Documentation: CS F437 - GAI - Assignment 2:

Names of students (with ID no):

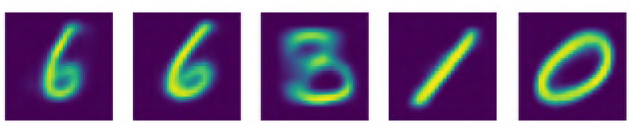
1. Vansh Agrawal - 2021A7PS2998H
2. Aditya Kumar Sharma - 2021A7PS3112H
3. Abhishek Joshi - 2021A7PS2727H

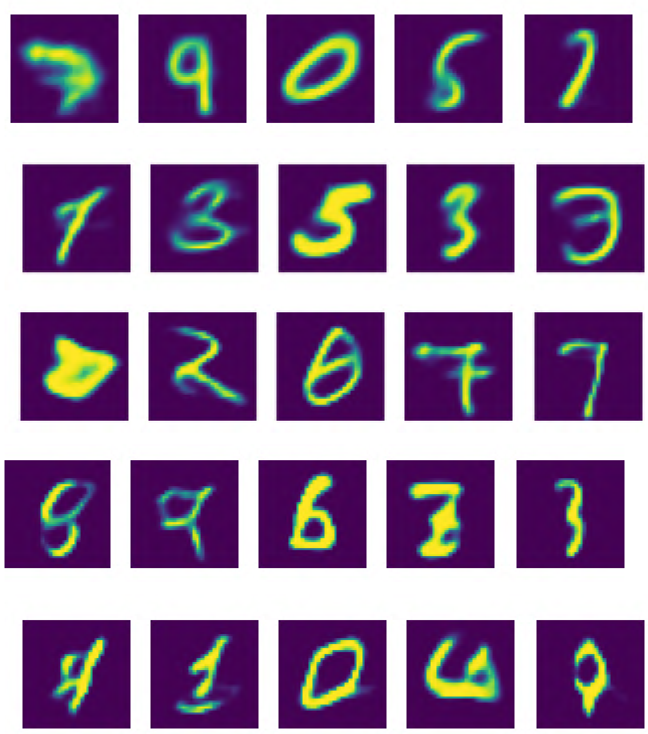
PART A

Question 1: Variational Autoencoder

Variational Autoencoders (VAEs) learn latent representations by combining neural networks with probabilistic models. The encoder maps input data to a probability distribution in latent space. The decoder reconstructs data samples based on sampled latent variables. Training involves optimizing parameters to maximize the likelihood of observed data and minimizing the Kullback-Leibler divergence.

Here are the 5 images from the dataset which we randomly samples from the dataset, to show our results variational autoencoder:





Question 2: Generative Adversarial Network

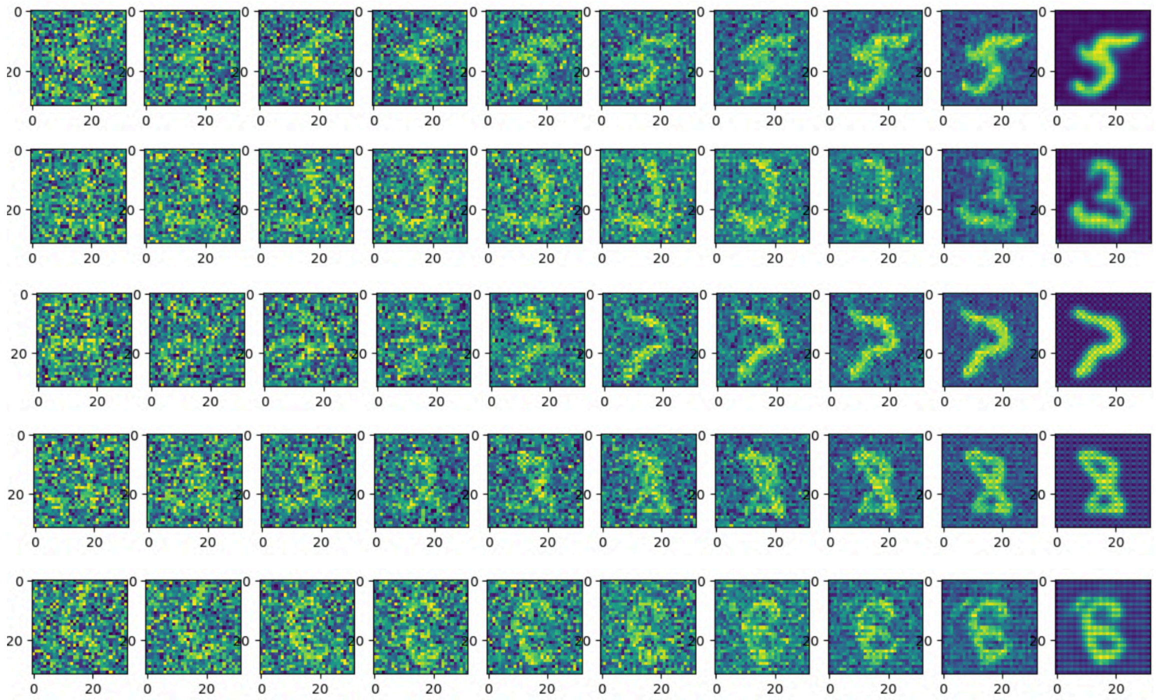
Generative Adversarial Networks (GANs) are a deep learning architecture composed of two networks, a generator and a discriminator, competing in a game-like setting. The generator creates synthetic data resembling real samples, while the discriminator distinguishes between real and fake data. Through adversarial training, GANs produce high-quality, realistic data, making them pivotal in tasks like image generation and data augmentation.

Here are the 5 images from the dataset which we randomly samples from the dataset, to show our results for GAN:

Question 3: Diffusion Model

Diffusion models are powerful probabilistic generative models that capture complex data distributions by iteratively diffusing noise. Unlike traditional generative models, diffusion models learn data distributions by sequentially transforming a noise source until it resembles the target data. During training, the model gradually refines the noise to generate realistic samples, leveraging reversible transformations and learned dynamics. Diffusion models have demonstrated exceptional performance in image generation tasks, offering high-fidelity results and enabling fine-grained control over generated samples.

Here are the 5 images from the dataset which we randomly samples from the dataset, to show our results for diffusion models:



Inferences from part A

Variational Autoencoders (VAEs)

- **Performance**: The VAE model demonstrated strong performance in capturing latent representations of the input data. It effectively learned to encode and decode data samples, producing reconstructions that closely resembled the original inputs.
- **Latent Space Exploration**: By traversing the latent space, we observed smooth transitions between different data representations, indicating meaningful latent representations learned by the model.
- **Training Stability**: The VAE model exhibited stable training behavior throughout the training process, with consistent convergence and reconstruction quality.

Generative Adversarial Networks (GANs)

- **Sample Quality**: The GAN model generated high-quality synthetic samples that closely resembled the real data distribution. Visual inspection revealed realistic images with fine details and diverse variations.
- **Mode Collapse**: We encountered challenges related to mode collapse, where the generator failed to capture the entire data distribution, resulting in limited diversity in generated samples.
- **Training Dynamics**: Despite mode collapse issues, the GAN model demonstrated dynamic training behavior, with fluctuations in generator and discriminator losses indicating ongoing learning.

Diffusion Models

- **Data Fidelity**: Diffusion models achieved impressive results in capturing complex data distributions, producing high-fidelity samples with fine-grained details.
- **Training Efficiency**: The training process of diffusion models was computationally intensive due to the iterative diffusion steps. However, it converged effectively, leading to significant improvements in sample quality over time.
- **Fine-Grained Control**: Diffusion models offered fine-grained control over generated samples, allowing manipulation of generated images by modifying diffusion steps or noise levels.

Overall Comparison

- **Performance Evaluation**: Across all three models, we evaluated performance based on sample quality, training stability, and the ability to capture meaningful latent representations.
- **Trade-offs**: Each model exhibited unique strengths and weaknesses. VAEs excelled in latent space exploration and stability, GANs in generating realistic samples with diversity, and diffusion models in capturing fine-grained details and offering control.
- **Future Directions**: Further investigation is needed to explore hybrid approaches or enhancements to address limitations and leverage the strengths of each model for improved generative modeling tasks.

