. T. D's [9] The paper presents a computer vision method for learning Indian languages, aiming to aid communication for the deaf. It utilizes deep learning, specifically Convolutional Neural Networks (CNNs), to increase accuracy in character classification. The system focuses on hand region input, achieving 99.56% accuracy for static images and 97.26% accuracy in low light. Future improvements will involve spatial analysis and enhanced low-light performance. Overall, the proposed approach aims to provide an accurate Indian Sign Language recognition system to improve communication for the deaf.

G. Rao, K. Syamala, P. V. V. Kishore, and A. S. C. S. Sastry [10] The document focuses on ISL gesture recognition using DeepCNN, achieving a recognition rate of 92.88% and outperforming other classifier models. A dataset of 200 sign language gestures performed by five subjects in five viewing angles was created. The advantages of using CNN models include improved performance, higher recognition rates, robustness, efficiency, and suitability for mobile platforms in sign language recognition tasks.

N.Pugeault & R.Bowden's [11] The paper focuses on developing an interactive hand shape recognition system for ASL finger-spelling, utilizing Microsoft Kinect for appearance and depth images. It employs OpenNI + NITE for hand detection, random forests for classification, and addresses challenges in hand shape recognition. Combining appearance and depth images enhances reliability and accuracy, with the system showing robustness to environmental conditions. Integrated into a user interface, it allows users to select between ambiguous detections, facilitating effective communication with an English vocabulary and additional features like word suggestions and validation.

Overall, the system improves usability and practicality in finger-spelling recognition.

Xu T, Zhang P, Huang Q, Zhang H, Gan Z, Huang X, et al.'s [12] The paper introduces the Attentional Generative Adversarial Network (AttnGAN) for fine-grained text-to-image generation. AttnGAN consists of an attentional generative network and a deep attentional multimodal similarity model. It significantly outperforms previous models in terms of image quality and accuracy, employing an attention mechanism to focus on relevant words in the description for generating fine-grained details in different sub-regions of the image. The generator utilizes word-context vectors and multimodal context vectors to refine image features at each stage, resulting in high-quality images aligned with the text description.

Yong Xuan Tan , Chin Poo Lee [13] The paper introduces SS-TiGAN, a self-supervised image-to-image generative adversarial network (GAN) for text-to-image synthesis. It utilizes a two-layer GAN architecture to combine images from 64x64 to 128x128 pixels, overcoming challenges in low data scenarios. The method facilitates training and reduces model complexity, reaching higher resolutions with a two-stage architecture. While it has limitations due to resource constraints, its simple design and use of sentence-level encoders make it promising for document synthesis.

B. Natarajan , E. Rajalakshmi , R. Elakkiya [14] The paper presents a system for Sign Language Recognition, Translation, and Video Generation, addressing significant challenges in development. It utilizes the Media Pipe library and a hybrid CNN + Bi-LSTM model for pose detail extraction and text generation, achieving high accuracy of 95% and above in classification. Additionally, for video generation, a model combining Neural Machine Translation (NMT), Media Pipe, and Dynamic Generative Adversarial Network (GAN) is employed, offering a comprehensive solution for sign language processing.

Hyunhee Lee, Gyeongmin Kim , Yuna Hur , and Heuiseok Lim [15] The paper introduces a system for text-to-image synthesis, addressing the challenge of conveying detailed visual information through text alone. It contrasts GANs with Neural Networks for this task, opting for the latter due to disadvantages of GANs. The CoDraw dataset is utilized, consisting of dialogue between a teller and a drawer who collaboratively create images based on textual descriptions. The system facilitates communication between the teller and drawer to generate images corresponding to the provided descriptions, enhancing understanding and visualization of textual content.

Uche Osahor , and Nasser M. Nasrabadi [16] The paper presents a method for synthesizing human face images from text descriptions and sketches using Generative Adversarial Networks (GAN) and an attention model. The model incorporates a position-wise-feed-forward network and layer normalization to extract attributes from text and focus on facial features. It introduces a text-guided model to translate sketches to images, facilitating image generation from sketches. The method produces high-quality and accurate images even from slightly degraded sketches. Ablation studies are conducted to assess the quality of sketch-to-image synthesis.

Nikhil Kasukurthi ,Brij Rokad, Shiv Bidani, Aju Dennisan [17] The paper introduces a model for recognizing American Sign Language Alphabet using RGB images, requiring only 2D input. It is trained on the SqueezeNet architecture for mobile device compatibility, achieving an accuracy of 83.29%. The architecture aims to reduce parameters and network size. Trained on the Surrey Finger Dataset, it achieves a maximum training accuracy of 87.47%.

