FACIAL EMOTION MODULATION: LEVERAGING STAR GENERATIVE ADVERSARIAL NETWORKS FOR EXPRESSION TRANSFER

[1]Syed Sama Sumayya, [2]Singamsetty Viswas
[1,2]UG Student, [1,2]Department of Computer Science & Engineering, Geethanjali Institute of Science and Technology, Gangavaram, Andhra Pradesh, India

ABSTRACT
The development of computer vision, image transfer a'nd facial expression transfer has been more and more widespread applications. But still there are some more problems, such as lack of realistic expression, poor retention of facial identity features and low synthesis efficiency. In order to solve these problems of facial expression transfer, our project proposes a facial expression transfer model based on Star Generative Adversarial Network, which can generate a highly realistic face image with source facial expression and target facial identity features.The model consists of four parts: The generator takes target domain labels, and style codes to produce output images that retain the content of the input while adopting the style of the target domain.The style encoder encodes reference images from the target domain into latent style codes.The mapping network generates additional style code and given to the discriminator to check whether it is real or fake. The model generates the image which is more realistic than the other models with source facial expression and target facial identity features.

INTRODUCTION
StarGAN v2 represents a significant breakthrough in the realm of image-to-image translation models, addressing key challenges faced by previous methodologies. Developed by Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha, StarGAN v2 introduces innovative techniques to enhance the diversity, scalability, and visual quality of generated images across multiple domains. This project provides a comprehensive overview of StarGAN v2, highlighting its architecture, key features, improvements over StarGAN v1, and its significance in the field of computer vision.

At the core of StarGAN v2 lies its innovative generator and discriminator architecture,which have been carefully designed to produce high-quality images with enhanced diversity. StarGAN v2 adopts a unified generator network capable of synthesizing images across various domains. This architectural refinement not only streamlines the model but also improves its performance in generating diverse and realistic images.

A key feature of StarGAN v2 is the incorporation of Adaptive Instance Normalization(AdaIN), a technique that enables precise control over style transfer between different domains. By dynamically adjusting the style of generated images while preserving their content, AdaIN helps to ensure that the synthesized images accurately reflect the desired characteristics of each target domain. StarGAN v2 introduces domain-adaptive initialization for the generator network,allowing it to capture the unique features of each domain more effectively during training. This initialization strategy enhances the diversity and realism of generated images

MOTIVATION
This project aims to enhance the diversity and realism of generated images compared to previous methods. Traditional approaches to image synthesis often struggle to produce diverse images, leading to limited variability and potentially unrealistic results. StarGAN v2 addresses this limitation by introducing novel techniques such as Adaptive Instance Normalization (AdaIN) and domain-adaptive initialization, which enable the model to generate images with enhanced diversity while preserving their realism and fidelity.

The development of StarGAN v2 is motivated by the need for more efficient, flexible, and realistic image synthesis techniques capable of meeting the demands of modern applicationsand research challenges. The key motivation for StarGAN v2 is its potential applications in various fields such as computer vision, graphics, and human-computer interaction.

By enabling the generation of diverse images across multiple domains, StarGAN v2 can facilitate tasks such as image translation, style transfer, data augmentation, and more. These capabilities have the potential to drive innovation and advancement in a wide range of domains, including entertainment, e-commerce, healthcare, and education. By pushing the boundaries of what is possible in image synthesis, StarGAN v2 opens up new opportunities for creativity, exploration, and innovation in the field of computer vision and beyond.

DEEP LEARNING

Artificial intelligence (AI) and machine learning techniques called deep learning model how people acquire specific types of information. Data science, which also encompasses statisticsand predictive modelling, contains deep learning as a key component. Deep learning makes this process quicker and simpler, which is very advantageous to data scientists who are entrusted with gathering, analysing, and interpreting massive amounts of data.

Deep learning can be viewed as a means to automate predictive analytics at its most basiclevel. Deep learning algorithms are piled in a hierarchy of increasing complexity and abstraction, as opposed to conventional machine learning algorithms, which are linear. Similar to how a toddler learns to recognize the dog, deep learning computer programmes go through similar stages. Each algorithm in the hierarchy performs a nonlinear transformation on its input and outputs a statistical model using what it has learned.

Deep learning computer programmes go through a similar process to a young child learning to recognize a dog. Each algorithm in the hierarchy performs a nonlinear transformationon the data it receives as input before using what it discovers to produce a statistical model as anoutput. Till the output is accurate enough to be relied upon, iterations are performed. Deep was given its name because of the quantity of processing layers that data must go through.

The learning process in typical machine learning is supervised, and the programmer must be very explicit when instructing the computer what kinds of things it should be looking for to determine whether or not an image contains a dog. The computer's success rate in this painstaking procedure, known as feature extraction, entirely hinges on the programmer's ability to precisely define a feature set for dogs. Deep learning has the advantage that the programme develops the feature set independently and without supervision. In addition to being quicker, unsupervised learning is typically more accurate.

