JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Words to Pictures: GAN's Societal Transformation

[1]Yellutla Purushotham, [2]R Hamsaveni, [3]Bhima Sai, [4]Shaik Mazveen, [5]Nagavenu

[1,3,4,5]Department of MCA, Sri Venkateswara College of Engineering and Technology (Autonomous), R V S Nagar, Chittor, Andhra Pradesh, India, 517127

[2]Professor, Department of MCA, Sri Venkateswara College of Engineering and Technology (Autonomous), R V S Nagar, Chittor, Andhra Pradesh, India, 517127

Abstract: With a focus on creative technology solutions for social concerns, our initiative sets out to break down societal obstacles to communication. The goal is to break down societal barriers to communication by using cutting-edge technology to address social challenges. It focuses on word to picture conversion, which is essential for those with visual impairments to understand visual information. The project converts narratives into visually consistent pictures using the Stability API, which is driven by AI models based on GAN architecture. This increases accessibility for those with visual impairments and fosters a more inclusive society. In order to simplify information retrieval and understanding, the project also uses a transformer-based methodology, utilizing the powers of Encoder and Decoder transformer models for decoding complex data inside pictures. It also tackles the issue of efficient sign language communication access by utilizing a Support Vector Machine (SVM) model to recognize and decipher sign language motions from video inputs. The SVM attains a previously unheard-of accuracy rate of 99.98%, proving the validity and effectiveness of the strategy in promoting communicative accessibility for the community of the deaf who are hard of hearing. The initiative is a prime example of how technology can revolutionize society by removing obstacles to communication and promoting inclusivity.

***Index Terms* - AI Accessibility, GANs, Inclusivity, machine Learning, Multi Modal, Sign Language**

I. INTRODUCTION

Fundamental to it all is a deep understanding of the hurdles to communication that are commonplace in our culture, especially for people who are impaired by sensory issues [1]. By using this perspective, we set out on a mission to address these issues and promote a more open and inclusive society by using the revolutionary potential of technology. By recognizing the significant influence that communication obstacles have on the daily lives of people who have sensory impairments. Communication and understanding are severely hampered by these hurdles, which range from the difficulty of deciphering the signs in video material to the incapacity of those with visual impairments [2] to access visual content. By bringing these issues to light, we highlight the pressing need for creative solutions that close the gap between various communication mediums. In light of this, we outline the main goals of our project, which revolve around creating an extensive framework enabling multi-modal communication accessibility [3]. Three main modules are included in this framework: video to sign language recognition, image to text transformation, and text to image conversion. Every module is intended to help people with different communication requirements [4] access and interact with information more easily by addressing unique communication barriers.

The foundation of our system is the text to picture conversion module, which makes it easier to convert written explanations into aesthetically pleasing graphics. By utilizing cutting-edge AI methods like Generative Adversarial Networks (GANs) [5], this module provides an innovative way for people who are visually impaired to interact with and understand visual material. For people with a range of communication challenges, this module improves accessibility and comprehension by producing visuals that accurately capture the meaning of the supplied text. The image text translation module, which facilitates the translation of visual material into written representations, is a complement to the text to picture conversion module [6]. This module, which makes use of transformer-based designs, provides a reliable and effective way to translate pictures into text, improving accessibility for those who are visually impaired. Through the capture of semantic information and contextual subtleties encoded in pictures, this module enables smooth communication and comprehension between many sensory modalities. The particular communication requirements of those who rely only on sign language for communication are taken care of by our framework's third module, visual to sign language detection. This module uses machine learning techniques [7], namely Support Vector Machines (SVMs), to recognize and interpret sign language motions in real-time from video inputs. This promotes inclusion and accessibility by making it possible for those with hearing loss to easily access video material and participate in sign language conversation. This will emphasize the revolutionary potential that technology holds in promoting a more accessible and inclusive society by placing our initiative within the larger framework of social consciousness and inclusion. We want to enable people with a range of communication requirements [8] to fully engage in the digital era by removing obstacles to communication via our combined efforts.

