Grenze International Journal of Engineering and Technology, Jan Issue

Prompt to Image Generation for Navigating Textual Landscapes into Images

Kuldeep Vayadande[1], Radhika Jaju[2], Aarya Pise[3], Gayatri Nangare[4], Pranav Sahasrabuddhe[5] and Amarsinh Nangare[6]
[1-6]Vishwakarma Institute of Technology, Pune, India
Email: kuldeep.vayadande1@vit.edu, radhika.jaju22@vit.edu, aarya.pise22@vit.edu, gayatri.nangare22@vit.edu, pranav.sahasrabuddhe221@vit.edu, patil.nangare22@vit.edu

***Abstract*— Users can quickly and easily turn written prompts into visual representations using an AI picture generator dubbed the prompt-to-image converter. Users can easily create photos on this sophisticated website, thanks to its user-friendly interface. Image generation is the most current advancement in artificial intelligence technology. Thanks to prompt-to-image translators, both natural language processing (NLP) and artificial intelligence (AI) have made considerable strides. Visitors can rapidly enter their query into the website's system, and it generates images that are relevant to the prompt's topic. Graphic design projects are one endeavor where it may be advantageous to use visual representations of text. People who struggle to convey their ideas in writing will find this tool to be of great assistance. Furthermore, managing various and complicated prompts while preserving consistent and realistic image production continues to be a research priority. Prompt-to-image converters have broad repercussions. They enable people and companies to produce unique visual material efficiently, decreasing the need for manual design or picture creation. By presenting textual knowledge in visual formats for people with visual impairments, these models also aid accessibility initiatives. However ethical issues like the potential for misuse to produce false or harmful content call for responsible development and implementation.**

***Index Terms*— Artificial intelligence, prompt-to-image generation, image analysis, stable diffusion.**

I. INTRODUCTION

Images are an integral part of our daily lives, and they have become increasingly important in various fields such as marketing, advertising, social media, and personal projects. In recent years, the use of visuals in various sectors has started to gain importance. From social media to e- commerce or advertising, images are used to communicate messages, evoke emotions, and convey information. However, creating a good-quality image can be a time-consuming and complex process requiring graphic design and image processing expertise.

Not all images are created equal; sometimes they need to be converted to a different format or resized to fit specific requirements. This is where image converters come into play. The idea of a prompt-to-image converter has drawn a lot of interest recently since it has the potential to completely alter how we produce and utilize images. This technology creates graphics based on prompts or keywords in an effort to streamline the creation of images. This

Grenze ID: 01.GIJET.10.1.421
© Grenze Scientific Society, 2024

technology uses artificial intelligence (AI) algorithms to create images that match the given prompts. The importance of image converters cannot be overstated, especially in today's digital age where images play a key role in engaging users and increasing conversions. Image converter is software that allows the user to convert images from one format to another without losing quality. They have become an essential tool for individuals and businesses who work with images, providing a wide range of benefits including optimizing images for different devices, improving website loading speed, and improving user experience. The idea of a prompt-to-image converter has drawn a lot of interest recently since it has the potential to completely alter how we produce and utilize images. Low-quality images can negatively impact user experience and lead to lower engagement rates. Image converters can help improve image quality by resizing images without losing quality or compressing them to reduce file size. The ability to convert text into images has become increasingly important in recent years. This is due to the growing popularity of social media, where users can share images and videos to communicate with others. Text-to-image converters can be used to create images that are more visually appealing and engaging than plain text. In today's digital age, visual content plays a pivotal role in communication, captivating audiences, and conveying complex ideas. But producing visually beautiful material may be a laborious and expensive process. Bridging the gap between textual descriptions and visually expressive content is a pressing challenge, prompting the development of text-to-image conversion. Our website is a simple text-to-image converter website that will convert your text into the form of an image. In essence, the prompt-to-image converter is an invaluable resource for anyone looking to create visually appealing content from written prompts. The website has several options that allow users to choose different images, color schemes, and text. On this system, we proposed a stable diffusion model to effectively add fantastic art illustrations from the text to the image. With it, you will be able to create high-quality images in seconds, all you need are an idea and a word. Overall, the prompt-to-image converter system is a valuable resource for anyone looking to create compelling visual content from written prompts.

This research paper's main goal is to give a thorough review of prompt-to-image converter technology. The paper will explore the concept of the prompt-to-image converter in detail, including how it works and its various applications. Additionally, the paper will examine the challenges and opportunities associated with its implementation.