Initially, the computer program might be provided with training data -- a set of images for which a human has labelled each image dog or not dog with metatags. The program uses theinformation it receives from the training data to create a feature set for dog and build a predictivemodel.

ABOUT THE PROJECT

This project aims to develop a unified framework capable of synthesizing diverse images across multiple domains using a single model. Building upon the success of its predecessor, StarGAN, this enhanced version introduces novel techniques and improvements to achieve better performance, flexibility, and realism in image synthesis. Developing a unified frameworkthat can generate high-quality images with diverse styles and characteristics across a wide rangeof domains.

Unlike traditional approaches that rely on separate models for each domain, StarGAN v2 offers a streamlined solution that simplifies the training and deployment of image synthesis systems. This project represents a collaborative effort to push the boundaries of image synthesis and domain adaptation. By combining advanced techniques from deep learning, computer vision, and image processing, the project aims to develop a versatile and efficient solution for generating diverse images across multiple domains, opening up new opportunities for research,development, and application in a wide range of fields.

OBJECTIVE

The objective of this project is to create an advanced image synthesis framework capableof generating diverse, high-quality images across multiple domains using a single model. This project aims to streamline the image synthesis process by developing a unified architecture thateliminates the need for separate models for each domain, enhancing efficiency and scalability. By incorporating novel techniques such as Adaptive Instance Normalization (AdaIN) and domain-adaptive initialization, the goal is to improve the diversity and realism of generated images compared to previous methods. This project aims to optimize training and inference processes to achieve better efficiency and make the framework suitable for real-world applications.

Literature Review

High fidelity facial expression transfer using part based local global conditional GANs

Muhammad Mamunur Rashid, Yongwei Nie & Guiqing Li

This model seems to be a method for transferring facial expressions from one imageto another with high fidelity. This part-based approach allows for more precise control and manipulation of facial expressions,different parts of the face contribute differently to overall expression. The system likely utilizes conditional GANs to achieve high-fidelity facial expression transfer, allowing for realistic manipulation of facial expressions in images.

Child GAN: Large scale synthetic child facial data using domain adaptation in StyleGAN

Muhammad Ali Farooq, Wang Yao, Gabriel cost ache, Peter Corcoran

This approach leverages domain adaptation techniques to bridge the gap between synthetic and real child faces, enhancing the realism and diversity of generated images. By integrating domain-specific knowledge into StyleGAN, the system aims to produce high quality synthetic child faces suitable for various applications, such as training facial recognition systems or augmenting datasets for child-related research.

Face image synthesis driven by geometric feature and attribute label

F. Y. Dai, J. Chi, M. G. Ren and Q. D. Zhang

The new face synthesis model can generate a highly realistic face image which ownsthe expression of the source face, the identity of the target face and the specified attribute. Thenew model consists of two parts: facial landmark generator(FLMG) and geometry and attribute aware generator(GAAG).FLMG uses the facial geometric feature points to encode the expression information, and transfers the expression from the source to the target face in the form of feature points.

Detailed features preserving 3D facial expression transfer

Z. Yu, J. Chi, Y. Ye and F. Dai

This model aims to accurately transfer facial expressions while preserving intricate details in a three-dimensional space. Leveraging advanced techniques, it captures subtle nuances like wrinkles and muscle movements. Through high-fidelity representation and expression mapping, it ensures faithful yet adapted transfer between different facial models. Special attention is given to preserving fine-grained details and distinctive facial features, enhancing realism.

Facial expression generation method based on enhanced conditionalgenerative adversarial networks

X. X. Wang, F. F. Li, and Q. Chen

This generation method utilizes conditional inputs to guide the generation process,ensuring desired expression outcomes. Through enhancements in architecture and training strategies, the method achieves improved fidelity and diversity in generated expressions. By conditioning on specific attributes or emotions, it enables precise control over the generated output. Deep learning techniques, possibly including attention mechanisms and advanced loss functions, enhance the network's capability to capture subtle expression variations.

Learning to Cartoonize Using White-box Cartoon Representations

Wang, Xinrui et al.

This model aims to generate cartoon-style images by leveraging white-box representations of cartoons. Through deep learning techniques, the system learns to map input images to cartoon-like representations while preserving essential features. White-box representations provide interpretability and control over the cartoonization process, allowing users to adjust style and detail. By training on a diverse dataset of real images thesystem learns to generalize cartoon styles effectively.

Discriminately deep fusion approach with improved conditional GANs for facial expression recognition
Zhe Sun, Hehao Zhang, Jiatong Bai, Mingyang Liu, Zhengping Hu

This model includes by fusing deep features discriminately, it captures richrepresentations of facial expressions. Advanced conditional GANs facilitate generating expressive and diverse facial images for training. Through iterative refinement and feedback, the system ensures robust recognition across various expressions and subjects. Leveraging deep learning techniques, including attention mechanisms and feature fusion, it enhances the model's ability to discern subtle expression variations.