Shagun Katoch, Varsha Singh, Uma Shanker Tiwary [18] The development of Indian Sign Language (ISL) recognition systems faces challenges like diverse gestures, regional variations, and historical lack of standardization. Machine learning and deep learning technologies have enabled efficient and accurate algorithms, including end-to-end models, enhancing communication for the hearing impaired. Automatic sign language recognition systems play a crucial role in bridging the communication gap between the hearing impaired and the rest of the population.

Satwik Ram Kodandaram, N. Pavan Kumar, Sunil G [19] The paper delves into sign language recognition using Deep Learning techniques, crucial for effective communication among individuals with hearing and speech impairments. It explores static and dynamic hand gestures, employing CNN algorithms like LeNet-5 and MobileNetV2 for efficient recognition from images or video frames. CNN's proficiency in recognizing complex gestures enhances communication accessibility for the deaf or hard of hearing.

Yuxuan Liu, Xijun Jiang , Xingge Yu, Huaidong Ye, Chao Ma, Wanyi Wang, Youfan Hu [20] The paper presents a wearable Sign Language Recognition system, utilizing Convolutional Neural Networks (CNN) integrated with stretchable sensors and inertial measurement units to capture hand postures and movements. This system aims to facilitate communication between the deaf/mute community and society, enabling effortless interaction and mutual understanding. Wearable technology enhances usability, allowing users to access sign recognition anywhere, independent of location.

Kaustubh Jadhav, Abhishek Jaiswal, Abbas Munshi, Mayuresh Yerendekar [21] The paper discusses challenges faced by the deaf/mute community and potential solutions, emphasizing the role of technologies like sign language recognition systems. It highlights the importance of recognizing sign languages' nuances and limitations in datasets, while advocating for advancements in machine learning and computer vision to aid communication and inclusion. CNNs, specialized for image and spatial data tasks, are briefly introduced, showcasing their effectiveness in various applications. Overall, addressing these challenges through technology and education can create a more inclusive world for everyone.

Lianyu Hu, Liqing Gao, Zekang Liu, Wei Feng [22] The paper discusses the significance of sign language as a communication tool for disabled individuals and the challenges hearing people face in learning it. It introduces vision-based continuous sign language recognition (CSLR) as a means to bridge the communication gap between the two groups by automatically translating sign videos into sentences. The PHOENIX14 dataset, recorded from German TV weather forecasts, is utilized for training and testing. It comprises 6841 sentences with a vocabulary of 1232 words, split into training, development, and testing samples for multi-signer setups. PHOENIX14-T serves as an extension for both CSLR and Sign Language Translation tasks.

Yao Du, Pan Xie, Mingye Wang, Xiaohui Hu, Zheng Zhao, Jiaqi Liu [23] The paper discusses the limitations of current deep learning-based approaches in single-level sign language recognition (SLR) and highlights the importance of capturing gestures and facial expressions for accurate interpretation. While convolutional neural networks (CNNs) are effective in extracting features from images, they lack the ability to create visual context awareness or understand video sequences. CNNs treat all parts of an image equally due to weight sharing, which is not ideal for extracting detailed information such as hands and faces. Thus, there is a need for advancements in feature extraction techniques to improve SLR systems.

Kartik Shenoy, Tejas Dastane, Varun Rao, Devendra Vyavaharkar [24] This paper presents a gesture recognition system for Indian Sign Language (ISL) to aid communication between hearing and speech impaired individuals and others. It achieves high accuracy without external devices, using smartphone cameras and facial recognition. The system recognizes 33 ISL characters and tokens, reaching a static character recognition rate of 99.7% and a correct character recognition rate of 97.23%.

Syed Yasser Arafat and Muhammad Javed Iqbal [25] This paper proposes a method for Urdu text detection in natural scene images using deep learning, specifically the Faster R-CNN algorithm combined with CNNs like Squeezenet, Googlenet, Resnet18, and Resnet50. The model is trained on extracted features from training images, achieving an accuracy of 79% for 4k images and 99.5% for 51k images.

Y. Watanabe, R. Togo [26] This paper introduces text-guided image manipulation, a concept where natural language descriptions are used to control image generation for user-friendly manipulation. Methods like CMPC-Refseg and text-guided feature exchange modules are proposed to semantically alter image appearance to meet user requirements, overcoming limitations of traditional image manipulation techniques.

Tao M, Tang H, Wu S, Sebe N, Jing X, Wu F, et al’s [27] DF-GAN is an innovative model for text-to-image synthesis that addresses three main challenges faced by existing GANs. Firstly, it introduces a single-stage architecture, eliminating the need for multiple renderers and enabling high-resolution image generation without entanglement. Secondly, DF-GAN incorporates an object-aware discriminator with match-aware richness to enhance text-image semantic compatibility without compromising network monitoring. Thirdly, it introduces a Deep Text-Image Fusion Block (DFBlock) for efficient fusion of text and visual features, improving the overall synthesis process. These advancements result in better performance compared to existing methods, as demonstrated on the CUB and COCO datasets. DF-GAN's single-level mapping for background images, object-aware discrimination, and Deep Text-Image Fusion Block contribute to its effectiveness in generating realistic images from text descriptions.