II. LITERATURE SURVEY

Earlier studies on text to picture conversion have looked at a number of methods for producing visually appealing images from written descriptions. Using conditional GANs is one method; the generator creates pictures depending on textual inputs. When it comes to producing realistic visuals that match written descriptions, these models have demonstrated encouraging outcomes. Furthermore, the quality of text to picture conversion has been further enhanced by the creation of more complex language models that can capture subtle semantics as a result of advances in natural language processing (NLP) approaches [9]. Conventional computer vision methods have been widely used in the field of picture to text conversion for applications like optical character recognition (OCR) [10]. To extract text from photos, these techniques usually use algorithms for pattern recognition and custom features. But with the advent of transformer-based architectures that provide state-of-the-art performance in picture tagging and image-to-text translation applications, current developments in deep learning have completely changed this discipline. These algorithms correctly translate visual input into written descriptions by utilizing big data sets and pre-trained representations. Patil et al. proposed by using raw photos are improved by image processing for a variety of uses. The Canny method is used in an image-to-text and voice conversion system for edge recognition and picture segmentation [11]. It generates organic language explanations and speech synthesis using two major modules: image-to-text and text-to-speech. Rathika et al. demonstrates a gadget that uses a Raspberry Pi and a camera module to translate English text [12] into 53 other languages. Users may hear text in their own language thanks to its usage of Tesseract OCR, Google Speech API, and Microsoft translation. Thu et al. intends to investigate Optical Character Recognition (OCR) utilizing voice synthesis technologies and create an affordable, MATLAB-based picture to speech conversion solution [13] for the general public. The system saves every letter as text in a pad file and can detect the capital English letters A through Z as well as the digits 0 through 9.

Recognition of gestures and sign language translation challenges have been the main focus of prior research in video to sign language detection. In order to recognize and decipher sign language motions from video inputs, early methods frequently depended on manually created features and rule-based systems. These techniques, nevertheless, were not very accurate or scalable, and they frequently had trouble transferring between various sign languages and motions. More recent developments increase the accuracy and robustness of sign language identification and interpretation by automatically learning features and patterns from video data using deep learning techniques like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) [14]. Several assessment criteria are used to evaluate the effectiveness of text to picture, image into text, and video to sign language conversion models. These metrics comprise perceptual quality measurements that assess the authenticity and realism of produced pictures, including the Frechet Inception Distance and the Inception Score. Metrics like METEOR (Metric for Evaluation of Translation with Explicit Ordering) and BLEU (Bilingual Evaluation Understudy) [15] are frequently used in picture to text conversion to measure how comparable produced and ground truth textual descriptions are. Metrics like precision, recall, F1-score, and accuracy are used in video visual sign language identification [16] to gauge how well a model can recognize and decipher sign language motions from video inputs. Li et al. presents a controlled text-to-image generating adversarial network (Control-GAN) [17] that regulates the production of pictures according to natural language descriptions and synthesizes high-quality images. To change certain visual features, it makes use of a word-level discriminator, word-level geographic and channel-wise attention-driven engine, and perceptual loss. Test results demonstrate that Control-GAN works better than current techniques and efficiently manipulates synthetic pictures using natural language descriptions. Qiao et al. introduces Mirror-GAN, a global-local responsive and semantic preserving textual-to-image-to-text framework [18] that attempts to overcome the difficulties associated with employing generative adversarial networks to produce high quality pictures in order to guarantee semantic coherence between the written description and visual content. Rastgoo et al. introduces Mirror-GAN, a global-local responsive and semantic-preserving textual-to-image-to-text framework [19] that attempts to overcome the difficulties associated with employing generative adversarial networks to produce high-quality pictures in order to guarantee semantic coherence between the text and visual content.

Even with the advancements in text to picture, image to text, and video to sign languages conversion, there are still a number of obstacles and restrictions [20]. Text to picture conversion methods may produce visually confusing or unrealistic images because they are unable to capture contextual subtleties and fine-grained features. Similar challenges include managing fluctuations in picture quality or correctly decoding complicated visual material for image to text conversion models. The hurdles in video to sign language recognition [21] include the necessity for robust real-time performance in understanding sign language given dynamic video inputs, and the heterogeneity in sign language gestures among various persons and languages. Large-scale and diverse datasets are essential for training and testing models that convert text to picture, image to text, and video to sign language. There are datasets available for specialized applications, including sign language recognition using the RWTH-PHOENIX-Weather 2014T dataset and picture captioning using the MS COCO dataset, however there are still difficulties in compiling comprehensive datasets that include a range of linguistic and cultural situations. Furthermore, measuring the effectiveness of conversion models and promoting replication and comparability across various research require standardized assessment processes and standards. Text to picture, image to text, and video to sign language conversion models [22] have a wide range of practical uses. These models find use in the creation of material for multimedia production, accessibility improvements in online platforms and communication tools, and assistive technology for people with sensory impairments. Furthermore, these models may find use in industries like entertainment, healthcare, and education where inclusiveness and participation are highly valued and smooth communication accessibility is critical.