Facial expression video generation based on spatio-temporalconvolutional GAN:FEV GAN
Hamza Bouzid, Lahoucine Ballihi

The Facial Expression Video Generation system, FEV GAN, utilizes spatio- temporal convolutional Generative Adversarial Networks (GANs) to generate realistic facial expression videos. By leveraging both spatial and temporal information, the model captures dynamic facial movements and expressions over time. The spatio-temporal convolutional architecture enables the model to capture both local facial features and theirtemporal dependencies, results in lifelike and synchronized expressions.

Facial expression transfer based on conditional generative adversarialnetworks(GANs)
Yang fan, Xingguo jiang, Shuxing lan, and jianghai lan

The proposed facial expression transfer system utilizes Conditional Generative Adversarial Networks (GANs) to seamlessly transfer facial expressions between images. Through conditional inputs, the model learns to map expressions from a source face to a target face while preserving identity and context. By conditioning on specific facial features or emotions, it ensures accurate and personalized expression transfer. Leveragingadversarial training, the system generates realistic and coherent expressions that blend naturally with the target face

EXISTING SYSTEMS

3-DIMENSIONAL MORPHABLE MODEL:

3D Morphable Models(3DMM) are a type of statistical model used to represent the shape and texture variations of 3D objects, typically faces in a simple and efficient manner. Each face can be thought of as being made up of two main components: shape and texture. The shape captures variations in facial geometry, such as differences in face width, length, and proportions.

3DMM works by anlyzing a large dataset od 3D scans of faces to understand the typicalvariations in both shape and texture across different faces. Through mathematical techniques like Principal Component Analysis (PCA),3DMM extracts the main patterns of variability in the dataset and represents them as principal components.

These principal components can then be combined in different ways to represent a widerange of faces. By adjusting the weights assigned to each principal component, you can generate new faces with varying shapes and textures. This allows 3DMMs to be used for taskslike face reconstruction from images, where the goal is to estimate the 3D shape and texture ofa face given only a 2D image. These models are constructed from a large dataset of 3D scans or images of objects captured under different lighting conditions, poses, and expressions.

PROPOSED SYSTEM

The proposed system of facial expression transfer based on Star Generative Adversarial Networks (GANs) offers a sophisticated approach to seamlessly transfer facial expressions between images. At its core, the system employs a Star GAN architecture, where the generator is conditioned on both the source and target faces along with the desired expression. This conditioning enables the model to learn the mapping between different facial expressions whilepreserving individual identity and context.

During training, the system learns to generate images that not only accurately reflect the desired expression but also maintain the unique facial characteristics of the target individual. Through adversarial training, the generator improves its ability to produce realistic and visuallyconvincing results,

The Key components include data preparation with diverse facial expression datasets,a Star GAN architecture for precise control over expression transfer, and training strategies employing techniques like minibatch discrimination and spectral normalization for stability and convergence. The system utilizes an encoder-decoder architecture for feature encoding and decoding, translating facial features into latent representations and generating output images with the desired expression. Loss functions such as adversarial loss, reconstruction loss, and identity loss are incorporated to ensure realism, consistency, and preservation of identity duringexpression transfer.

Fine-tuning and evaluation processes optimize system performance, considering both quantitative metrics and qualitative assessments by human observers. This proposed system holds promise for applications in animation, virtual communication, and emotional analysis, offering realistic and natural facial expression transfer capabilities.

SYSTEM ARCHITECTURE

The System consists of the following steps :-

1. Generator
2. Style Encoder
3. Mapping Network
4. Discriminator

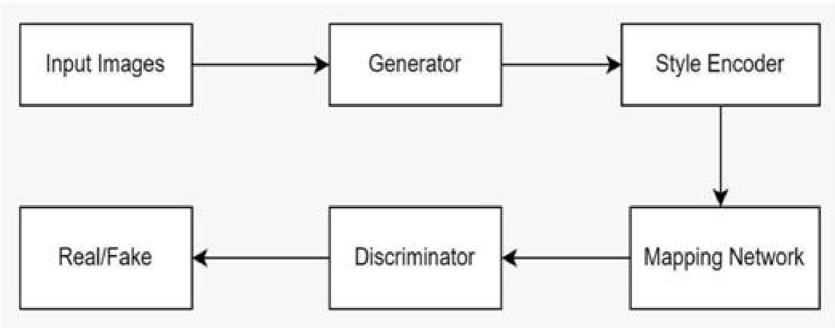

Fig: 1 Architecture of Facial Expression Transfer

GENERATOR

The generator takes input images from different source domains, along with target domain labels, as conditional inputs. It learns to generate images that not only resemble the input images but also belong to the specified target domain. The generator learns to adapt the visual characteristics of input images to the target domain while preserving essential features and details. It achieves this by incorporating domain-specific information from both the input images and target domain labels. Unlike traditional GANarchitectures where each domain has its own generator, StarGANs use a single unified generator for all domains. The generator undergoes adversarial training alongside the discriminator, aiming to generate images that are indistinguishable from real images in thetarget domain.