Zhu MP, Chen W, Yang Y. [28] DMGAN (Dynamic Memory Generative Adversarial Network) is designed to overcome key challenges in text-to-image synthesis, specifically addressing dependence on initial image quality and maintaining consistency in text representation during image processing. It introduces a dynamic memory module to correct image blur and a memory writing system to select important text based on the initial image. By adaptively fusing information from memories and image features through a response gate, DMGAN produces high-quality images from text descriptions. Evaluation on datasets like CUB and COCO demonstrates DMGAN's superior performance in terms of image quality and annotation accuracy compared to existing methods, making it a state-of-the-art approach for text-to-image synthesis.

Zhang H, Xu T, Li H, Zhang S, Wang X, Huang X's [29] StackGAN is a text-to-image synthesis model that generates high-quality images based on text descriptions. It consists of two stages: Stage-I GAN sketches the basic shape and colors of the object, producing low-resolution images, while Stage-II GAN refines these images, adding realistic details. By decomposing the text-to-image process into two stages, StackGAN improves the quality and diversity of generated images. The model also introduces a Conditioning Augmentation technique to enhance training dynamics and stabilize the output. Overall, StackGAN significantly advances the generation of realistic images from text annotations.

A. K. Tushar, A. A. Ashiquzzaman, and M. R. Islam's [30] This research paper focuses on improving digital character recognition, particularly for American Sign Language (ASL), using deep convolutional neural networks (DCNN). The model integrates batch normalization and dropout techniques to enhance training speed and reduce overfitting. The proposed model achieves 98.50% accuracy on ASL codes, outperforming other methods. The ASL-Numerical dataset, comprising 500 images per category across 10 digital categories, is used for evaluation. The combination of batch normalization and dropout in the model enhances digital character recognition performance.

3. CONS OF DEAF SIGN LANGUAGE RECOGNITION

Technical limitations: Current sign language recognition technology may not be highly accurate, leading to misunderstandings and frustration.

Privacy Concerns: Privacy concerns can arise when sign language conversations are recorded or transcribed without consent.

Gaps in accessibility: Not all deaf individuals may have access to sign language recognition technology due to cost, availability, or technical barriers.

Dependence on technology: There is a risk of over-reliance on technology, which can potentially reduce motivation to learn and use sign language.

Limited vocabulary: Recognition systems can struggle with complex or unusual characters, which limits their practicality. Cultural Implications: Reliance on technology could potentially erode the cultural significance of sign languages and deaf communities.

Development and maintenance costs: Implementation and maintenance of sign language recognition systems can be expensive for organizations and governments.

Training Requirements: Deaf individuals and those who interact with them may need training to effectively use sign language recognition technology.

Accessibility for low-income communities: Technology may not be accessible to low-income deaf individuals or people in developing countries. Integration challenges: Integrating sign language recognition into different applications and devices can present technical and compatibility issues.

4. DISCUSSION

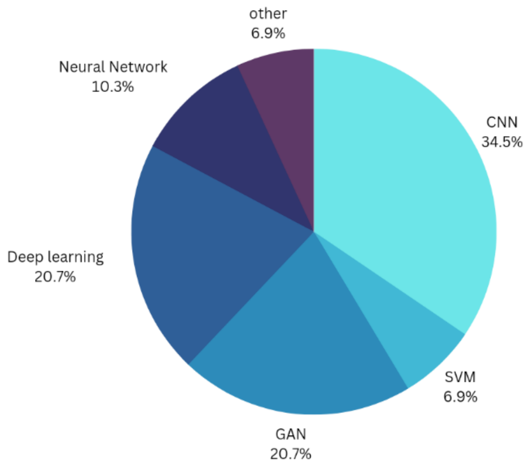


Figure 1: Use of Algorithm in Paper

In the above diagram, it is evident that convolutional neural networks (CNN) have emerged as a dominant force, accounting for a substantial 34.5% of the algorithms discussed. This widespread adoption highlights their effectiveness in various applications, particularly in the field of computer vision and image analysis. Additionally, deep learning, encompassing a diverse range of neural network architectures, contributes significantly with 20.7%, reflecting its versatile utility. Generative Adversial Network is a well-known text to image generation algorithm that holds a prominent place at 20.7%, illustrating their significant role in generating synthetic data and fostering creative applications. Furthermore, traditional methods like Support Vector Machines (SVM) remain relevant at 6.9%, offering a complementary approach. These statistics reveal the important role of algorithms in many areas of today's research and applications and show that their impact on education is gradually increasing.