PART B

Question 1: Deep Convolutional GAN (DCGAN)

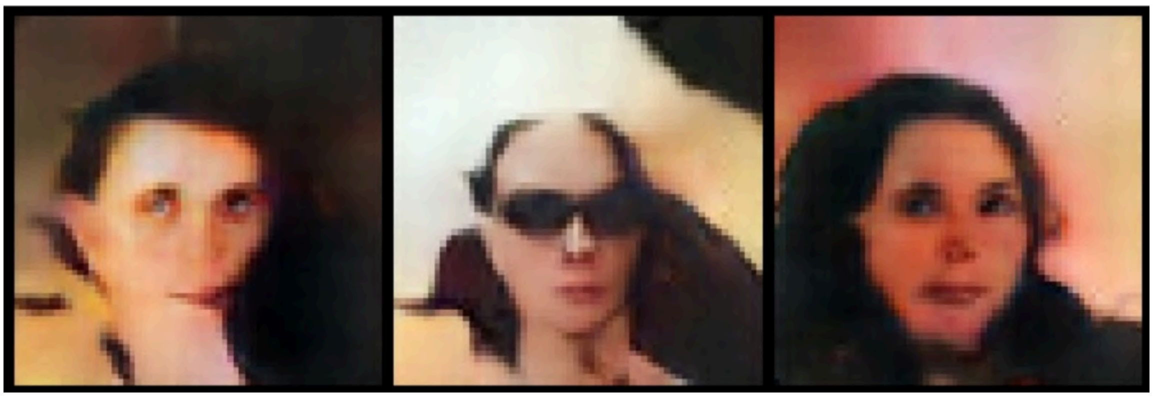
We adopted the architecture proposed in the paper "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks" by Alec Radford, Luke Metz, and Soumith Chintala. Our implementation of Deep Convolutional Generative Adversarial Networks (DCGANs) adhered closely to the prescribed architectural constraints. By following this framework, we aimed to leverage the strengths of convolutional networks (CNNs) for unsupervised learning tasks in computer vision. Through training on diverse image datasets, our DCGAN model demonstrated the ability to learn hierarchical representations spanning from object parts to scenes within both the generator and discriminator networks. Additionally, we validated the effectiveness of the learned features for various downstream tasks, echoing the findings of the referenced paper and showcasing the potential of DCGANs as versatile tools for unsupervised representation learning.

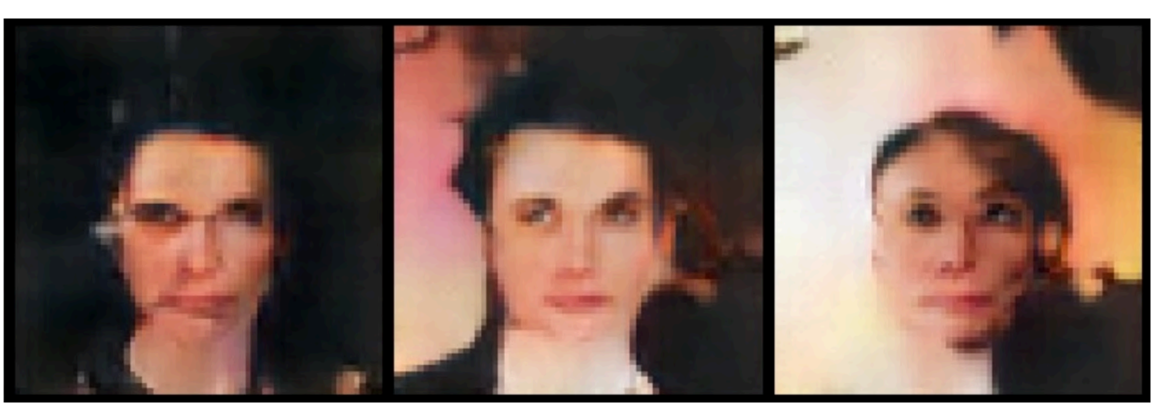
Results for DCGAN

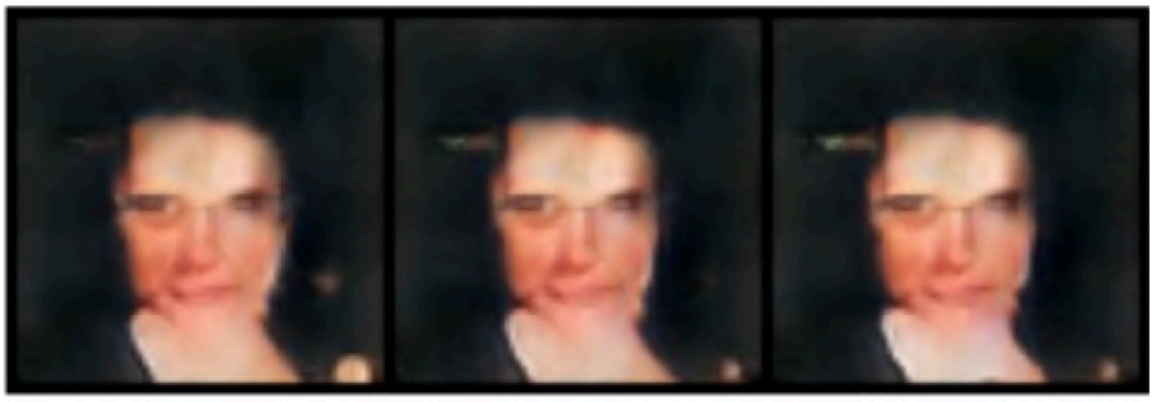
Vector Arithmetics

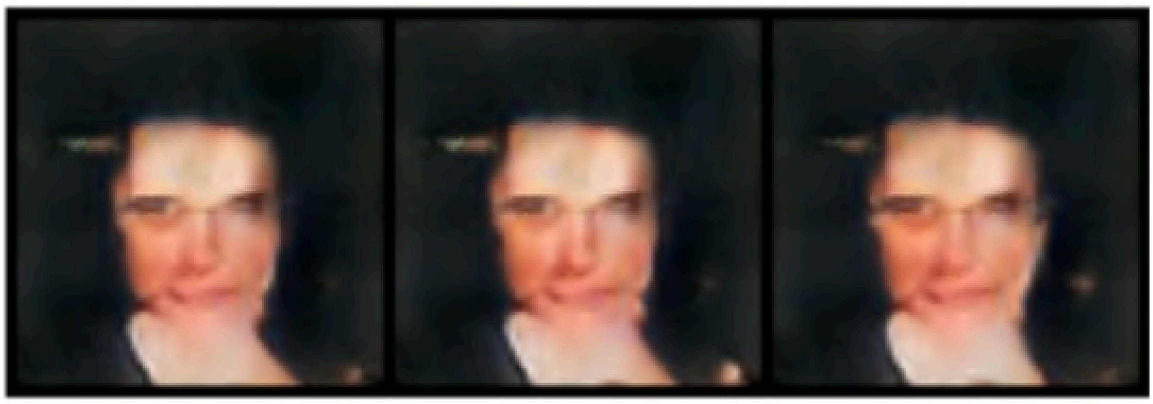
For each of the vector arithmetic operations:

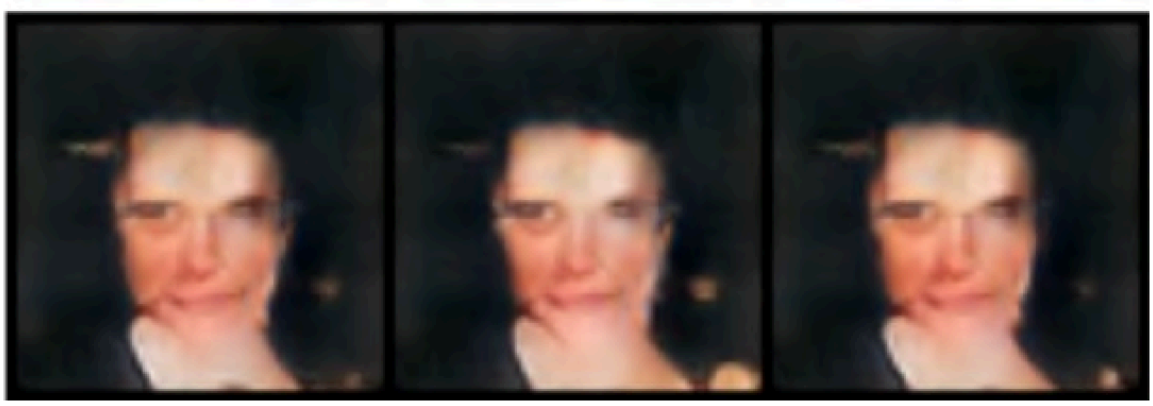
- **Operation 1: Men without glasses + People with glasses - People without glasses**:
  - For all 5 outputs, we display:
    - The average of the three images in each category.
    - The 3x3 grid output.