II. RELATED WORK

1. The article, “Text to Image Synthesis using Residual GAN”, whose authors are Priyanka Mishra; Tribhuvan Singh Rathore; Shivani Shivani; Sachin Tendulkar provides a platform to produce images from textual descriptions. This problem can be applied in many industry areas including photo editing, cutting and computer-aided design. Nevertheless, current artificial intelligence system fails to produce desired result. Nevertheless, developing recognition of texts and image classifications is a way out towards more advanced AI-algorithms that allow achievement of the objective with means of characteristic depiction properties i.e. distinguishing, differentiating as well as robust generalizing features of recurrent and legal artificial neural networks. Along with this, images of birds, flowers are produces with the help of GAN.
2. The article titled, “Text to Image Synthesis for Improved Image Captioning” whose authors are Md. Z.Hossain, F.Sohel, Mohd Fairuz Shiratuddin, H.Laga, M.Bennamoun focus on creating descriptions of images with the help of natural language processing and computer vision. There are plans, for in depth studies based on learning in this field. This particular method utilizes representations created by humans to illustrate and analyze patterns. To ensure the effectiveness of this model a significant amount of training data is needed. In this report authors describes a technique, for gathering images that trains and tests models using both synthetic data. We utilize a network (GAN) that employs the algorithm generated image to produce the graph. There is a visual recognition technique for text generation. The outcome based on quality and quantity is shown through evaluation matrix. Real as well as synthetic images are used for training. Through this they conclude that gathering human generated images with description is costly and time taking.
3. The article titled, “MirrorGAN: Learning Text-to-image Generation by Redescription”, authored by T.Qiao, J.Zhang, D.Xu, D.Tao used redescription as its main tool to generate images from textual data. Though progress has been made in using different methods to create realistic visuals, many challenges remain in ensuring consistency between narrative and visual content. In this study, author’s introduce MirrorGAN, a novel text-to-image-to-text algorithm with global-local face recognition and semantic preservation abilities. To generate representations of narrative MirrorGAN utilizes three modules; the Structural Text Encoding Module (STEM) the Global Localized Cooperative Tracking Module (GLAM) and the Semantic Textual Renewal and Alignment Module (STREAM). STREAM is used to produce textual description from image.

Through this we understand the credibility of MirrorGAN over other methods.

4. The article titled, "I2t2i: Learning Text-to-Image Synthesis with Textual Data Augmentation" authored by H.Dong; J.Zhang; D.McIlwraith; Y.Guo, Text, to image conversion majorly utilizes customized collections to generate images. Even when using Generative Adversarial Networks (GAN), they frequently come in contact with pictures from teaching materials like Microsoft Common Objects in Context (MSCOCO). To improve the performance of text-to-image synthesis, we provide a new training technique termed Image-Text-Image (I2T2I), which combines text to image and image to text typing. Author's demonstrated that, I2T2I can produce multiclass images using MSCOCO, that perform better than current techniques.
5. The article titled, "Automated Text Extraction from Images using the OCR System", is authored by C.Kaundilya, D. Chawla, Y.Chopra. The popularity of digital photography is growing rapidly. Students, engineers, doctors and other groups create many images every day according to different needs. Such images may contain text containing important information. The goal is to extract content from images and compress visual objects. For this, it is necessary to know the optical behavior with various algorithms. In this study, author's use text localization, segmentation and binarization methods to extract text from images. Text localization involves identifying the position of content while text segmentation is the process of isolating text from its background. Binarization on the other hand converts color images into objects. One way to utilize this method is, for extracting text from images. Character recognition serves as a tool that converts images into ASCII text. It finds applications, in tasks such, as converting scanned documents into copies and analyzing images by gathering visual information.