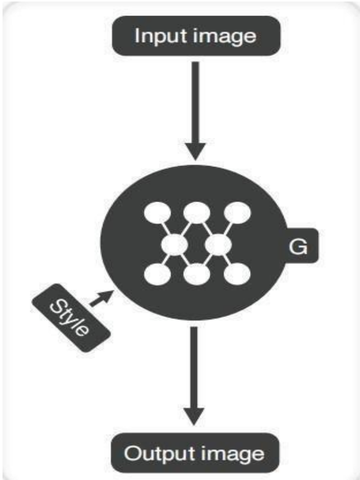

Fig :2 Generator

STYLE ENCODER

The style encoder is responsible for extracting style features from input images. It playsa crucial role in the image-to-image translation process, facilitating the generation of images that exhibit desired attributes or characteristics.The style encoder takes input images from different source domains and extracts latent style features from them.These style features capture essential characteristics of the input images, such as colors, textures, shapes, and otherdomain-specific attributes.

The style encoder maps input images to a shared latent space, where style features are represented as vectors or embeddings.This shared latent space allows for the transfer of style information across different domains, enabling flexible image translation between arbitrary domain pairs. By encoding input images into a shared latent space, the style encoder facilitatesdomain adaptation, enabling the model to learn domain-invariant representations. This helps ingenerating realistic images in target domains, even when there is a lack of paired training data.

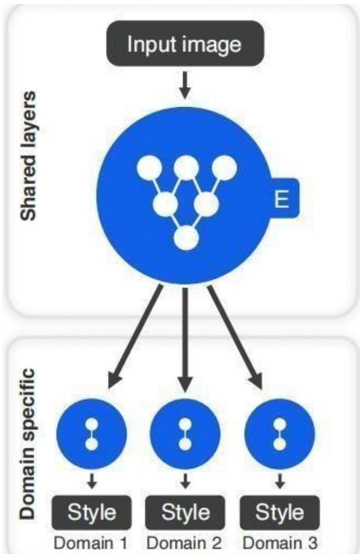

Fig : 3 Style Encoder

MAPPING NETWORK

The mapping network serves to translate latent vectors from the input domain to latentvectors in the target domain. The mapping network takes as input a latent vector sampled fromthe input domain's latent space. This latent vector represents the style characteristics of an input image from a particular domain. This adaptation process enables the model to translate style information from one domain to another, facilitating image-to-image translation across multiple domains.

The mapping network maps the latent vectors from the input domain to a shared latent space, where style features are represented consistently across all domains. This shared latent space allows for the transfer of style information between different domains, enabling flexible image translation between arbitrary domain pairs. The latent vectors generated by themapping network serve as conditional inputs to the generator network.

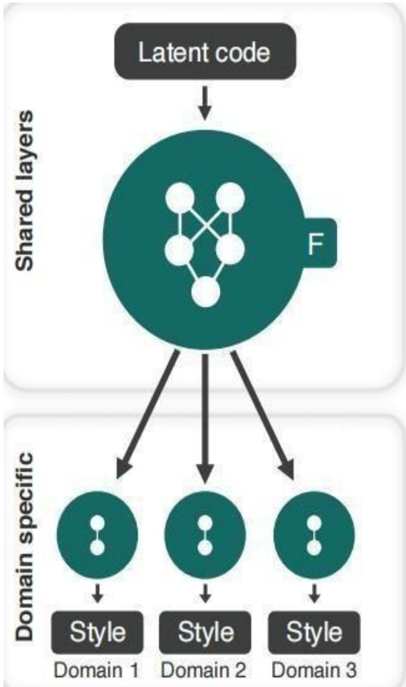

Fig: 4 Mapping Network

The mapping network provides fine-grained control over image generation by translating latent vectors between different domains. This allows for precise manipulation of image attributes during the translation process, such as modifying facial expressions, changinghair styles, or altering clothing types. The mapping network undergoes adversarial training alongside the generator and discriminator networks. It learns to translate latent vectors in a waythat maximizes the realism of generated images in the target domain, while also fooling the discriminator into classifying them as real.

Overall, the mapping network in the StarGAN architecture plays a critical role in translating latent vectors between different domains, facilitating flexible and efficient image- to-image translation across multiple domains. The mapping network generates another style code froma noise vector.They guide the generation process, allowing the generator to produceimages that exhibit the desired attributes specified by the target domain label. The mapping network learns to adapt the latent vectors from the input domain to the latent vectors in the target domain.

DISCRIMINATOR

The discriminator plays a crucial role in the adversarial training process, guiding the generator to produce realistic images in target domains. The primary role of the discriminatoris to distinguish between real images from the target domain and fake images generated by the generator. It is trained to classify whether an image is real (from the target domain) or fake (generated by the generator). During training, the discriminator provides feedback to both the generator and itself. It provides feedback to the generator by indicating how well it has managed to generate realistic images, encouraging the generator to produce more realisticimages over time.Additionally, the discriminator itself is trained to improve its ability to distinguish between real and generated images, leading to a more effective discriminator overtime.