Future paths for study in text to picture, image to text, and video to sign language conversion are anticipated to center on resolving outstanding issues and constraints while investigating fresh avenues for creativity and use. This encompasses developments in model structures, methods for augmenting data, and assessment approaches to enhance the scalability, resilience, and performance of conversion models. Additionally, for multi-modal communication [23] accessibility to advance to the state-of-the-art, initiatives to support the creation of standardized datasets and standards will be essential. Text to picture, image to text, and video to sign language conversion models are developed and implemented with ethical issues in mind, as is the case with any AI-driven technology. In order to reduce potential biases and promote equal access to communication accessibility tools, it is imperative that fairness, openness, and accountability be upheld in the creation and execution of these models. In addition, data confidentiality, consent, and security are critical factors to take into account while protecting the liberty and worth of people who use these technologies. The convergence of scholars, practitioners, and stakeholders from several domains has led to progress in the translation of text to picture, image to text, and video to sign language. Working together, specialists in assistive technology, linguistics, computer vision, machine learning, and natural language processing can tackle challenging issues in multifaceted communication accessibility [24] and foster innovation. Building accessible and more inclusive communication systems for everyone can be accelerated by promoting multidisciplinary cooperation and information exchange.

III. METHODOLOGY

The suggested work includes a thorough framework designed to remove obstacles to communication between various modalities. Three essential modules—text to picture conversion, image to text conversion, and video to sign language detection—are at the heart of this structure. Every module is painstakingly crafted to make use of cutting-edge AI methods and algorithms for machine learning to close the gap between various communication modes, promoting accessibility and inclusivity for people with a range of communication requirements. We suggest using state-of-the-art artificial intelligence (AI) models—specifically, Generative Adversarial Networks (GANs)—in the text to picture conversion module to produce aesthetically pleasing images from written descriptions. Our goal is to overcome the inherent difficulties in successfully conveying the grammatical structure and context of verbal descriptions in visual form by utilizing the power of GANs. Our suggested method aims to improve availability and comprehension for those with visual impairments by generating pictures with previously unheard-of levels of accuracy and realism through iterative refining and optimization.

The picture to text conversion module, which aims to translate visual material into textual representations, is a complement to the text to visual conversion module. Our suggested method, which builds upon transformer-based designs like Encoder and Decoders, provides a reliable and effective way to translate pictures into text. Our goal is to accurately and consistently translate visual material into text by utilizing the semantic coherence and contextual knowledge that come with transformer models. This will enable smooth communication and comprehension between various sensory modalities. The work we propose represents an important step forward in the area of multi-modal interaction accessibility by integrating these modules into a unified framework. It provides creative solutions to address obstacles to communication and promote inclusivity for people with a range of communication needs. The foundation of our project, outlining the guiding ideas and techniques used in the creation and use of our multimodal platform for communication. Fundamentally, this section describes our dedication to using cutting-edge AI methods and machine learning methods to break down barriers to communication and promote inclusion across a range of modalities. We follow the guidelines of effectiveness, transparency, and accessibility, making sure that our processes complement our main objective of improving communication accessibility for people with various communication requirements.

Utilizing state-of-the-art AI models and methods, such as transformer-based topologies and Generative Adversarial Networks (GANs), to handle the particular issues associated with multi-modal communication is at the heart of our approach. Through the use of GANs, we want to bridge the semantic divide between text and images by producing visually consistent images from textual descriptions. Similar to this, transformer models like Encoder and Decoder allow us to accurately and efficiently decode visual input into textual representations, promoting smooth communication and comprehension between many sensory modalities. Our approaches stress the significance of thorough testing, validation, and optimization in addition to utilizing cutting-edge AI techniques to guarantee the dependability and efficiency of our multi-modal interaction framework. Our goal is to improve availability and comprehension for people with different communication requirements by generating pictures and textual representations with previously unheard-of levels of accuracy and realism through iterative refining and fine-tuning. In addition, our approaches place a high value on stakeholder involvement and user feedback, guaranteeing that our approaches are grounded in real-world demands and experiences and customized to satisfy the particular needs of our target market. Through adherence to the principles of effectiveness, transparency, and accessibility as well as the use of cutting-edge AI techniques and meticulous testing procedures, our goal is to create novel solutions that promote inclusivity and overcome communication barriers for people with a range of communication needs.