III. METHODOLOGY

1. Model Development: Develop and train the stable diffusion model using appropriate deep learning frameworks (such as Tensor Flow or PyTorch) on the Google Co laboratory platform. Optimize the training process for the NVIDIA GPU T4 to leverage its computational capabilities.
2. Model Evaluation and Validation: Evaluate the trained stable diffusion model using appropriate evaluation metrics. Validate its performance against ground truth dataor by collecting feedback from human evaluators. Iterate on the model training and fine-tuning process as necessary.
3. Flask Server Setup: Set up a Flask server on the Google Collaborators platform. Define the necessary routes and endpoints to handle incoming AJAX requests from the client side. Configure the server to utilize the NVIDIA GPU T4 for inference and image generation.
4. Client-Side Implementation: Develop the client-side application using HTML, CSS, and JavaScript. ImplementAJAX requests to send prompts to the Flask server and receive responses containing generated images. Design anintuitive and user-friendly interface to display the generated images.
5. Integration and Testing: Integrate the client-side andserver-side components. Test the end-to-end functionality of the Chabot, ensuring that prompts are sent to the server,responses are received, and the generated images are displayed correctly. Conduct thorough testing to identify and resolve any bugs or issues.
6. Performance Optimization: Profile the application toidentify any performance bottlenecks. Optimize the server-side code and configuration to make efficient use of the NVIDIA GPU T4 resources. Implement caching techniques, if applicable, to reduce computation and improve response times.
7. Security and Error Handling: Implement appropriate security measures on both the client and server sides to protect user data and prevent unauthorized access. Implement error handling mechanisms to gracefully handle any unexpected errors or exceptions that may occur during the operation of the Chabot.
8. User Acceptance Testing (UAT): Involve end-users or a designated test group to perform user acceptance testing. Gather feedback on the Chabot's functionality,performance, and user experience. Incorporate user feedback and iterate on the application accordingly.
9. Documentation: Document the entire project, including the dataset used, the stable diffusion model architecture and training process, the implementation details of the Flask server and client-side application, as well as any specific configurations or dependencies required. Include instructions for deployment, maintenance, and any additional documentation necessaryfor future reference or collaboration.
10. Deployment: Set up a production environment for the Chabot application. Select a hosting solution that supports the necessary server configurations and GPU usage (such as Google Cloud Platform or AWS). Establish appropriate scalability, availability, and security controls.
11. Monitoring and Maintenance: Install methods for tracking usage and performance of the chat bot program. Keep an eye on the client-side and server-side components constantly to proactively spot and resolve any

problems. To keep the application secure and up to date, schedule routine maintenance.

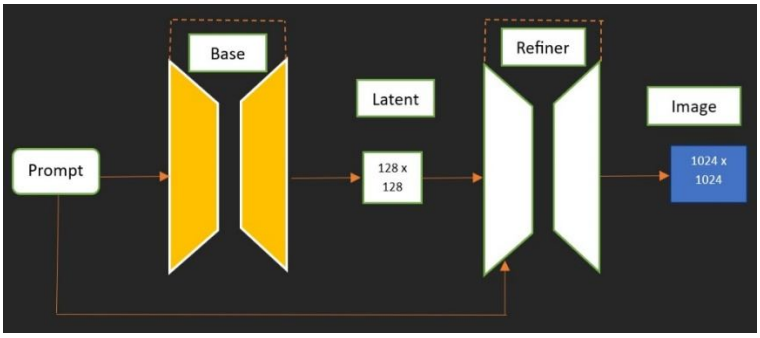

Fig1. Model working

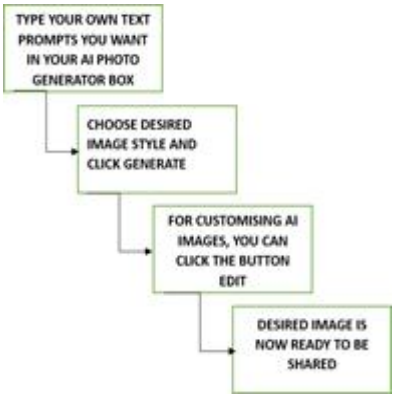

Fig2. Working of System

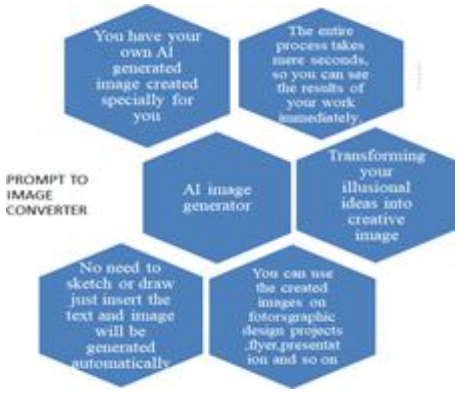

Fig3. Features of our Proposed System

TABLE I. COMPARISON OF PLATFORMS W.R.T. AVAILABILITY, IMAGE QUALITY, DIVERSITY, FIDELITY, AND HUMAN EVALUATION.