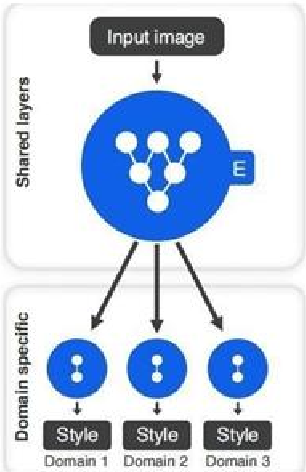

Fig.5. Discriminator

The discriminator also aids in domain adaptation by learning to recognize domain specific features and characteristics. By discriminating between real images from different domains, it helps the model learn domain-invariant representations that facilitate image translation across multiple domains. During training, the discriminator's parameters are optimized using gradient descent to minimize its classification loss. This optimization process involves adjusting the discriminator's parameters to improve its ability to accurately classify real and generated images. The discriminator plays a crucial role in stabilizing the training process of the entire StarGAN model. Its adversarial feedback helps prevent mode collapse andencourages the generator to produce diverse and realistic images.

Results & Analysis

Evaluation Criteria

In this study, we construct a system for generating a realistic image.Assess the quality of the generated images visually. In this study, we construct a system for generating a realistic image.Assess the quality of the generated images visually. Look for artifacts, blurriness, and overall fidelity to the target expression. In this method, we: a) Use Identity preservation metrics evaluatewhether the generated images maintain the identity features of the person depicted in the input images. b) Evaluate how well the generated images convey the target expression. Ensure that thefacial features and expressions align with the intended emotion. c) Assess whether the generated images appear contextually consistent with the surrounding environment and facial context. d) Assess the time taken to generate facial expressions from input images in real-time or near-real-time scenarios. Consider the model's speed and efficiency for deployment in interactive applications.

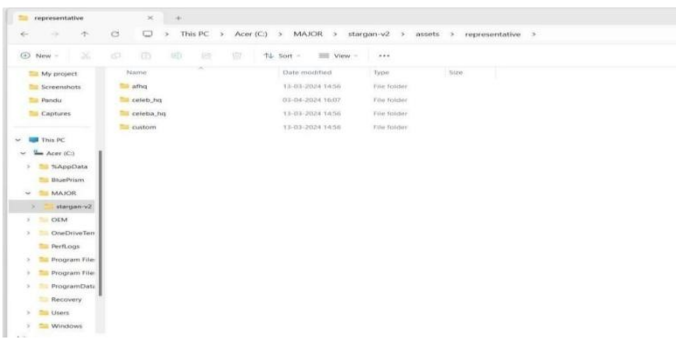

Fig.6. The folder which consists of source and reference images

Description : Fig 6 describes the folder which consists of the source images and referenceimages is taken from the online and uploaded.

This folder includes two sub folders whose names are ‘src’ for source and ‘ref’ for reference. Itis a sub folder of a representative folder which consists of all the images.

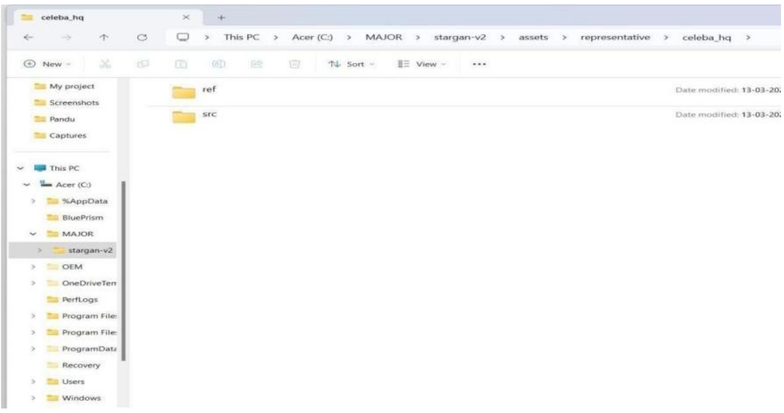

Fig : 7. The source and reference folders

Description : Fig 7. describes the folders of the source and reference where theimages will be uploaded.

The source folder consists of two sub folders named as female and male. The reference folderalso consists of two sub folders named as female and male.

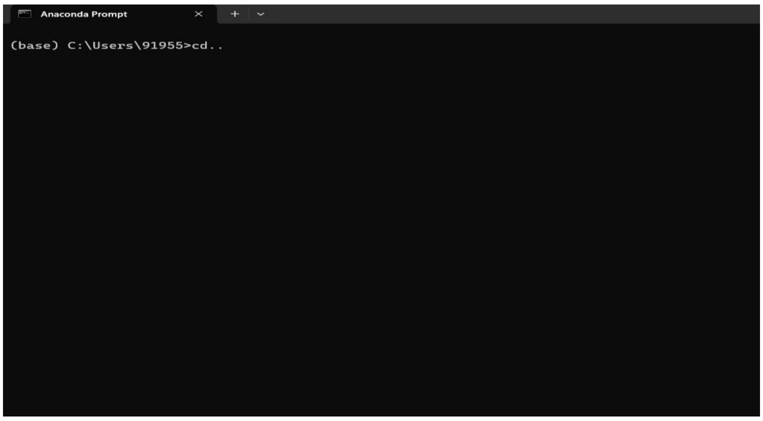

Fig 8 Anaconda Prompt

Description : Fig 8, describes the Anaconda prompt where the execution starts , By using‘cd..’ command we can change the directory.