5. PROJECT IMPLEMENTATION

Introduction

The implementation of Reverse Sign Language Recogniton involves 4 modules.

- **Text Embedding:**

In this module the input text description is converted into vector of 300 elements that can be understood by the machine. The text is converted to vector with the help of Word2Vec model which is a pre-trained model used for text embedding. These vector elements are then stored in a pickle file.

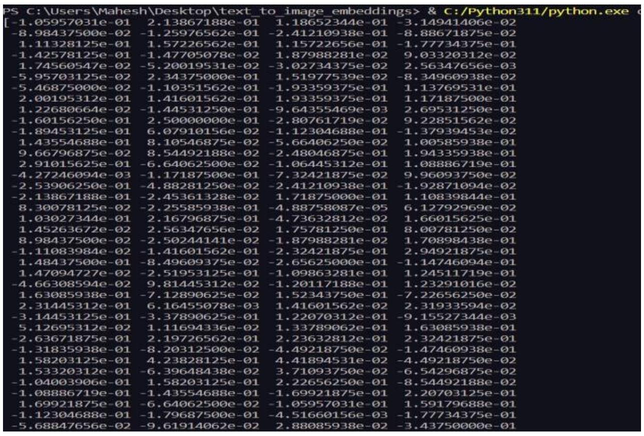


Figure 1.1 Vector elements for letter A

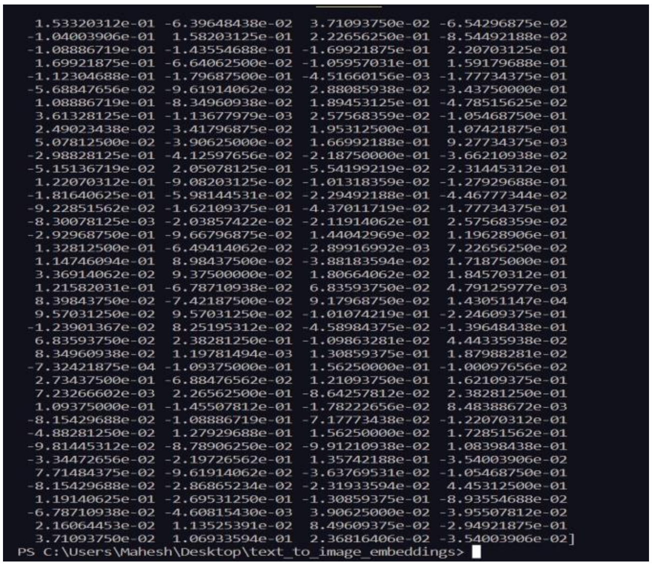


Figure 1.2 Vector elements for letter A

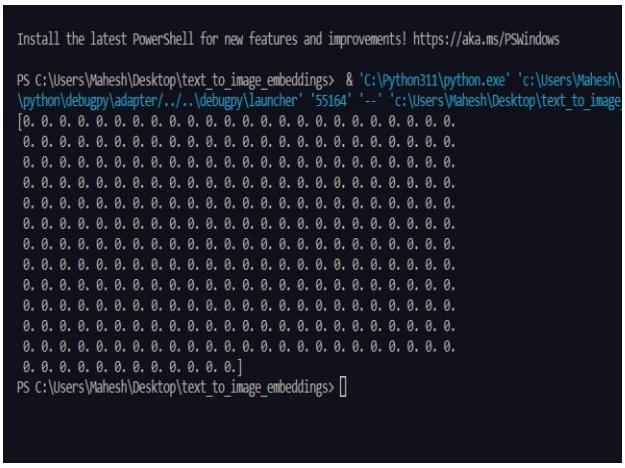


Figure 1.3 0 is displayed when the input text is not recognised

- **Stage I:**

In Stage-I, the process begins with taking the text embedding generated in the 1st module as input which is the pickle file. These encoded text embeddings are then combined with random noise vectors to form conditional inputs for the generator network. This generator network is responsible for producing low-resolution images of sign language gestures based on the input textual descriptions. Through adversarial training, where the generated images are compared against real images by a discriminator network, both networks iteratively improve. The generator aims to create images that are indistinguishable from real sign language gestures, while the discriminator learns to accurately distinguish between real and synthesized images.