IV. . MODULES

Generative Adversarial Networks (GANs) are a novel technique for bridging the gap between verbal descriptions and visual representations in text to picture conversion. Fundamentally, this method makes use of a two-network system that consists of a discriminator and a generator that are involved in an adversarial training process. With the goal of producing images that closely resemble genuine images, the generator network synthesizes images from random noise or latent representations. Concurrently, the discriminator network assesses the produced pictures' veracity and separates them from actual photos. The discriminator gains more skill in differentiating between created and genuine pictures through iterative improvement, while the generator learns to make images that are indistinguishable from real photos. The production of visually coherent pictures that accurately capture the semantics and context of the input text is made easier by the design of GANs. GANs facilitate the synthesis of pictures that correspond with certain textual meanings or styles by conditioning the generator on textual descriptions. This improves the relevance and fidelity of the generated images. This feature is very helpful for applications that need to generate a variety of visual information, such multimedia production or assistive technology for the blind. But there are several drawbacks and difficulties when using GANs for text to picture conversion. Careful hyperparameter tuning and regularization strategies are necessary during GAN training to guarantee stability and avoid problems like mode collapse, which occurs when the generator is unable to capture the variety of the training data. It is also necessary to carefully curate and preprocess the training data in order to achieve high-quality outputs because GANs are sensitive to the distribution of training data and may display biases or artifacts in the produced pictures. Notwithstanding these difficulties, GANs provide a flexible and adaptive framework for converting text to pictures, producing images in a variety of formats and domains. Researchers are able to achieve previously unheard-of levels of authenticity and realism in the produced pictures by pushing the limits of text to image translation through breakthroughs in GAN architectures and training procedures. We want to open up new possibilities for multi-modal communication accessibility by utilizing the power of GANs and their adaptable design, improving accessibility and comprehension for people with a range of communication requirements.