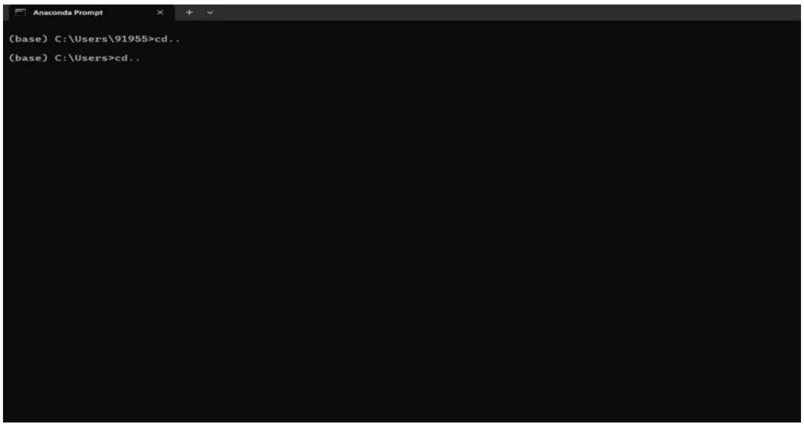

Fig : 8 Changing Directory to the ‘users’

Description : Fig 8, describes the Anaconda prompt from the users directory we arenavigating into the ‘C’ drive with the help of ‘cd..’ command.

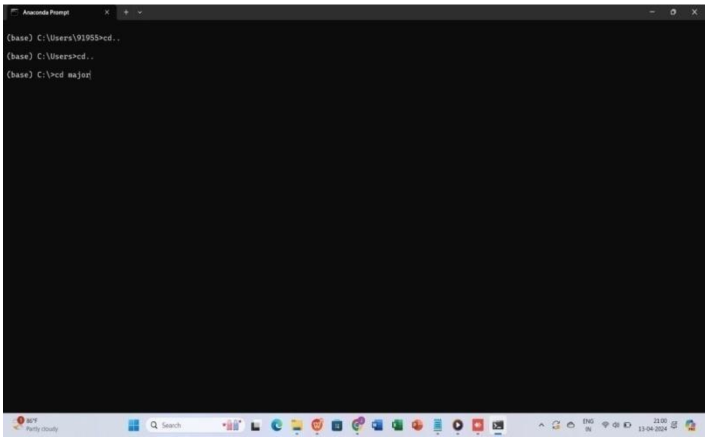

Fig 9 Changing Directory to the major folder

Description : Fig 9, describes the Anaconda prompt after navigation to the ‘c’ drive we arenavigating to the major folder. This is the folder where our project is saved.

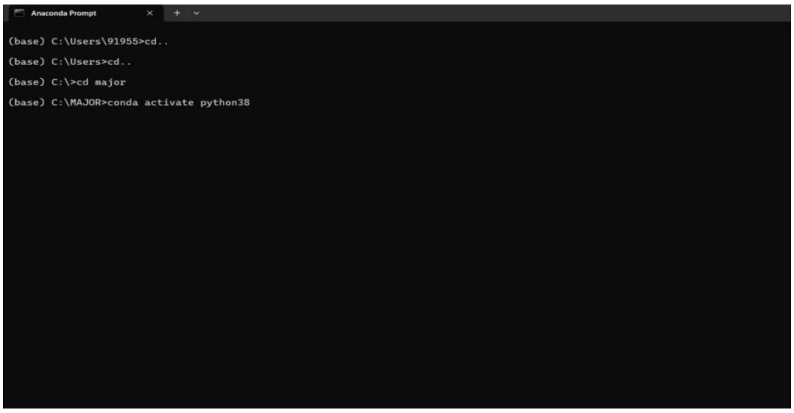

Fig 10. Activating the python in Anaconda

Description : Fig 10, describes the activation of python which is in Anaconda. With the helpof this command we can activate the python which is placed in Anaconda.