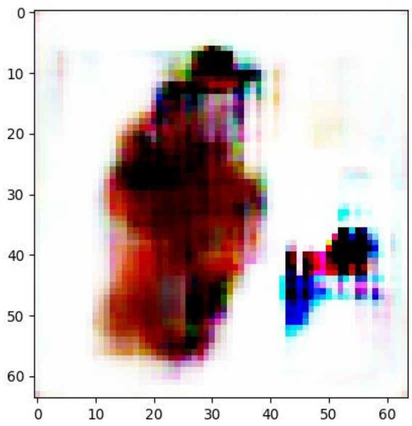


Figure 1.4 Low Resolution Image

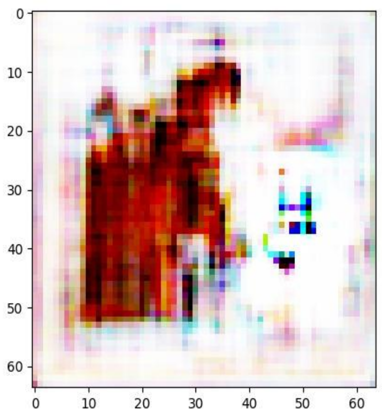


Figure 1.5 Low Resolution Image

- **Stage II:**

In Stage II, the focus shifts to refining the low-resolution images generated in Stage-1 into higher-resolution and more detailed representations. The initial low-resolution images serve as the starting point, undergoing an upsampling process to increase their spatial resolution. Alongside, the text embeddings derived from the input textual descriptions are incorporated to create conditional inputs for the refinement network. This network, known as the Refiner, then takes these inputs and generates high-resolution images with enhanced clarity, sharpness, and realism. Through adversarial training, the Refiner is trained to produce images that are perceptually indistinguishable from real sign language gestures, with a discriminator network providing feedback on the realism of the synthesized images. A loss function guides the optimization process, ensuring that the refinement network learns to generate high-quality images that align closely with the input textual descriptions.

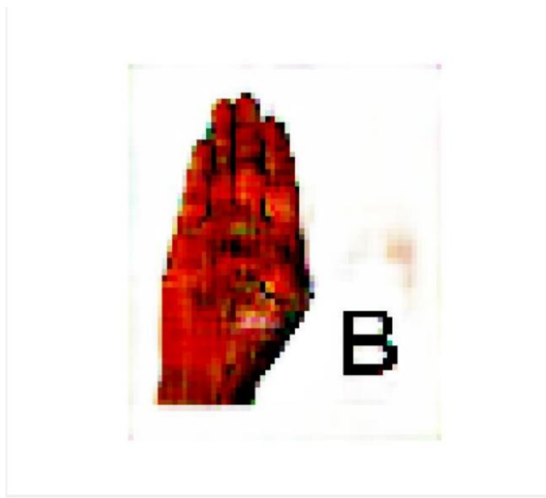


Figure 8.1.6 High Resolution Image for letter B

Figure 1.7 High Resolution Image for letter Z

- **GUI (GRAPHICAL USER INTERFACE):**

A Graphical User Interface (GUI) for reverse sign language recognition provides an intuitive platform for users to interact with the system. The GUI features an input section where users can input textual descriptions through a text input box and initiate the image generation process with a submit button. Upon submission, the output section displays the generated image representations of the sign language gestures corresponding to the input text, offering users a visual interpretation of the provided descriptions.