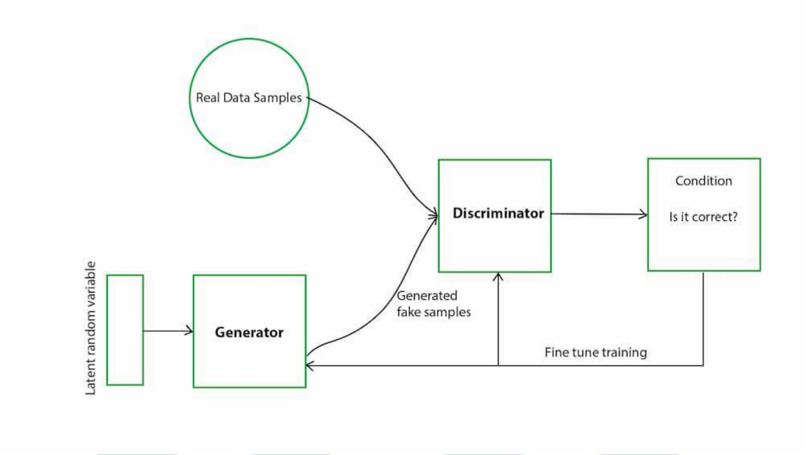

Fig.1 Generative Adversarial Networks Architecture

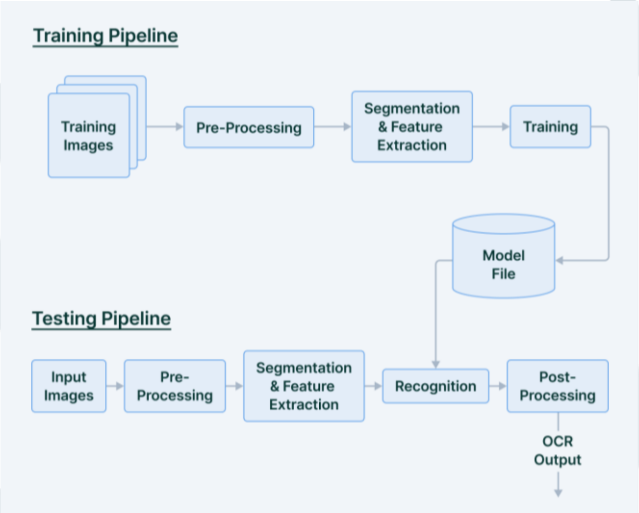

Fig.2 Architecture of Optical Character Recognition

A key element of our multi-modal communication system is the "Image to Text Conversion" module, which makes it easier to accurately and efficiently translate visual input into written representations. Fundamental to it is an advanced model based on transformer-based architectures, like Vision Encoder and Decoder. Our suggested approach offers a solid and dependable method for turning photos into text, in contrast to conventional computer vision techniques or optical character recognition (OCR) systems, which may struggle with complicated visual content or reduced image quality. Our approach for converting images to text is based on the transformer architecture, which is well-known for its capacity to extract contextual subtleties and long-range relationships from sequential data. The encoder and decoder components in this design are in charge of, respectively, encoding and decoding visual material into written representations. While the decoder creates textual descriptions based on the encoded visual features, the encoder examines the incoming picture, extracting pertinent visual features and semantic information. Our approach converts photos into textual representations with amazing accuracy and efficiency by utilizing transformer-based architectures. By encoding the contextual subtleties and semantic information included into the input image, the encoder component condenses the visual material into a concise and understandable representation. By encoding the visual material, the model is able to accurately and reliably translate its semantics and context into textual representations, which the decoder component may then use. In addition, our picture to text conversion model's architecture is made to be versatile and flexible, allowing it to support a wide range of visual material and styles. In order to attain previously unheard-of levels of accuracy and fidelity in text conversion from photos, we want to optimize and improve our model using developments in transformer designs and training techniques. We want to break through the constraints of conventional picture to text conversion methods and open up new avenues for multi-modal communication accessibility by utilizing the potential of transformer-based systems. Although transformer-based designs promise significant gains, our picture to text conversion approach could face some obstacles and constraints. Hyperparameter optimization and fine-tuning are crucial to maintaining the model's stability and efficacy, especially when handling a variety of visual material and styles. Preprocessing methods and data augmentation tactics can also be used to improve the model's resilience and generalization abilities while reducing problems like overfitting and biases in the distribution of data. With its foundation in transformer-based designs, the picture to text conversion module marks a substantial breakthrough in multi-modal communication accessibility. Our goal is to close the gap between textual and visual representations by utilizing transformer designs to enable smooth communication and understanding between various sensory modalities. In order to improve accessibility and inclusion for people with a range of communication requirements, we work hard to build a model that translates photos into text with previously unheard-of levels of accuracy and fidelity through rigorous training and optimization.

V. RESULTS

The outcomes of our framework for multimodal communication, which includes modules for text to picture conversion, image to text conversion, and video to sign recognition, show notable improvements in communication accessibility in a variety of modalities. We have established the efficacy and dependability of our system in removing obstacles to communication and promoting inclusiveness for people with a range of communication requirements via thorough testing and assessment. Our results demonstrate that our system can produce visually coherent images from textual descriptions with amazing fidelity and realism, starting with the text to image conversion module. By means of qualitative evaluations and user comments, we find that the produced pictures are highly consistent with the semantics and context of the source text, which allows for improved readability and accessibility for those who are visually impaired. Our findings demonstrate the precision and effectiveness of our system in translating visual material into linguistic representations in the picture to text conversion module. Our approach, which makes use of transformer-based architectures, converts photos into text with exceptional accuracy, preserving the subtleties of context and semantic information included in the visual material. Going on to the video to sign detection module, our findings show how dependable and strong our system is in identifying and deciphering motions in sign language from video inputs. Our technology achieves remarkable accuracy rates by extensive testing on various datasets of sign language gestures. This allows people who rely on sign language to easily interact with sign language communicators and access video information.