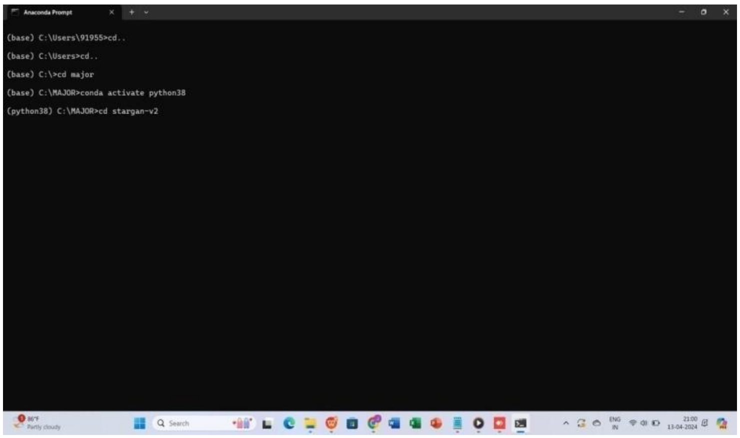

Fig.11. Changing Directory to the StarGAN-v2

Description : Fig 11, describes the navigation of StarGAN-v2 folder. This folder is a sub folderof the major folder. So that we navigating from the major folder to the StarGAN-v2 folder.

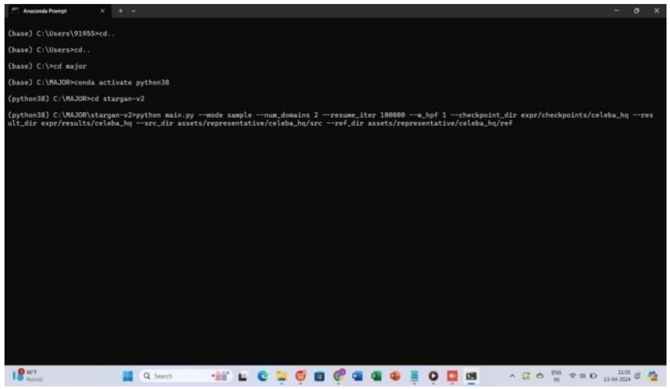

Fig.12. Executing the command

Description : Fig 12, describes the command and this is the major command of our project.By executing this command we can get the results of our project.

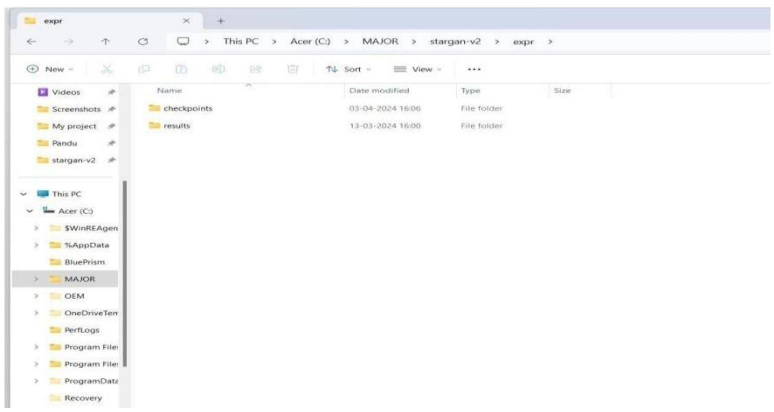

Fig.13. Opening the results folder

Description : Fig 13, describes the folders where our results will be saved. It means that in the‘expr’ folder there are two sub folders whose names are checkpoints and results.

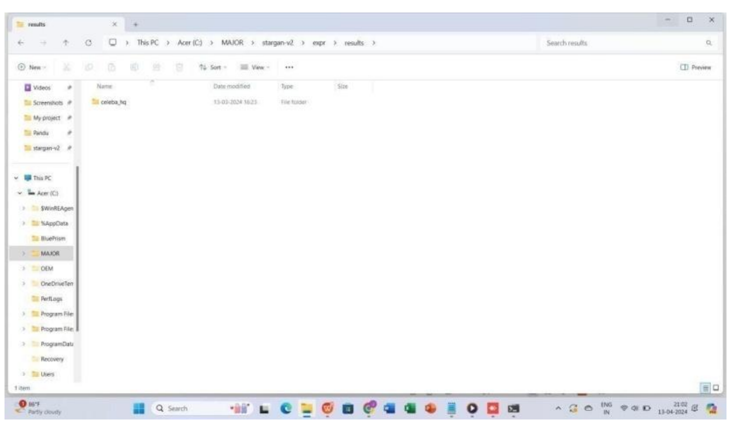

Fig.14. Opening the celeba_hq folder

Description : Fig 14, describes the folder where our results will be saved. It means that in the‘results’ folder, there is one sub folder whose name are celeb_hq, which consists of our generated images.

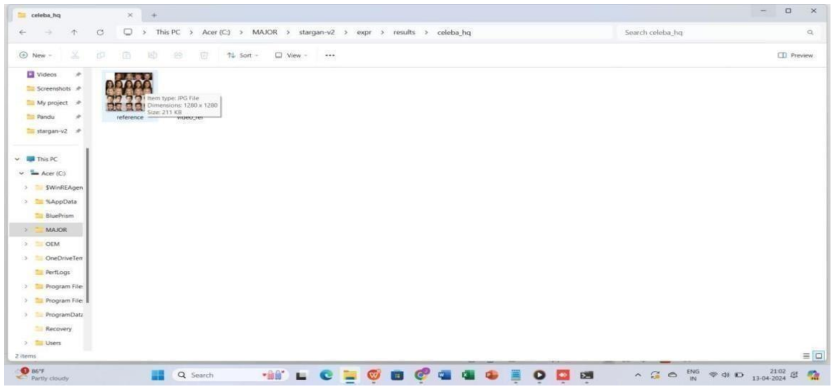

Fig.15. Opening the generated images

Description : Fig 15, contains the images and video which are generated in a matrix format.Itconsists of both source images and reference images.