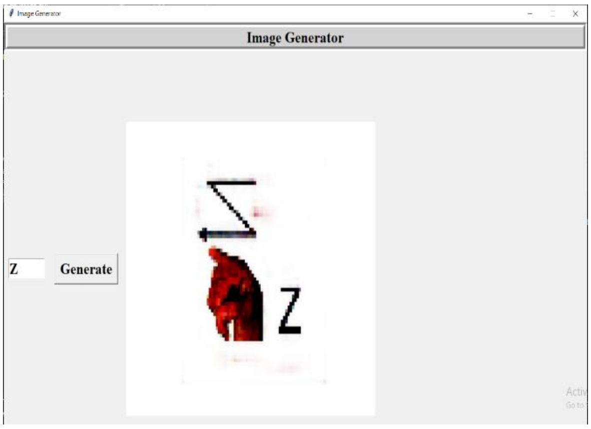


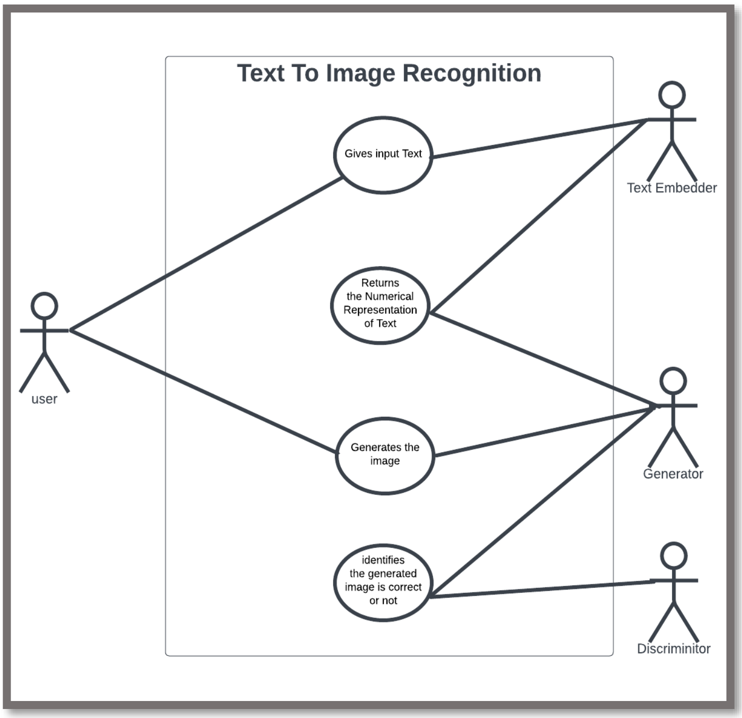
Figure 8.1.8 GUI

3. CONCLUSION

The research papers presented demonstrate significant advances and challenges in language learning, with emphasis on the important role played by deep learning models, especially communication neural networks (CNN). These studies include languages such as Arabic, Pakistani, Chinese, Hindi and Indian Sign Language, and emphasize the necessity of fast communication for the deaf. The point is that CNN has done a good job of identifying hand gestures, exceeding the capabilities of traditional products and showing good results, being powerful and accessible for the mobile platform. Scientists achieved the complexity of descriptions, such as different hands and heads, different directions and emotions, by carefully building materials and layers through the deep learning process.

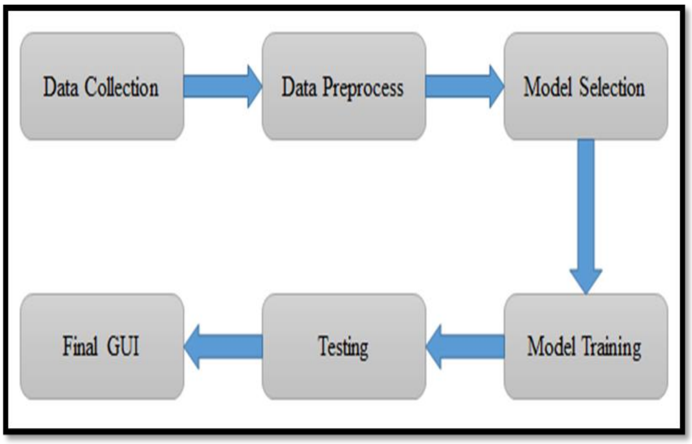
DIAGRAMS

1.Use Case Diagram



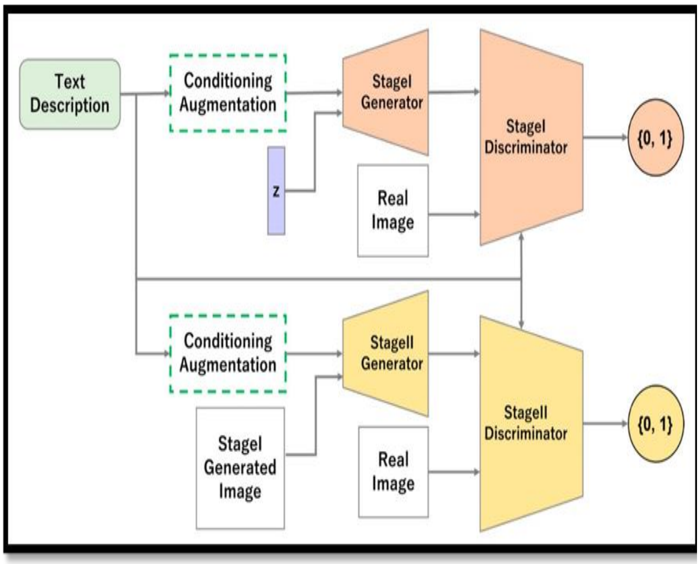
The user gives its desired input in the text inserting module. Further the text is taken to the text embedding and it converts into a vector of fixed size then the generator will take the input of the vector of a fixed length and conditional augment to generate an image and the image is sent to discriminator for verification of the image generated by generator is correct or not is acknowledged to the generator and the generator will display the output to the user.

2. Process Flow diagram



In the process flow diagram, collection of data from various sources after that data will go through the preprocessing stage in which the data will be cleaned transformed and organized. Then the next step is selecting a suitable model that is satisfying all our needs. Then we will train the model in such a way that it will understand relationship and pattern. Then we will test the model to perform its accuracy. And deploy with a suitable GUI.

3.Implementation of StackGan



1.Data collection: Collect data about images, such as images and their corresponding descriptions.

2.Preprocessing: Prepare the data by resizing the image, tagging the description, and creating training pairs. Phase 1: Text Embedding and Image Generation (Phase 1 GAN word embedding).
Image generation: Teach GAN to generate low-level images from descriptions.