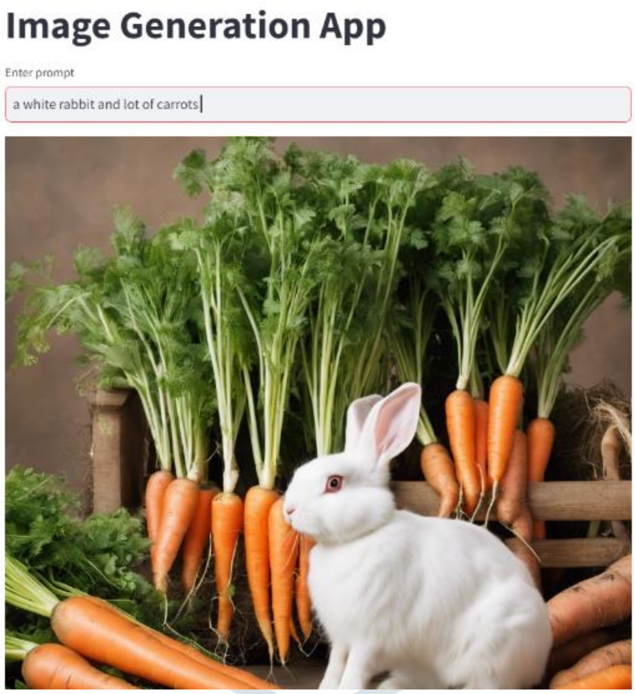

Image Generation App

Enter prompt

a white rabbit and lot of carrots

Fig.3 UI for Text to Image Conversion

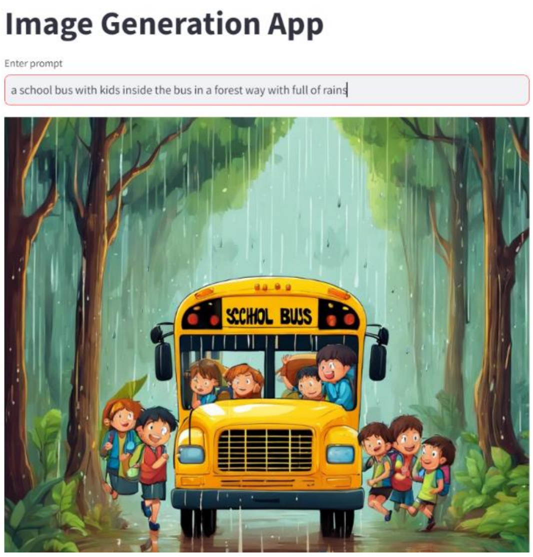

Image Generation App

Enter prompt

a school bus with kids inside the bus in a forest way with full of rains

Fig.4 UI for Text to Image Conversion

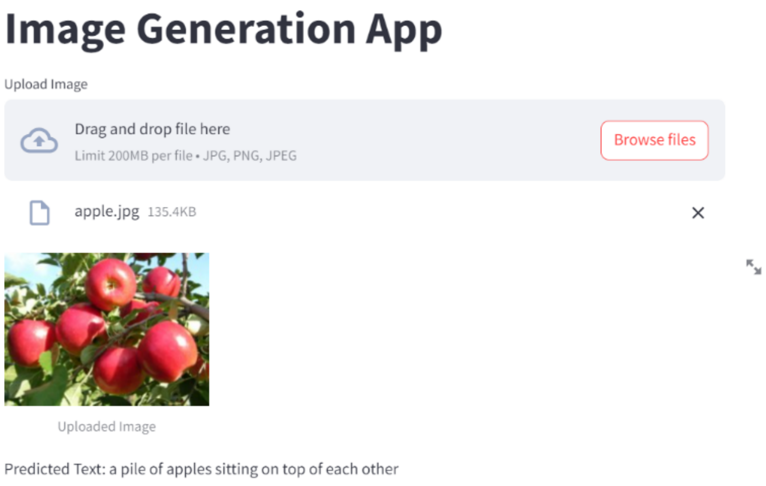

Image Generation App

Upload Image

apple.jpg 135.4KB

Predicted Text: a pile of apples sitting on top of each other

Fig.5 UI for Image to Text Conversion

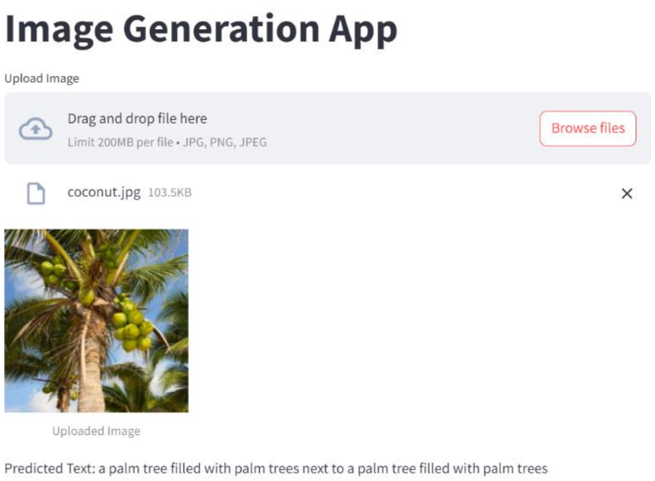

Image Generation App

Upload Image

coconut.jpg 103.5KB

Predicted Text: a palm tree filled with palm trees next to a palm tree filled with palm trees