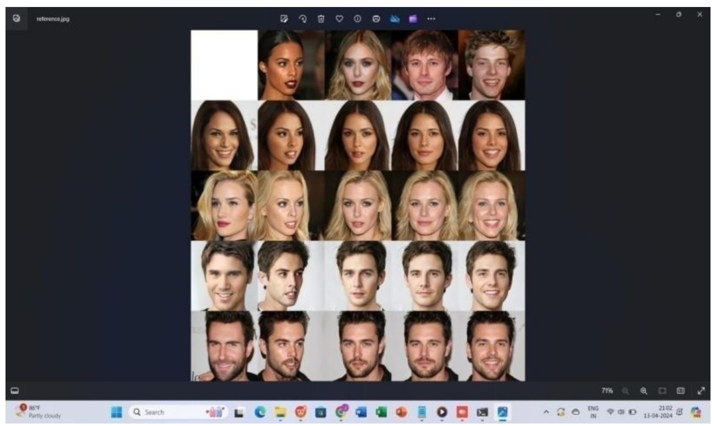

Fig.16 Generated images

Description : Fig 16, describes the images which are generated in a matrix format.It consistsof both source images and reference images.

CONCLUSION

In this project we have created a model which generates the images along with the video. First we have to upload the images of source and reference. Then the images which are uploaded should be cropped and then those will be restored in separate folders. All the source images of the male will be stored in one folder. And also all the female source images will be stored in a separate folder. Same process will be repeated for the reference images. The images will be generated with the help of StarGANs, Python, and the libraries which are present in the python. The generator takes target domain labels, and style codes to produce output images that retain the content of the input while adopting the style of the target domain. The style encoder encodes reference images from the target domain into latent style codes. The mapping generates additionalstyle and given to the discriminator to check whether it is real or fake. The generated images will be stored in a separate folder. Facial expression transfer enables users to modify facial expressions in images, videos, and even in real-time interactions. This capability has the potential to enhance storytelling, character animation, and user- generated content creation in various domains. facial expression transfer can foster a community of users and developers interested in the technology. Providing open-sourceresources can encourage collaboration, knowledge sharing, and innovation in the field. It is useful in the fields such as entertainment, gaming, virtual reality, telecommunication, and human-computer interaction.

FUTURE ENHANCEMENTS

Future enhancements could focus on enabling interactive and real-time facial expression transfer applications. This includes optimizing StarGAN models for low- latency inference, developing intuitive user interfaces. In StarGAN architectures, loss functions, and training methodologies are expected to improve the realism and fidelity of transferred facial expressions. This includes capturing finer details, textures, and nuances in expressions to produce more lifelike results. Users will have more control over the transferred expressions, including the ability to adjust expression intensity, duration, style, and other attributes. Fine- grained control mechanisms will enable users to customize and personalize the transferred expressions according to their preferences.

References

[1] Mohammad Mamunur Rashid, Yingwei Nie &Guiqing Li(2019). High-fidelity facial expression transfer through a part-based approach and conditional GANs, The International Journal of Computer Graphics 3635–3646

[2] Muhammad Ali Farooq, Dan Bigioi, Rishabh Jain, Wang Yao, Mariam Yiwere, Peter Corcoran, "Synthetic Speaking Children – Why We Need Them and How to Make Them", 2023 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), pp.36-41, 2023.

[3] F.Y. Dai,J.Chi,M.G.Ren,Q.D.Zang(2022). Face Image Synthesis Driven by Geometric Feature and Attribute Label[J].Computer Science, 2022, 49(10): 214-223.

[4] Z. Yu, J. Chi, Y. Ye and F. Dai (2021).Detailed features-preserving 3D facial expressiontransfer Journal of Computer-Aided Design & Computer Graphics , Volume 33, Issue 2: 186 - 198 (2021) |

[5] X.X. Wang, F.F.L, Q.Chen(2020). Facial expression generation method based on enhanced conditional generative adversarial networks

[6] Wang, Xinrui et al(2020). Learning to Cartoonize Using White-box Cartoon Representations.

[7] Zhe sun, Hehao Zhang, Ziatong Bai, Mingyang Liu, Zhengping Hu (2021). Discriminately deep fusion approach with improved conditional GANs for facial expressionrecognition.

[8] Hamza Bouzid, Lahoucine Ballihi (2021). Facial expression video generation based on spatio-temporal convolutional GAN: FEV GAN

[9] Yang Fan, Xingguo jiang, Shuxing lan, Zianghai lan (2023). Facial expression transfer based on conditional generative adversarial n