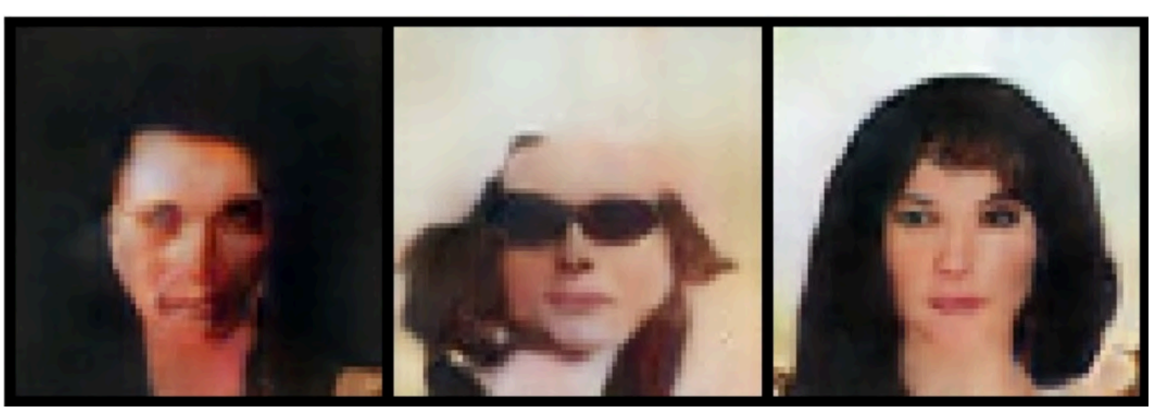


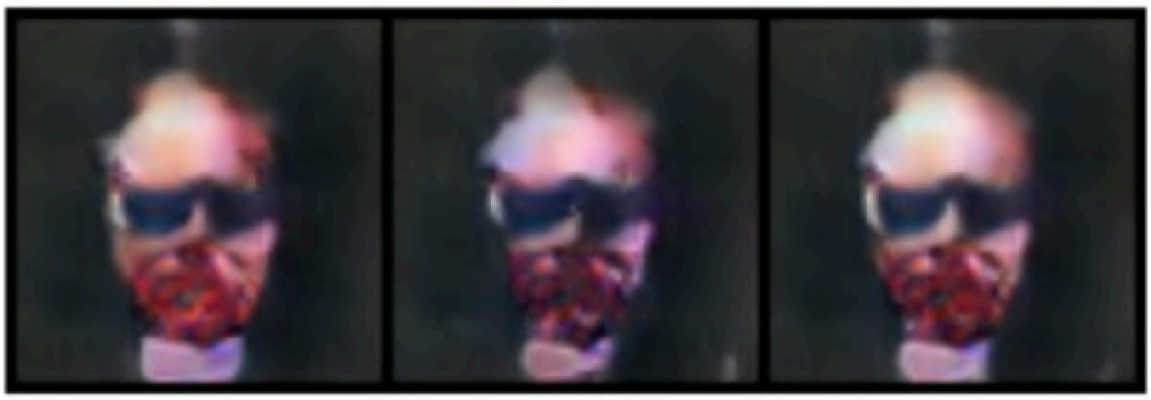


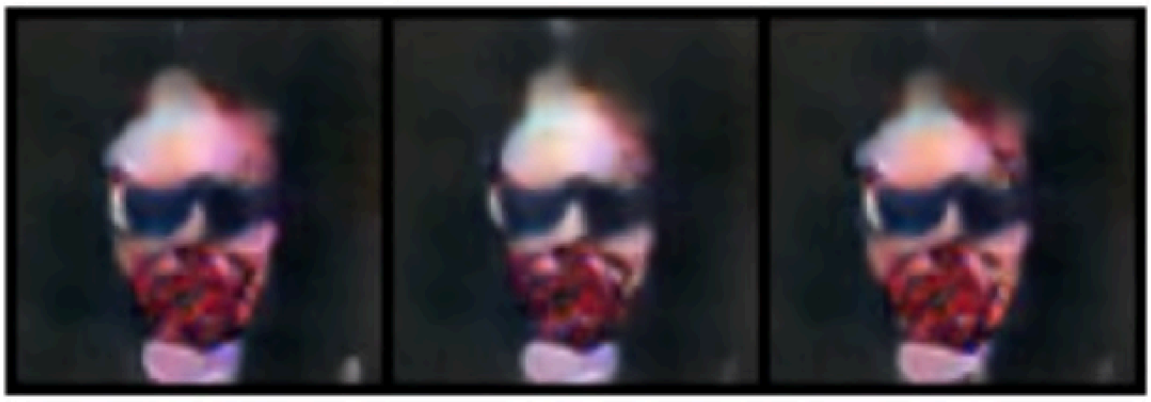


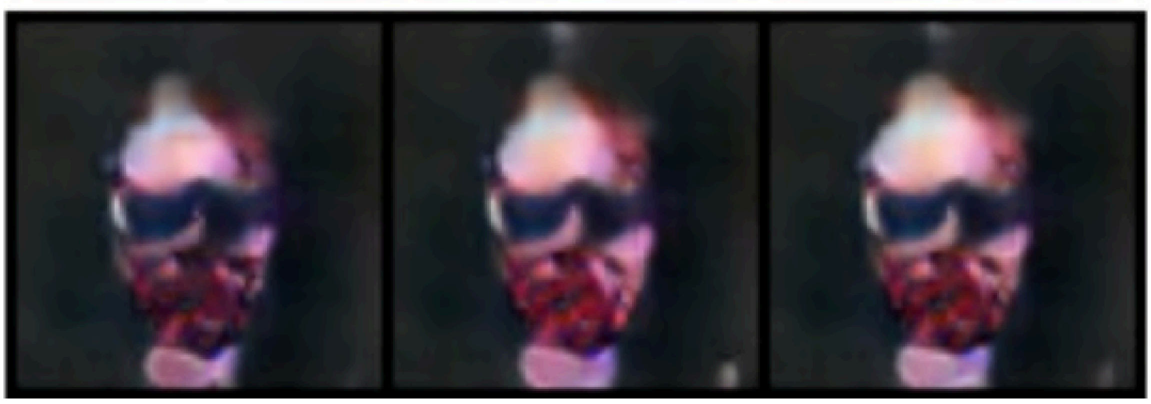


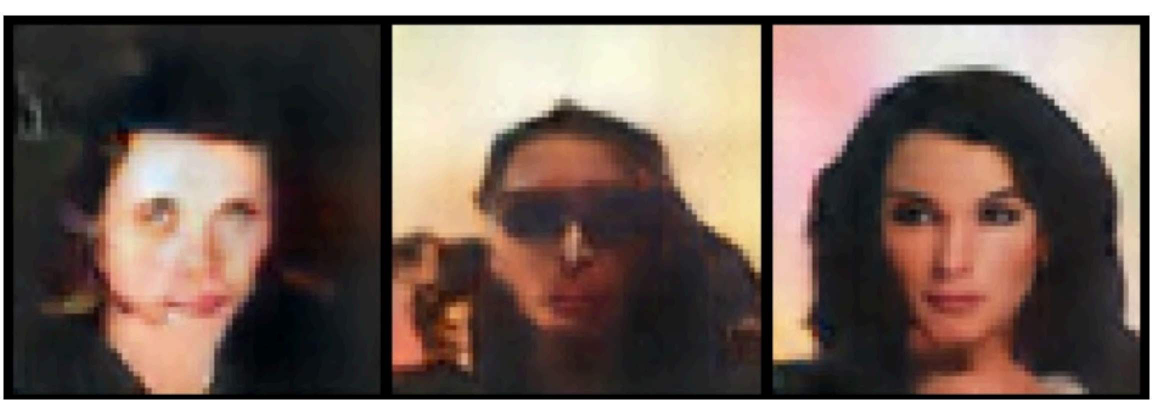


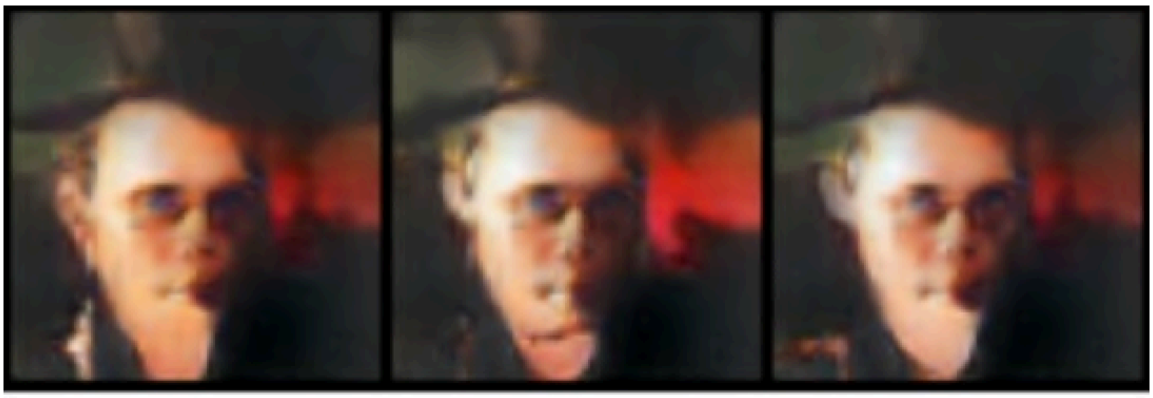