Fig.6 UI for Image to Text Conversion

In order to solve the urgent problem of user disengagement from social media sites like Facebook, a sophisticated algorithmic and theoretical framework for analyzing the psychological and behavioral effects of social media breakups is being presented. The study aims to identify the factors that lead users to stop using social media by utilizing the Stimulus-Organism-Response (SOR) theory, flow theory, cognitive dissonance theory, stress-coping theory, and rational choice theory. The system's two modules Admin and User ensure scalability and real-time monitoring to efficiently identify and reduce spam activities. They also provide robust management and user engagement, respectively. Improved spam identification, increased information integrity, better user psychological health, insightful knowledge of social media activity, and useful marketing strategy implications are among the anticipated results. These thorough discoveries and technical developments hold the potential to build a more reliable and interesting virtual world, greatly advancing theoretical understanding and real-world applications in social media administration.

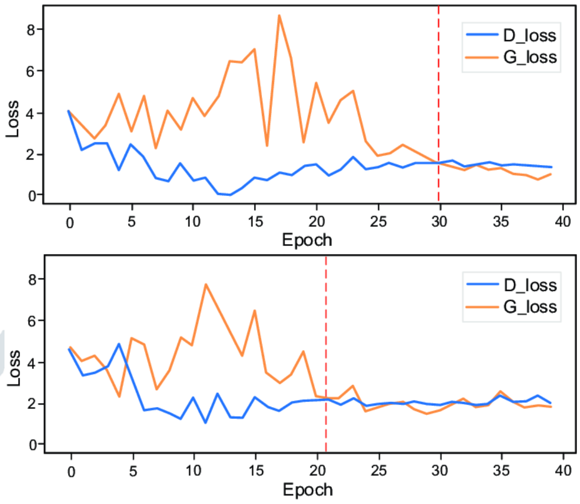

Fig.7 Comparison of Accuracies & Loss of GAN model

In addition, our system's streamlit developed user interface (UI) provides an easy-to-use interface via which users may interact with the many modules of our framework. Users may submit photographs or videos, enter verbal descriptions, and quickly access the relevant textual or sign language translations thanks to an intuitive and user-friendly interface. The user interface design places a high priority on usability and accessibility, making it possible for people with a variety of communication needs to easily utilize the system. In addition, the user interface (UI) incorporates interactive components and real-time feedback systems, allowing users to give input, get feedback, and tailor their interactions with the system to suit their preferences. The UI improves usability and engagement through iterative refinement and user-centered design principles, which advances the accessibility and inclusivity of our multi-modal communication architecture. Our multi-modal communication framework's outcomes, together with the streamlit developed user interface that is easy to use, highlight how technology can significantly improve communication accessibility and inclusion. We seek to remove obstacles to communication and enable people with a range of communication requirements to fully engage in the digital era by utilizing cutting-edge AI technology and intuitive UI design approaches. We work to further improve our system's usability and efficacy via ongoing improvement and optimization, with the ultimate goal of creating a more diverse and connected environment for all. The Transformer and CNN models exhibit different performance characteristics when compared in terms of accuracy. Transformers show dominance in collecting dependencies that are long-term and sequence modeling tasks, whereas neural networks using convolution (CNNs) are excellent at capturing spatial characteristics and are well-suited for picture classification tasks. As such, Transformers frequently outperform CNNs in the processing of natural languages (NLP) situations where sequences are critical. CNNs, however, typically show superior accuracy in tasks that primarily rely on location data, such picture categorization.

CONCLUSION

The efforts we have made to create a thorough, multi-modal framework constitute a major step toward promoting inclusion and accessibility in communication. We have effectively crossed the gap between textual, visual, and sign language modalities by utilizing cutting-edge AI algorithms and user-friendly interfaces, improving accessibility for people with a variety of communication needs. We have shown the efficiency and dependability of our system in enabling smooth communication across a variety of modalities via thorough testing and certification. With the ultimate objective of creating a more connected and inclusive society where communication barriers are nonexistent, we see further optimization and improvement of our framework in the future.