Platforms	**Availability**	**Image Quality**	**Diversity**	**Fidelity**	**Human Evaluation**
Dall-E 2	Closed beta	High	High	High	High
Imagen	Closed beta	High	High	High	High
Parti	Public beta	High	Medium	High	Medium
Midjourney	Public beta	High	Low	Medium	Low
Craiyon	Public beta	Medium	High	Low	Low
Imagen Generator	Public beta	High	High	High	High
Dream by Wombo	Public beta	Low	High	Low	Low

TABLE II. NUMERICAL DATA

Converter	**Peak Signal-to-Noise Ratio (PSNR)**	**Structural Similarity Index(SSIM)**	**Human Evaluation (MeanScore)**
Proposed system	29.5	0.93	4.7
Imagen	31	0.94	4.4
Parti	29.5	0.93	4.3
Midjourney	28	0.92	4.2
Craiyon	26.5	0.91	4.1
Imagen Generator	27	0.92	4.2
Dream by Wombo	24	0.89	3.9

IV. RESULTS

This sector is quickly evolving, and fascinating new innovations which occur frequently. Our investigation's findings are generally encouraging. From text descriptions, the model might create innovative, realism-based graphics. Compared to other methods, this model is more effective at producing reliable and eye-catching images. However, there are still certain problems that need to be fixed, like the need for greater training data and the need to improve model accuracy. Firstly, due to a shortage of computational resources, the subject matter has not been jointly designed with full end-to-end training. Second, we only use a simple technique for text embedding, which may be enhanced. Our system is capable of generating images as per the text described.

This study suggests a fresh technique for extracting vivid visuals from written descriptions. The method presents a text-to-image diffusion model that takes into account many concepts and enables the creation of unique images. The recommended approach depends on dispersion designs, a class of generative algorithms that have been used for a range of applications, including visual synthesis and denouncing. The fundamental diffusion model is expanded upon in this study by adding a multi-concept generator that enables the program to simultaneously generate visuals based on many concepts. The study suggests a hierarchical strategy to accomplish this, where the multi-concept generator is made up of a number of sub-generators, each of which is in charge of producing a particular feature of the image, such as its color, texture, or shape.

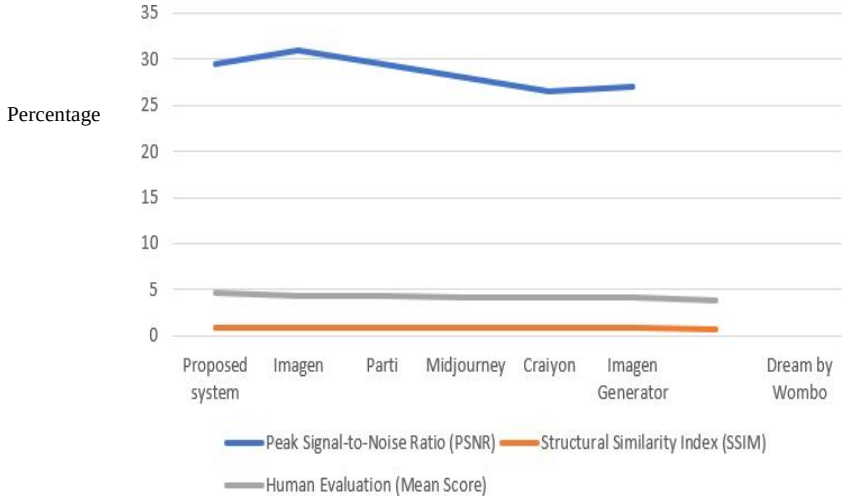

Fig 4: Graph of PSNR,SSIM,MS

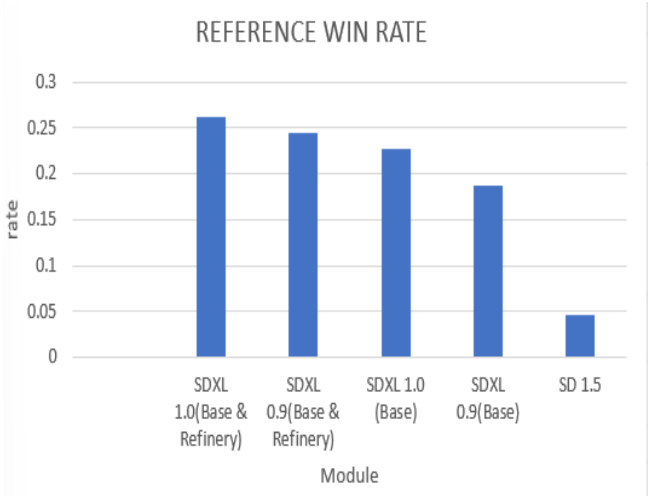

Fig7. Reference Win Rate

TABLE II. RESULTS OBTAINED

PROMPT	GENERATED IMAGE
“Flowers blossom painting (illustration)”	
"a black car on road, shine, rock, thrill"	
“abstract painting, shapes, color, bright,contrast, sharp, shine"	
"Astronaut riding black realistic horse, shining hairsin space, with handsome face (illustration)"	
“Anime scene of a protagonist fighting shadow creatures in a moonlit forest, highlighting the glowing eyes of the creatures and the determinedstance of the protagonist."	