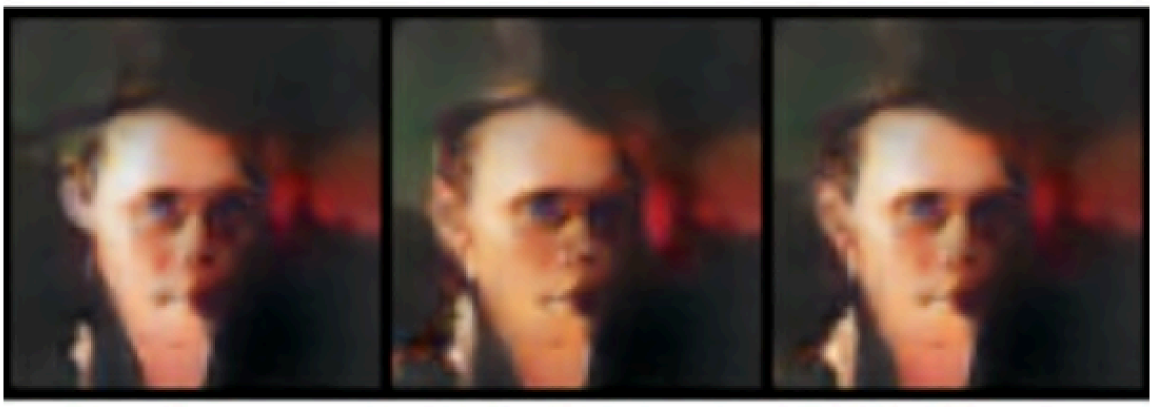


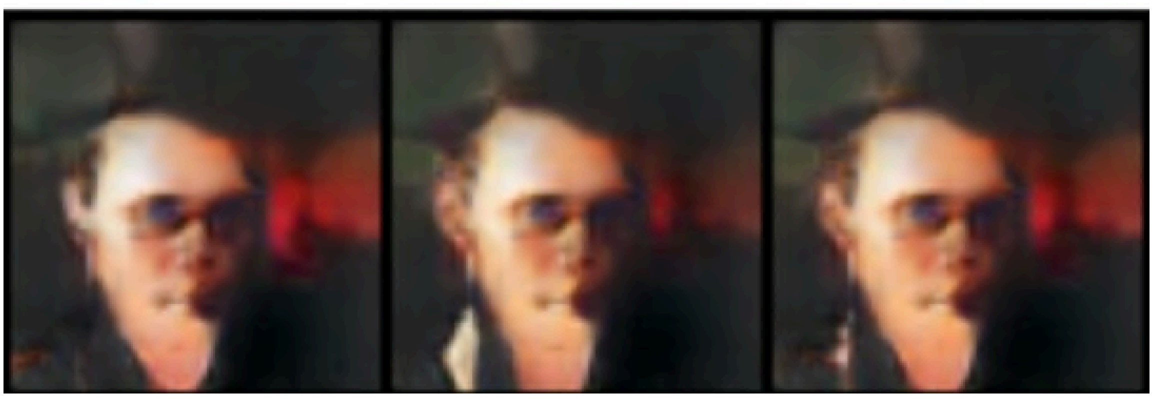


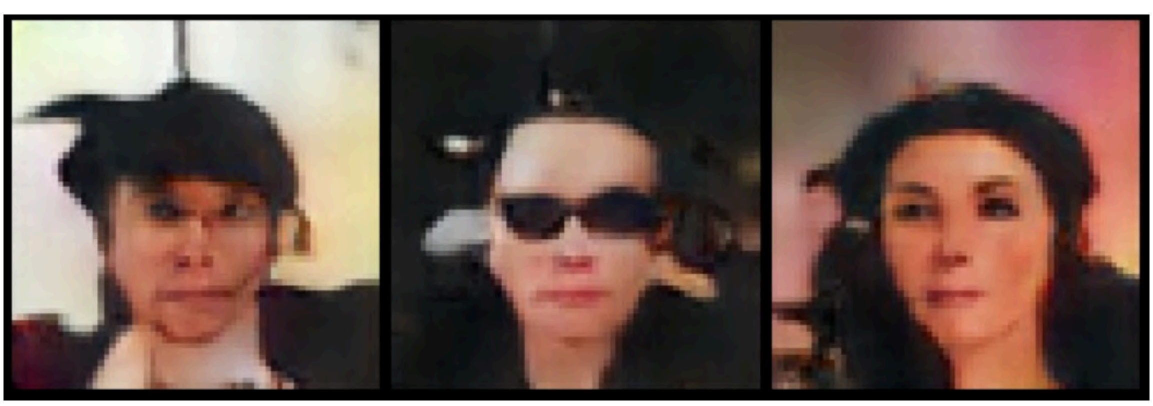












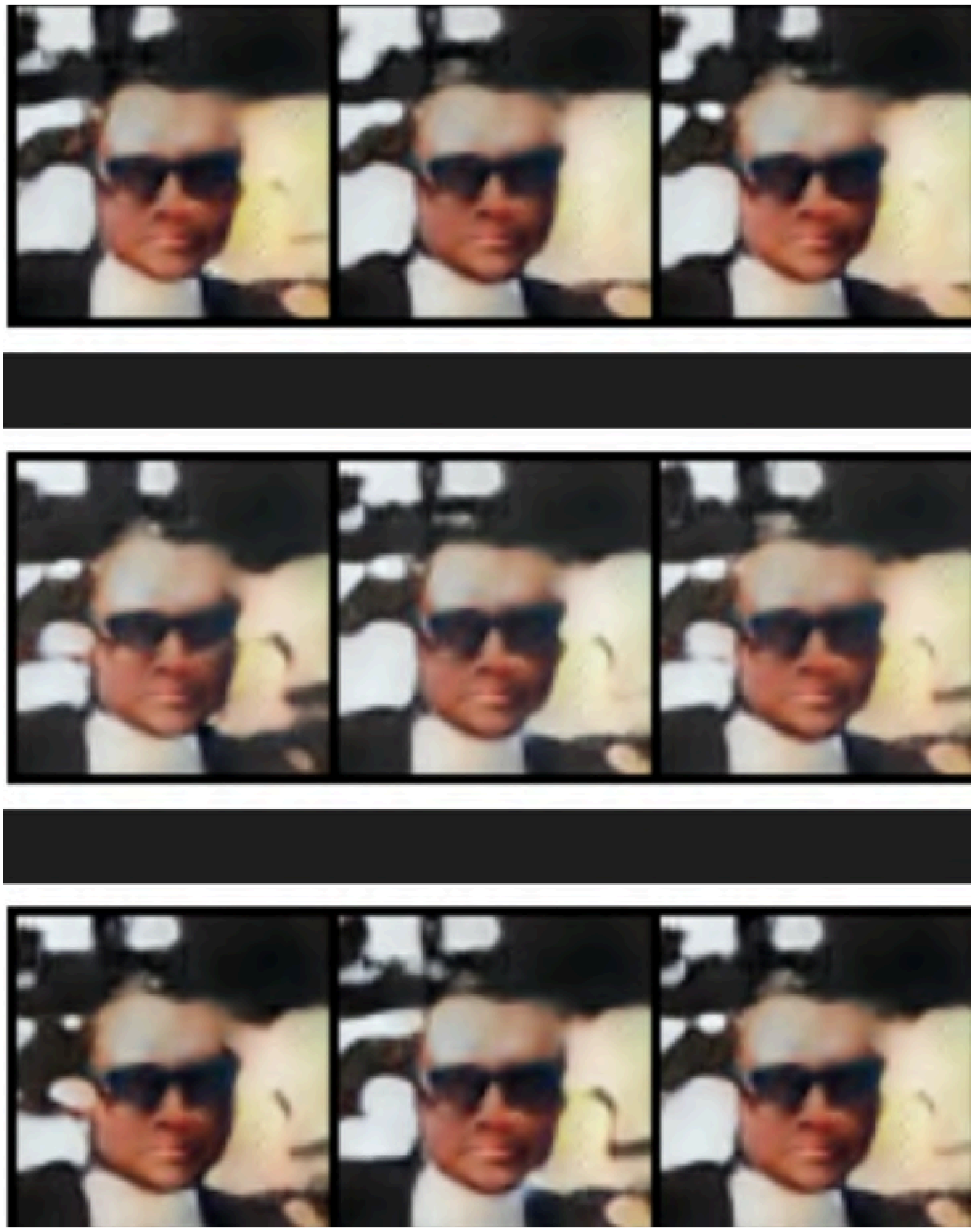
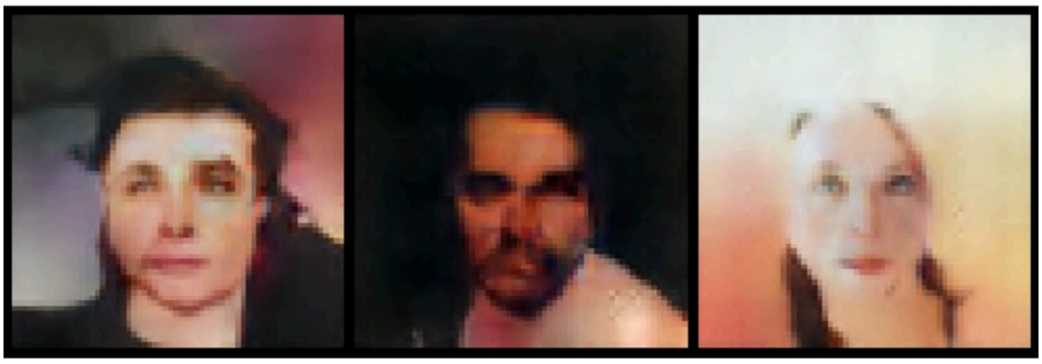
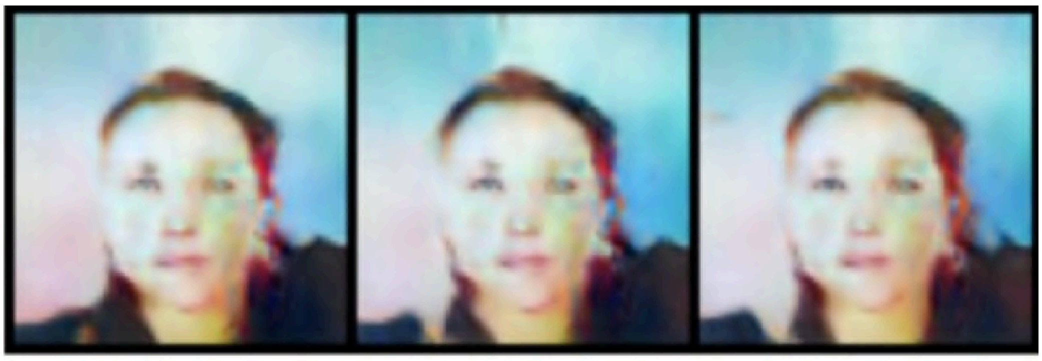