Stage 2: Conditional Image Enhancement (Second Stage GAN):

Conditional GAN: Use images to create negative images and text-like images to enhance and create high-resolution images.
Refine: Add a separate GAN to improve the quality and detail of the negative image created in Stage 1.
Stacked GAN: The output of phase-II GAN and phase-I GAN is used to enhance low images to high image Images as per description.

Training and optimization: Two-level StackGAN architecture training, fine-tuning parameters and optimizing the network for better image synthesis.

Assessment and evaluation: Evaluate the quality of the created image using indicators such as threshold scores, Frechet threshold distance, or by manual evaluation.

Use and improve: Use the training model and continue to improve it based on feedback and training to create better images.

REFERENCES

[1] Mourad Bahani, Aziza El Ouaazizi, Khalil Maalmi "Ara-BERT and DF-GAN fusion for Arabic text-to-image generation", 2022.
[2] M. Alsulaiman, M. Faisal, M. Mekhtiche, M. Bencherif, T. Alrayes, G. Muhammad, H. Mathkour, W. Abdul, Y. Alohali, M. Alqahtani, H. Al-Habib, H. Alhalafi, M. Algabri, M. Al-hammadi, H. Altaheri, and T. Alfakih, "Facilitating the communication with deaf people: Building a largest Saudi sign language dataset," 2023.
[3] K. Kudrinko, E. Flavin, X. Zhu, and Q. Li, "Wearable Sensor-Based Sign Language Recognition: A Comprehensive Review," 2020.
[4] F. Shah, M. S. Shah, W. Akram, A. Manzoor, R. O. Mahmoud, and D. S. Abdelminaam, "Sign Language Recognition Using Multiple Kernel Learning: A Case Study Of Pakistan Sign Language," 2021.
[5] S. M. Kamal, Y. Chen, S. Li, X. Shi, and J. Zheng, "Technical Approaches To Chinese Sign Language Processing," 2020.
[6] H. D. Yang, S. Sclaroff, and S. W. Lee, "Sign Language Spotting With A Threshold Model Based On Conditional," 2009.
[7] C. M. Sharma, K. Tomar, R. K. Mishra, and V. M. Chariar, "Indian Sign Language Recognition Using Fine-Tuned Deep Transfer Learning Mode," 2021.
[8] S. C. J and L. A, "Signet: A Deep Learning Based Indian Sign Language Recognition System," 2019.
[9] B. M. V and S. T. D, "Indian Sign Language Numeral Recognition Using Region Of Interest Convolutional Neural Network," 2018.
[10] G. Rao, K. Syamala, P. V. V. Kishore, and A. S. C. S. Sastry, "Deep Convolutional Neural Networks For Sign Language Recognition," 2020.
[11] N.Pugeault & R.Bowden "Spelling It Out: Real–Time ASL Fingerspelling Recognition" 2019
[12] Xu T, Zhang P, Huang Q, Zhang H, Gan Z, Huang X, et al. “Attngan: Fine- grained text to image generation with attentional generative adversarial networks”,2018
[13] Yong Xuan Tan , Chin Poo Lee , Mai Neo , Kian Ming Lim , Jit Yan Lim,“Text-to-image synthesis with self-supervised bi-stage generative adversarial network” 2023
[14] B. Natarajan , E. Rajalakshmi , R. Elakkiya , Ketan Kotecha , Ajith Abraham , (Senior Member, IEEE), Lubna Abdelkareim Gabralla and V. Subramaniyaswamy, “Development of an End-to-End Deep Learning Framework for Sign Language Recognition, Translation, and Video Generation” 2022
[15] Hyunhee Lee , Gyeongmin Kim , Yuna Hur , And Heuiseok Lim, “Visual Thinking of Neural Networks: Interactive Text to Image Synthesis” 2021.
[16] Uche Osahor , (Student Member, Ieee), And Nasser M. Nasrabadi, “Text-Guided Sketch-to-Photo Image Synthesis” 2023.
[17] Nikhil Kasukurthi ,Brij Rokad, Shiv Bidani, Aju Dennisan, “American Sign Language Alphabet Recognition using Deep Learning” 2019.
[18] Shagun Katoch, Varsha Singh, Uma Shankar Tiwary,” Indian Sign Language recognition system using SURF with SVM and CNN” 2022.
[19] Satwik Ram Kodandaram, N. Pavan Kumar, Sunil G ,” Sign Language Recognition using Deep learning.” ,2021
[20] Yuxuan Liu, Xijun Jiang, Xingge Yu, Huaidong Ye, Chao Ma, Wanyi Wang, Youfan Hu, “A wearable system for sign language recognition enabled by a convolutional neural network”, 2023
[21] Kaustubh Jadhav, Abhishek Jaiswal, Abbas Munshi, Mayuresh Yerendekar, “Sign language recognition using neural network”, 2020
[22] Lianyu Hu, Liqing Gao, Zekang Liu, Wei Feng.“Scalable frame resolution for efficient continuous sign language recognition”, 2023
[23] Yao Du, Pan Xie, Mingye Wang, Xiaohui Hu, Zheng Zhao, Jiaqi Liu,” Full transformer network with masking future for word-level sign language recognition”, 2022
[24] Kartik Shenoy, Tejas Dastane, Varun Rao, Devendra Vyavaharkar,” Real-time Indian Sign Language (ISL) Recognition” 2021
[25] Syed Yasser Arafat and Muhammad Javed Iqbal,“Urdu-Text Detection and Recognition in Natural Scene Images Using Deep Learning” 2020.
[26] Y. Watanabe, R. Togo, K. Maeda, T. Ogawa and M. Haseyama, "Text-Guided Image Manipulation via Generative Adversarial Network With Referring Image Segmentation-Based Guidance," 2023.
[27] Tao M, Tang H, Wu S, Sebe N, Jing X, Wu F, et al. “Deep fusion generative adversarial networks for text-to-image synthesis”, 2020
[28] Zhu MP, Chen W, Yang Y. Dm-gan: “Dynamic memory generative adversarial networks for text-to-image synthesis.” 2019
[29] Zhang H, Xu T, Li H, Zhang S, Wang X, Huang X, “Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks” ,2017.
[30] A. K. Tushar, A. A. Ashiquzzaman, and M. R. Islam, "Faster Convergence and Reduction of Overfitting in Numerical Hand Sign Recognition using DCNN," Computer Science and Engineering Department, University of Asia Pacific, Dhaka, Bangladesh, 2017