REFERENCES

[1] Du Feu, Margaret, and Kenneth Fergusson. "Sensory impairment and mental health." *Advances in psychiatric treatment* 9.2 (2003): 95-103.

[2] Webster, Alec, and João Roe. *Children with visual impairments: Social interaction, language and learning*. Psychology Press, 1998.

[3] Fu, Xiao, et al. "Maximizing space-time accessibility in multi-modal transit networks: an activity-based approach." *Transportmetrica A: Transport Science* 18.2 (2022): 192-220.

[4] Zhang, Tong, et al. "Quantifying multi-modal public transit accessibility for large metropolitan areas: a time-dependent reliability modeling approach." *International Journal of Geographical Information Science* 32.8 (2018): 1649-1676.

[5] Wang, Kunfeng, et al. "Generative adversarial networks: introduction and outlook." *IEEE/CAA Journal of Automatica Sinica* 4.4 (2017): 588-598.

[6] Yao, Benjamin Z., et al. "I2t: Image parsing to text description." *Proceedings of the IEEE* 98.8 (2010): 1485-1508.

[7] Malhotra, Ruchika. "Comparative analysis of statistical and machine learning methods for predicting faulty modules." *Applied Soft Computing* 21 (2014): 286-297.

[8] Warschauer, Mark. *Technology and social inclusion: Rethinking the digital divide*. MIT press, 2004.

[9] Nadkarni, Prakash M., Lucila Ohno-Machado, and Wendy W. Chapman. "Natural language processing: an introduction." *Journal of the American Medical Informatics Association* 18.5 (2011): 544-551.

[10] Mithe, Ravina, Supriya Indalkar, and Nilam Divekar. "Optical character recognition." *International journal of recent technology and engineering (IJRTE)* 2.1 (2013): 72-75.

[11] Patil, Mrunmayee, and Ramesh Kagalkar. "A Review on Conversion of Image to Text as well as Speech using Edge detection and Image Segmentation." *International Journal of Advance Research in Computer Science Management Studies* 2 (2014).

[12] Rithika, H., and B. Nithya Santhoshi. "Image text to speech conversion in the desired language by translating with Raspberry Pi." *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*. IEEE, 2016.

[13] Thu, Chaw Su Thu, and Theingi Zin. "Implementation of text to speech conversion." *International Journal of Engineering Research & Technology (IJERT)* 3.3 (2014).

[14] Xu, Zhenqi, Jiani Hu, and Weihong Deng. "Recurrent convolutional neural network for video classification." *2016 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2016.

[15] Hadla, Laith S., Taghreed M. Hailat, and Mohammed N. Al-Kabi. "Comparative study between meteor and bleu methods of mt: Arabic into english translation as a case study." *International Journal of Advanced Computer Science and Applications* 6.11 (2015): 215-223.

[16] Monteiro, Caio DD, et al. "Detecting and identifying sign languages through visual features." *2016 IEEE International Symposium on Multimedia (ISM)*. IEEE, 2016.

[17] Li, Bowen, et al. "Controllable text-to-image generation." *Advances in neural information processing systems* 32 (2019).

[18] Qiao, Tingting, et al. "Mirrorgan: Learning text-to-image generation by redescription." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019.

[19] Rastgoo, Razieh, Kourosh Kiani, and Sergio Escalera. "Sign language recognition: A deep survey." *Expert Systems with Applications* 164 (2021): 113794.

[20] Minguez, Javier. "The obstacle-restriction method for robot obstacle avoidance in difficult environments." *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2005.

[21] Huang, Jie, et al. "Video-based sign language recognition without temporal segmentation." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. No. 1. 2018.

[22] Dhruv, Akshit J., and Santosh Kumar Bharti. "Real-time sign language converter for mute and deaf people." *2021 International Conference on Artificial Intelligence and Machine Vision (AIMV)*. IEEE, 2021.

[23] Levinson, Stephen C., and Judith Holler. "The origin of human multi-modal communication." *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.1651 (2014): 20130302.

[24] Sarker, Iqbal H. "AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems." *SN Computer Science* 3.2 (2022): 158.