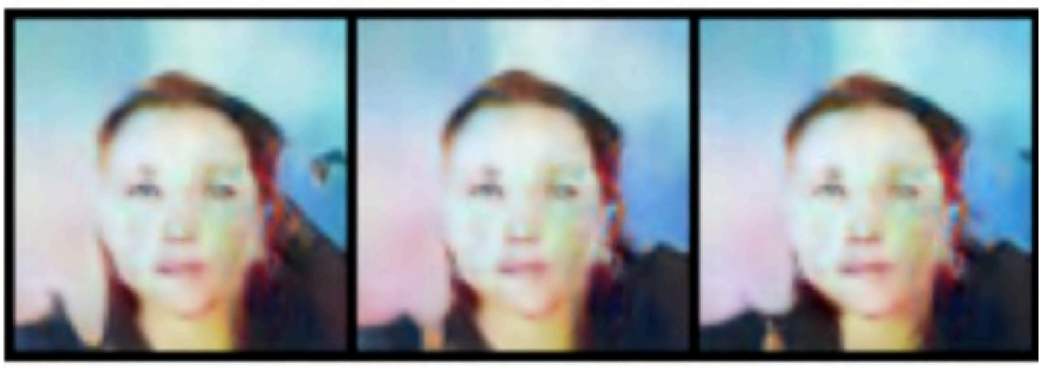
Image Arithmetic

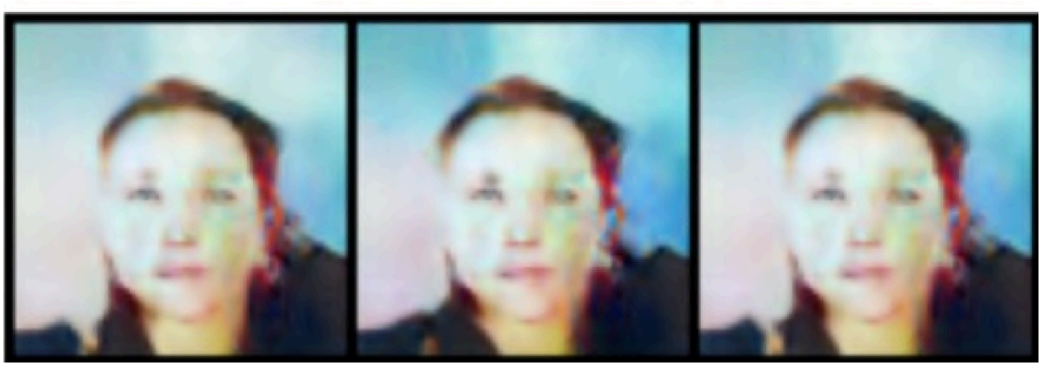
- **Operation 2: Men with glasses - Men without glasses + Women without glasses**:
    - For all 5 outputs, we display:
        - The average of the three images in each category.

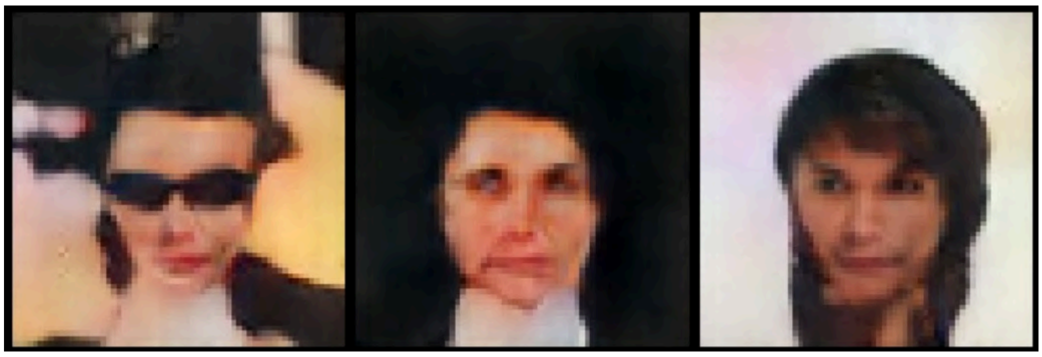
- The 3x3 grid output.