LITERATURE SURVEY TABLE :

	PAPER NAME	YEAR	AUTHORS	REMARK
1.	Text-to-image synthesis with self-supervised bi-stage generative adversarial network.	2023	Yong Xuan Tan , Chin Poo Lee , Mai Neo , Kian Ming Lim , Jit Yan Lim	This paper makes use of SS-TiGAN ,which is a network architecture designed to generate high quality images from text descriptions
2.	Visual Thinking of Neural Networks: Interactive Text to Image Synthesis	2021	Hyunhee Lee , Gyeongmin Kim , Yuna Hur , And Heuiseok Lim	The authors have made use of Neural Networks to overcome some drawbacks of GANs.
3.	Text-Guided Sketch-to-Photo Image Synthesis	2022	Uche Osahor , (Student Member, Ieee), And Nasser M. Nasrabadi	The authors show that their model can generate much accurate images of human faces from text description and sketches even when sketches are slightly degraded.
4.	Indian Sign Language recognition system using SURF with SVM and CNN	2022	Shagun Katoch, Varsha Singh, Uma Shanker Tiwary	A novel approach to classify and recognize Indian sign language signs (A-Z) and (0–9.) using the SVM and CNN is presented in the paper.
5.	Sign Language Recognition using Deep learning.	2021	Satwik Ram Kodandaram, N. Pavan Kumar, Sunil G	It focuses on static sign language hand gestures, using Deep Neural networks
6.	AraBERT and DF-GAN fusion for Arabic textto-image generation	2022	Mourad Bahani, Aziza El Ouaazizi, Khalil Maalmi	The paper introduces a novel text-to-image generation framework, combining DF-GAN and AraBERT, tailored for Arabic text descriptions and demonstrates its success in producing high-resolution, realistic images, showcasing its uniqueness in the field.
7.	Signet: A Deep Learning Based Indian Sign Language Recognition System	2019	Sruthi C. J And Lijiya A Member, IEEE	A Deep Learning-Based System For Indian Sign Language Recognition Achieving An Impressive 98.64% Accuracy By Addressing Challenges Like Hand Region Extraction And Background Elimination Using A CNN Architecture With Six Hidden Layers.
8.	Recognition in Natural Scene Images Using Deep Learning.	2020	Syed Yasser Arafat, Muhammad Javed Iqbal	This system was able to achieve accuracy of 99.06%
9.	Technical Approaches To Chinese Sign Language Processing.	2020	Suhail Muhammad Kamal, Yidong Chen, Shaozi Li,Xiaodong Shi, Jiangbin Zheng	This Technical Review Highlighting The Importance Of Efficient Sign Language Processing Technologies And Offering Valuable Insights For Researchers New To CSL
10.	Indian Sign Language Numeral Recognition Using Region Of Interest Convolutional Neural Network	2018	Beena M V, Sajanraj T D	The Article Presents A Real-Time ISL-To-Text Conversion System Achieving Impressive Accuracies Of 99.56% For The Same Subject And 97.26% In Low Light Conditions, Utilizing A Region Of Interest CNN, Skin Segmentation, And Extensive Training.