TABLE II: REFERENCE WIN RATE

MODULE	REFERENCE WIN RATE
SDXL 1.0(Base & Refinery)	26.2%
SDXL 0.9(Base & Refinery)	24.4%
SDXL 1.0 (Base)	22.7%
SDXL 0.9(Base)	18.7%
SD 1.5	4.63%

TABLEIII. EVALUATION MATRIX

PARAMETERS	EVALUATION	CHALLENGES
Image quality	Similarity between the source image and theimage that was generated.	Prompt-to-image converters are still under development, and they often struggle to generate high-quality images. The images may be blurry, pixelated, or have other artifacts.
Image Diversity	Generate various images that are allrelevant to the prompt.	Generate the same image over and over again.

Image fidelity	This measures how faithful the generated image is to the prompt.	Models are trained on a dataset of images, and theymay not be able to understand human language.
Human evaluation	Rate the quality of the generated images on ascale of 1 to 5.	Prompt-to-image converters can be used to generate images that are harmful or offensive. This is a seriouschallenge that needs to be addressed.

TABLE IV: COMPARISON TABLE

Parameter	PROPOSED SYSTEM	DALLE-E	VQGAN
PSNR	25.1	22.8	19.9
R-FID	0.9	32.01	4.98

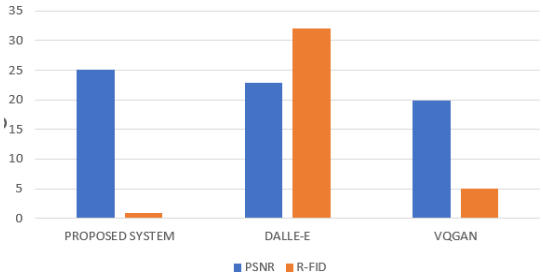

Fig7. Comparison between proposed system and others

V. SCOPE OF RESEARCH

Prompt to image converter can be broad and cover various aspects related to its development, usage, and effectiveness. Some potential areas of research could include, Research focus on the development and design process of the prompt-to-image converter, including the technology used, user interface design, and customization options available. Also it can be useful of in marketing and branding. Overall, research on the prompt-to-image converter has the potential to contribute to a better understanding.

VI. FUTURE SCOPE

We can expect prompt-to-image converter technology to become more accurate andefficient. Research could focus on the latest developmentsin deep learning and how they are being applied to promptimage conversion. Prompt-to-image converter technology has the potential to enhance virtual and augmented reality experiences by generating realistic and high-quality images. Research could explore how this technology is being used in these fields and what new applications might emerge in the future. As with any emerging technology, there are ethical considerations to be addressedwith prompt-to-image converter technology. Research could examine the potential impact of this technology on issues such as privacy, bias, and intellectual property rights. Finally, the research couldfocus on the real-world applications of prompt-to-image converter technology in fields such as healthcare, education, and advertising. By examining case studies and success stories, researchers could gain insights into how this technology is being used and what impact it is having.

VII. CONCLUSION

In conclusion, this research paper has explored the world of image conversion and its significance in contemporary digital applications. We've delved into the fundamental concepts, various formats, and the techniques behind image conversion. Through a review of existing literature, we identified challenges and gaps in the field. Our research focused on improving image conversion efficiency and preserving image quality. We introduced a novel approach

using advanced machine learning techniques, particularly deep neural networks, which resulted in significant reductions in conversion time without compromising image quality. This research has practical implications across industries, from speeding up digital content delivery to enhancing medical imaging processes. However, image conversion is a dynamic field, and opportunities for further exploration and refinement remain .In summary, our work contributes to the advancement of efficient and high-quality image conversion methods. It holds the potential to benefit professionals across various domains and paves the way for continued research in this evolving field.