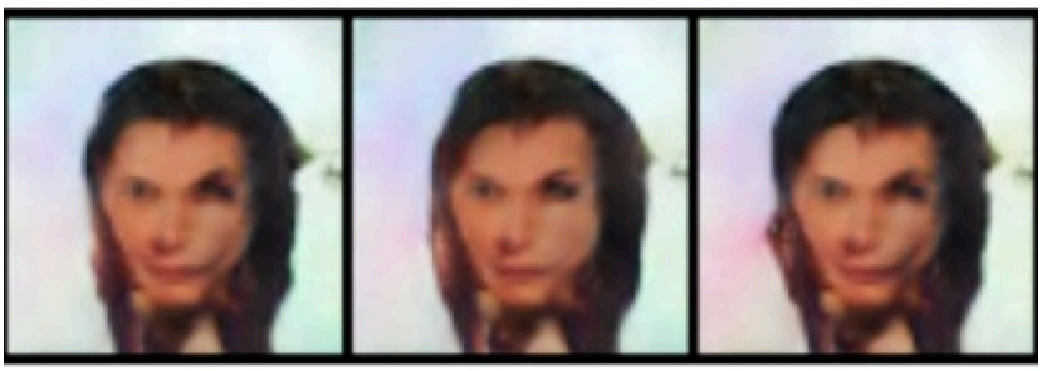


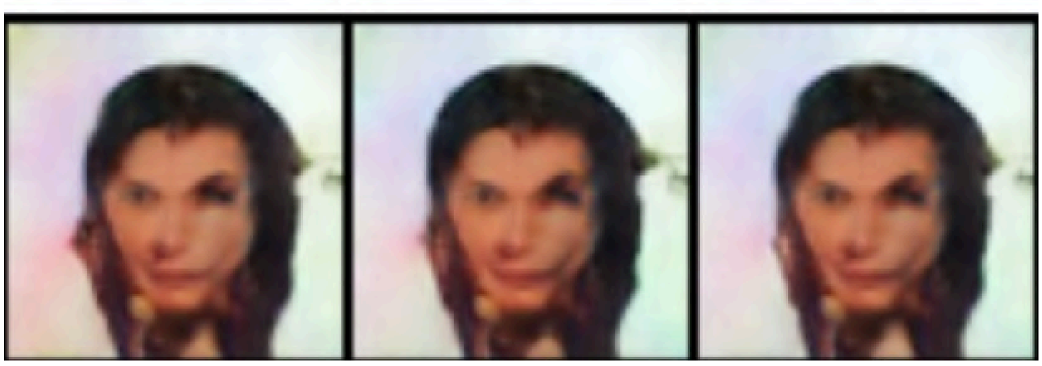


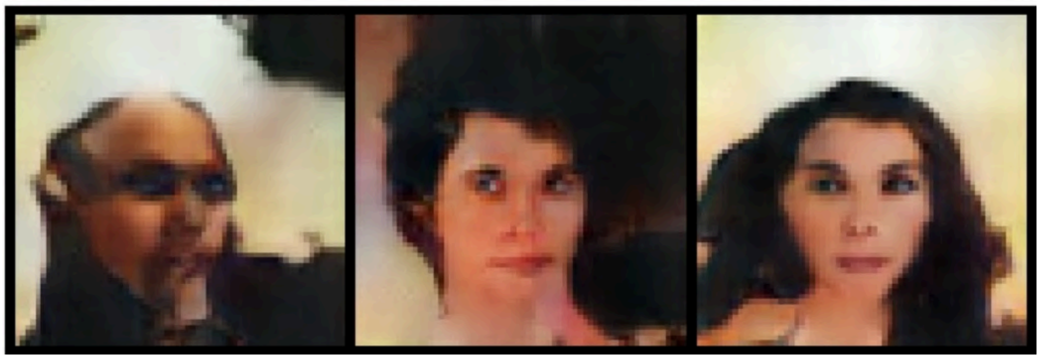


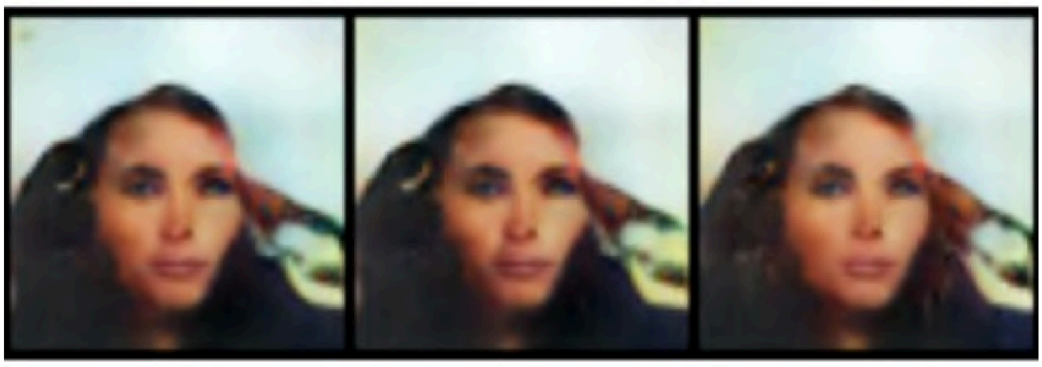


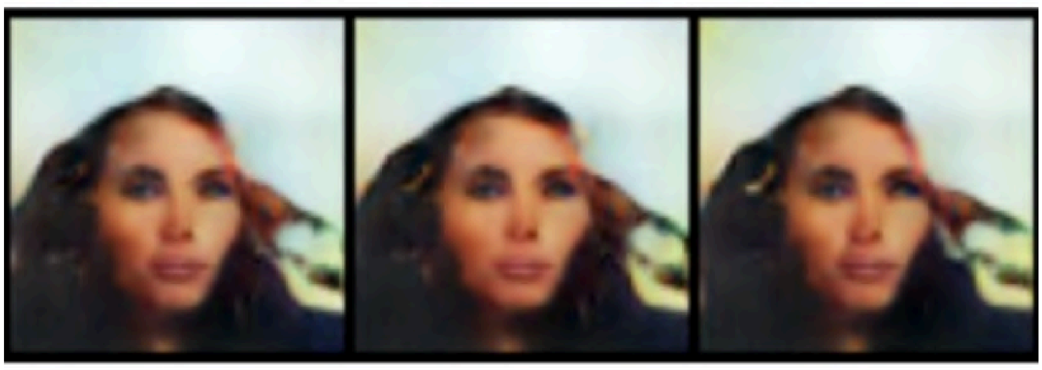


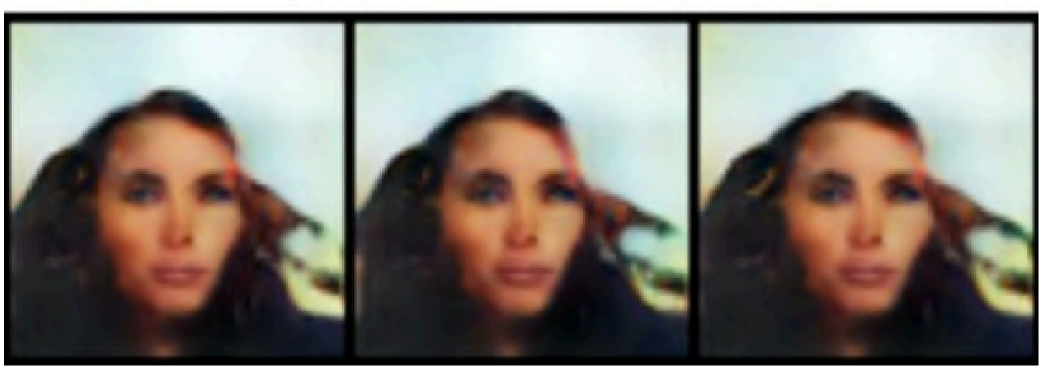


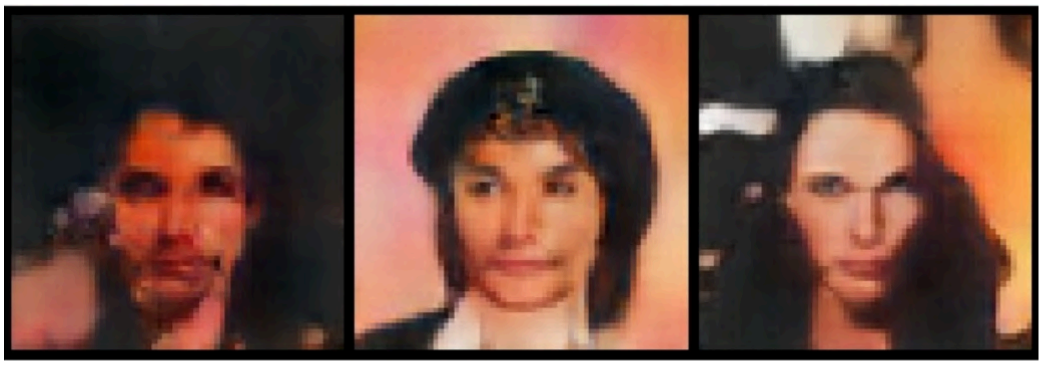


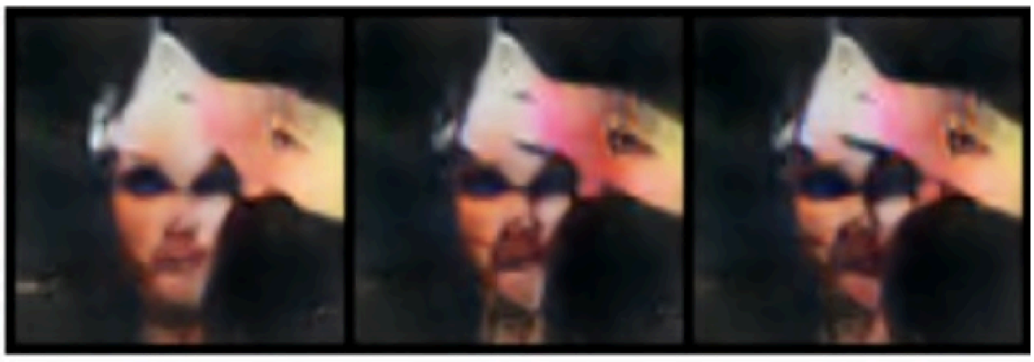


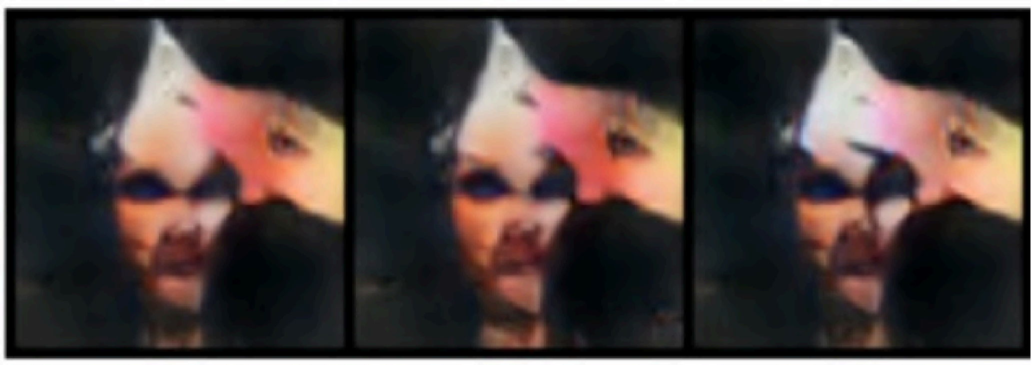


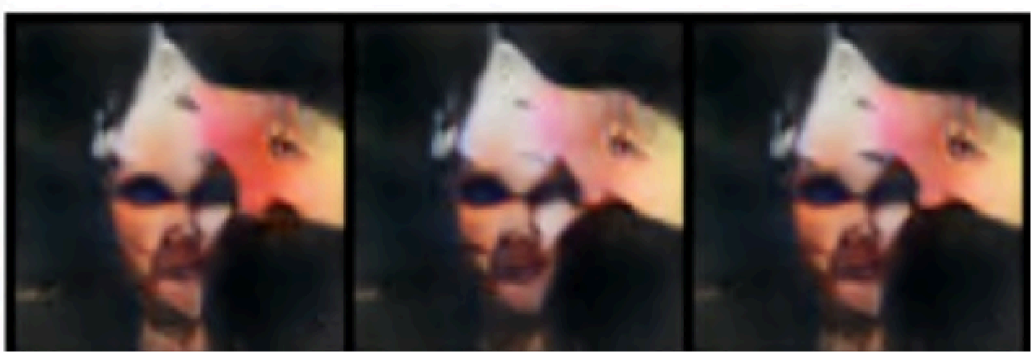


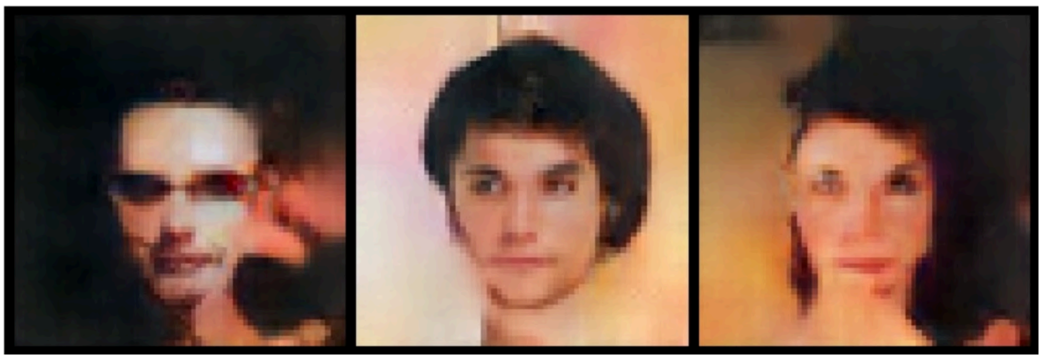


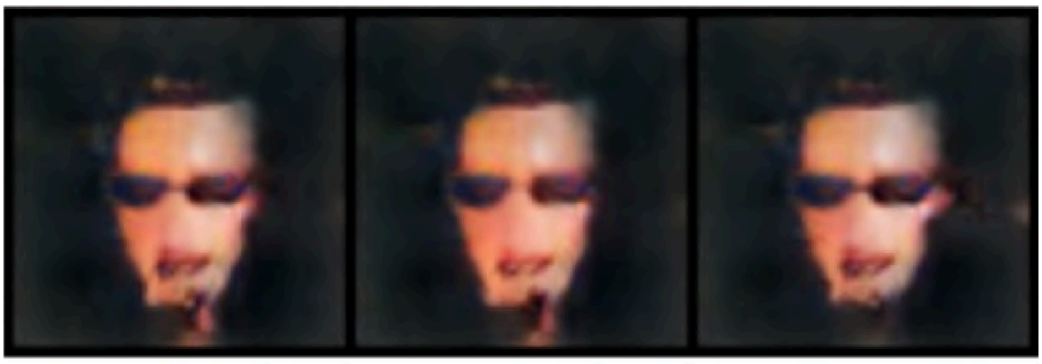


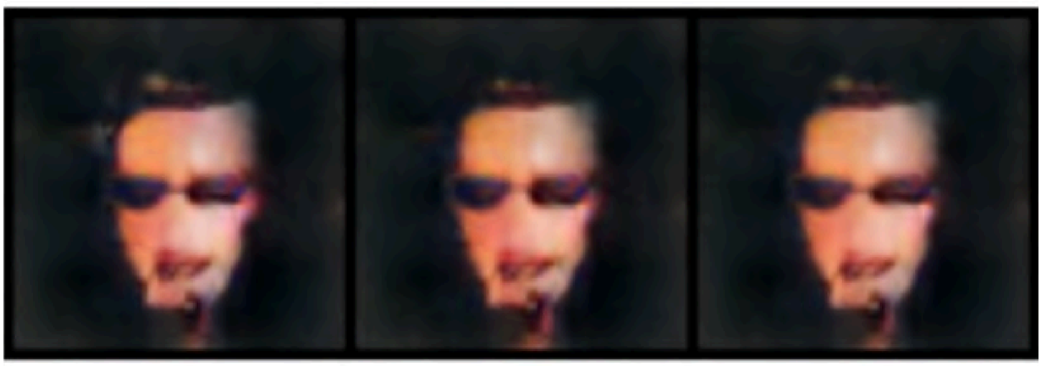












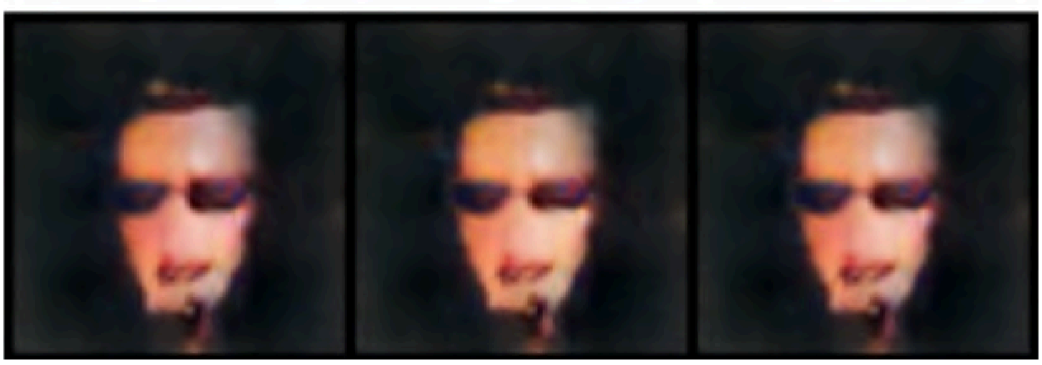
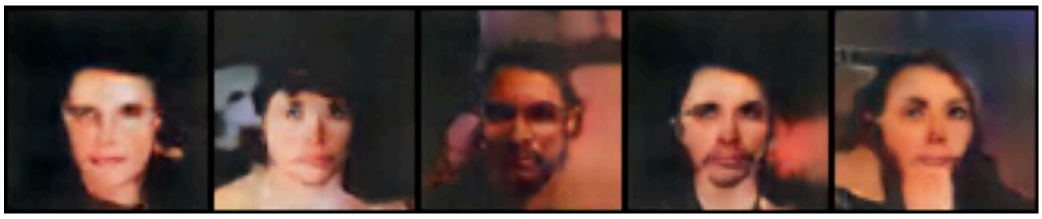
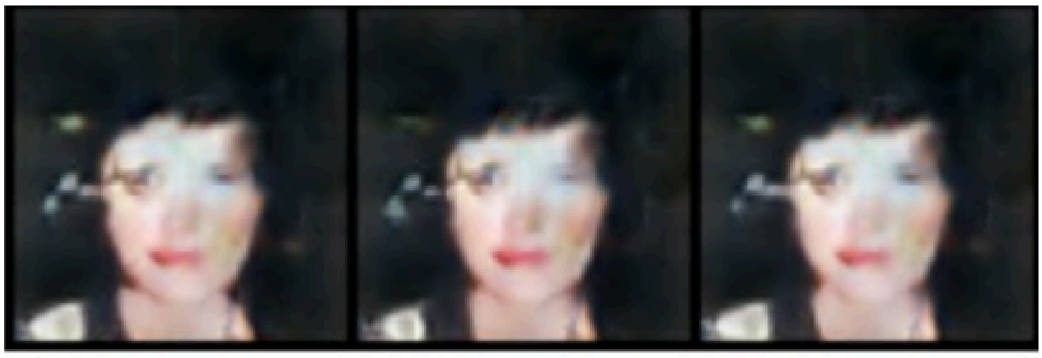
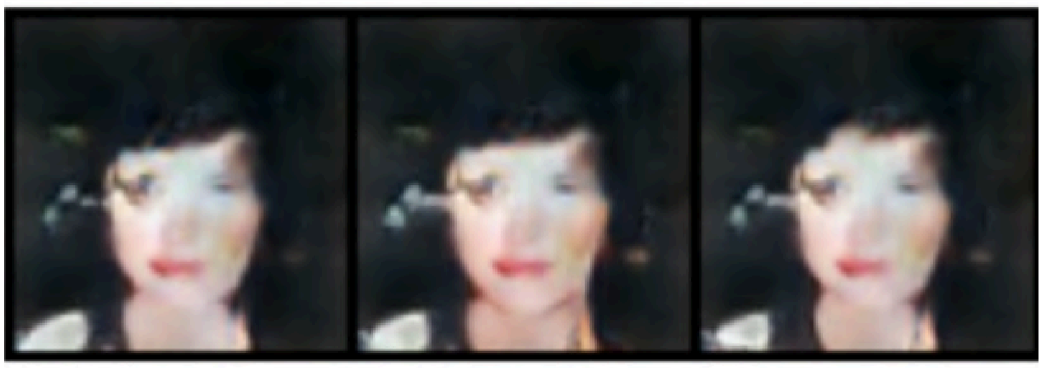


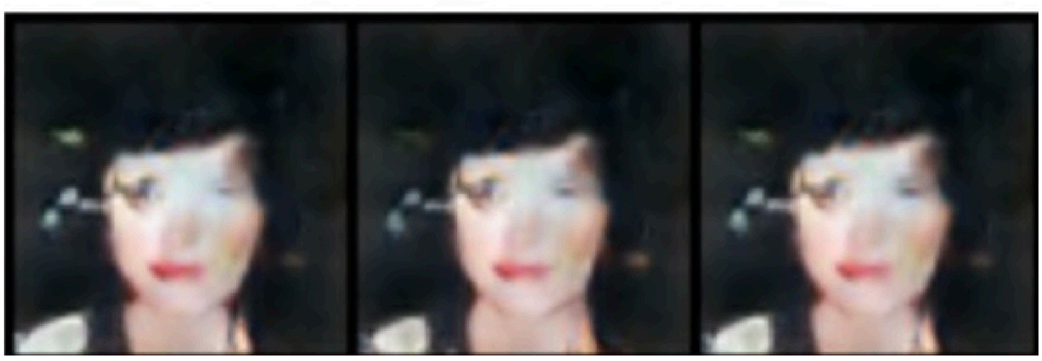
Image Arithmetic

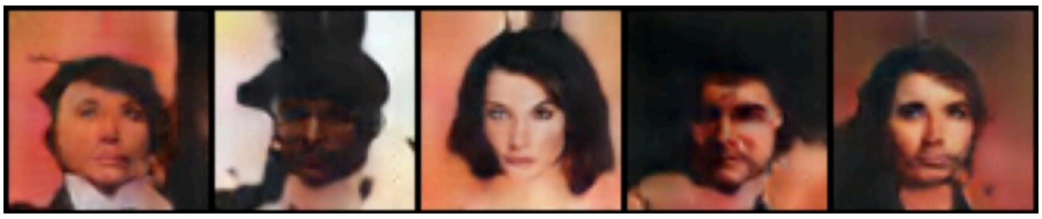
- **Operation 3: Smiling Men + People with Hat - People with Hat + People with Mustache - People without Mustache**:
  - For all 5 outputs, we display:
    - The average of the three images in each category.
    - The 3x3 grid output.