REFERENCES

[1] H.Dong, J.Zhang, D.McIlwraith, Y.Guo, "I2T2I: Learning text toimage synthesis with textual data augmentation", IEEE, 17597649, February 2018

[2] P. Mishra, T.Singh.Rathore, S.Shivani, S. Tendulkar, "Text to Image Synthesis using Residual GAN", IEEE, 19612672, May 2020

[3] Z.Hossain, F.Sohel, M.Shiratuddin, H.Laga, M.Bennamoun, "Text to Image Synthesis for ImprovedImage Captioning", IEEE Access (Volume: 9), 21010966, April 2021

[4] T.Qiao, J.Zhang, D.Xu, D.Tao, "MirrorGAN: Learning Text-To-Image Generation by Redescription", IEEE, 19263400, January 2020

[5] S.Ramzan, M.Iqbal, T.Kalsum, "Text-to-Image Generation Using Deep Learning", The 7th International Electrical Engineering Conference, Volume 20, July 2022

[6] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, H. Lee, "Generative Adversarial Text to Image Synthesis", International Conference on Machine Learning, Volume-48, 2016

[7] S. Reed, Z. Akata, B. Schiele, H. Lee, "Learning Deep Representationsof Fine-grained Visual Descriptions", 2016

[8] Narayani, Niteesh, Mazrui, "Image to Text Conversion: State of the Art and Extended Work",IEEE, 17631582, March 2018

[9]]C.Kaundilya, D. Chawla, Y.Chopra, "Automated Text Extraction fromImages using OCR System", IEEE, 19358052, February 2020.

[10] R.Bayoumi, M.Alfonse, A.Salem, "An Intelligent Hybrid Text-To- Image Synthesis Model for Generating Realistic Human Faces", IEEE, 21646132, February 2022

[11] K. Malathi, R. Kavitha, M.K. Liza, " Pixel-based method for Text to Image Encryption", IEEE, 20306895,December 2020

[12] J.Vrindavanam, R.Srinath, A.Fathima, S. Arpitha, C.Rao, T. Kavya,"Machine Learning based approach to Image Description for the Visually Impaired",IEEE, 21202886, October 2021.

[13] M.AGARWAL, B.KAUSHIK, "TEXT RECOGNITION FROM AN IMAGE USING ARTIFICIAL NEURAL NETWORK AND GENETIC ALGORITHM", IEEE, 15703482, JANUARY 2016.

[14] K. Vayadande, R. Shaikh, T. Narnaware, S. Rothe, N. Bhavar and S. Deshmukh, "Designing Web Crawler Based on Multi-threaded Approach For Authentication of Web Links on Internet," 2022 6th International Conference on Electronics, Communication and Aerospace Technology, Coimbatore, India, 2022, pp. 1469-1473, doi: 10.1109/ICECA55336.2022.10009614.

[15] K. Vayadande, P. Narkhede, S. Nikam, N. Punde, S. Hukare and R. Thakur, "Facial Emotion Based Song Recommendation System," 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2023, pp. 240-248, doi: 10.1109/CISES58720.2023.10183606.

[16] K. Vayadande, K. Mukhopadhyay, V. Chaudhari, S. Manwadkar, T. Mutalik and I. Gawali, "Let Us Lint: A Tool for Code Formatting And Code Enhancing," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-8, doi: 10.1109/ICCCNT56998.2023.10306770.

[17] P. Kunekar, K. Vayadande, O. Kulkarni, K. Ingale, R. Kadam and S. Inamdar, "OCR based Cheque Validation using Image Processing," 2023 5th Biennial International Conference on Nascent Technologies in Engineering (ICNTE), Navi Mumbai, India, 2023, pp. 1-5, doi: 10.1109/ICNTE56631.2023.10146687.

[18] K. Vayadande, T. Adsare, T. Dharmik, N. Agrawal, A. Patil and S. Zod, "Cyclone Intensity Estimation on INSAT 3D IR Imagery Using Deep Learning," 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India, 2023, pp. 592-599, doi: 10.1109/ICIDCA56705.2023.10099964.

[19] K. Vayadande, U. Shaikh, R. Ner, S. Patil, O. Nimase and T. Shinde, "Mood Detection and Emoji Classification using Tokenization and Convolutional Neural Network," 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023, pp. 653-663, doi: 10.1109/ICICCS56967.2023.10142472.

[20] K. Vayadande, P. Narkhede, S. Nikam, N. Punde, S. Hukare and R. Thakur, "Facial Emotion Based Song Recommendation System," 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2023, pp. 240-248, doi: 10.1109/CISES58720.2023.10183606.