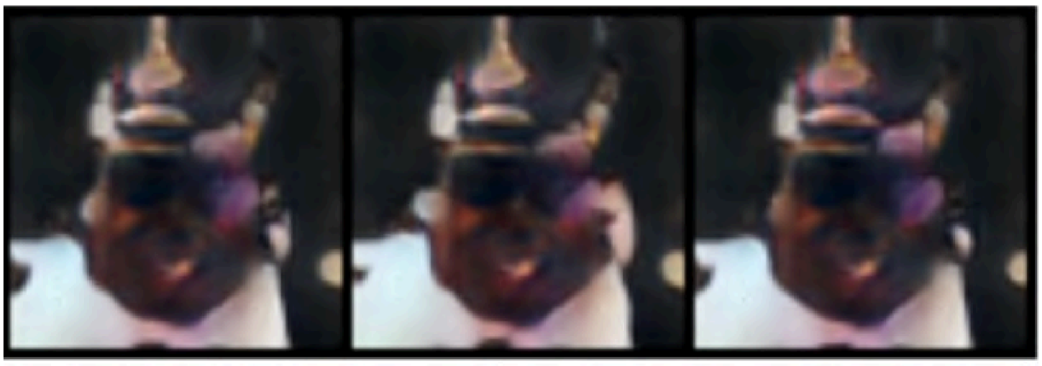


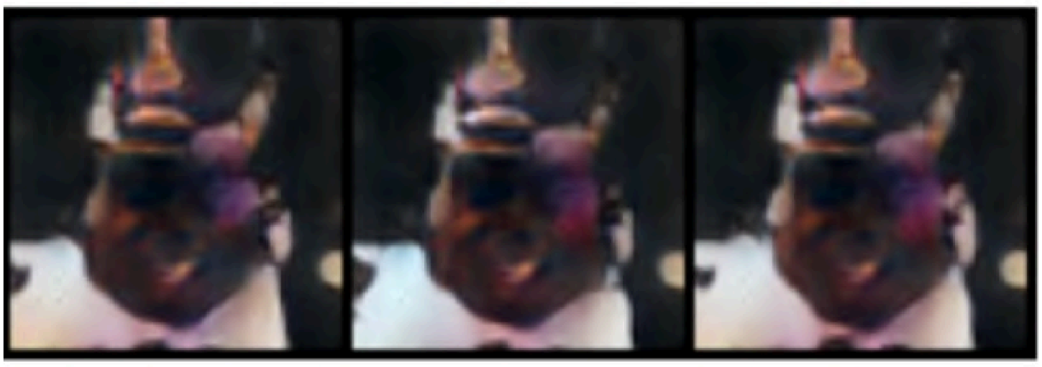


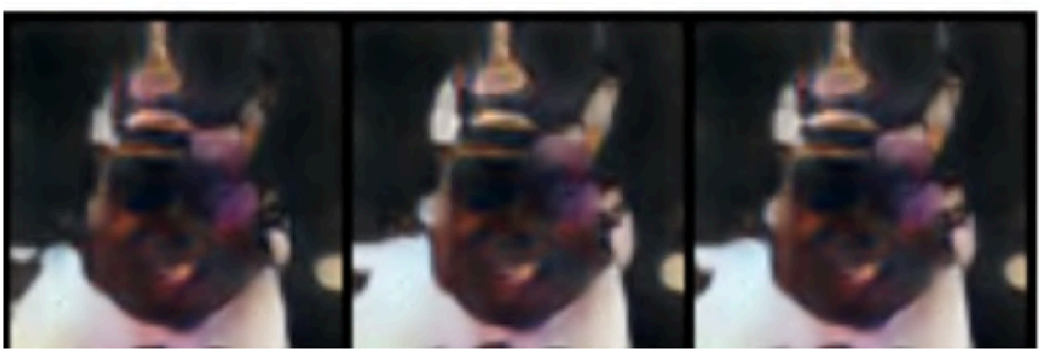


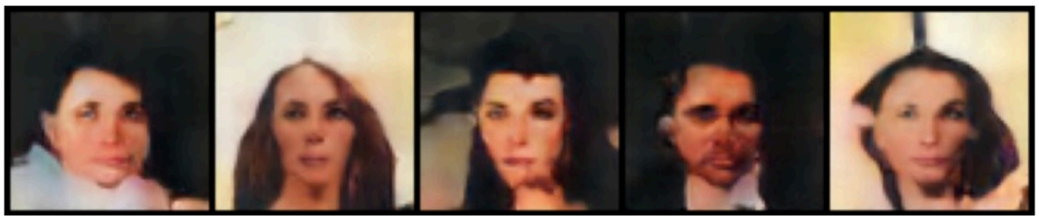


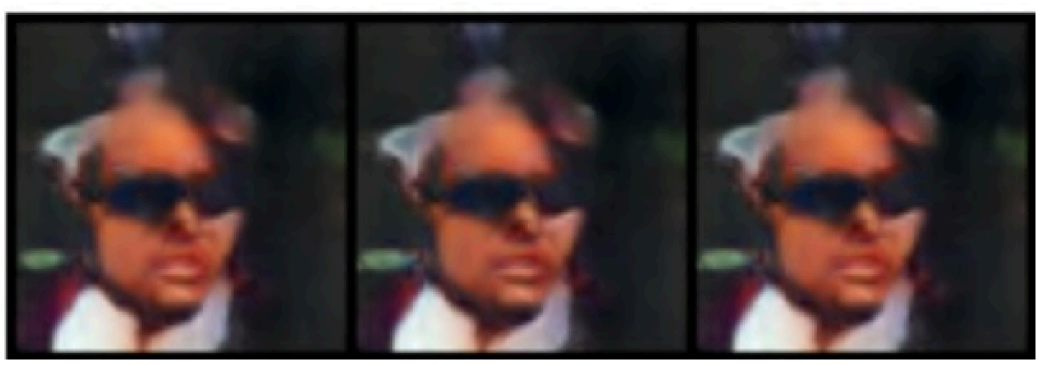


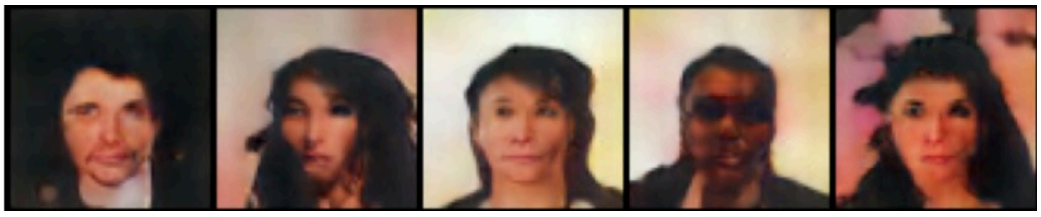


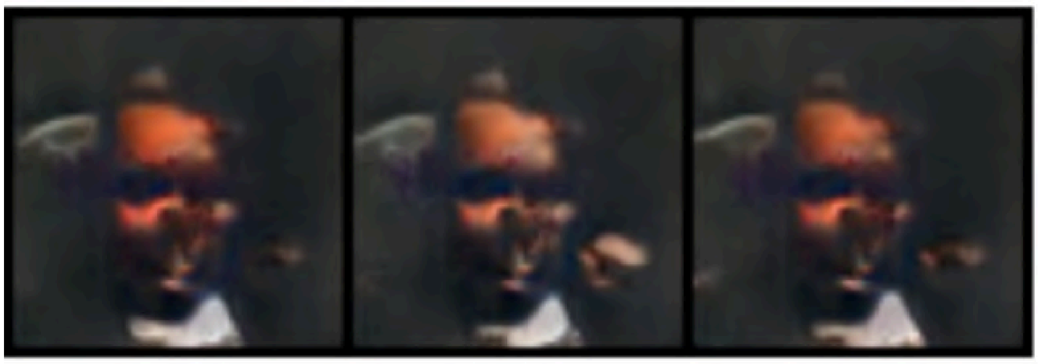


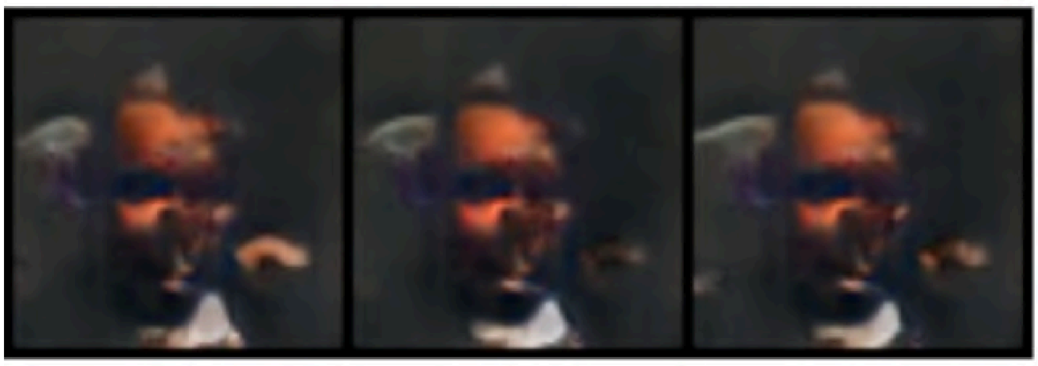


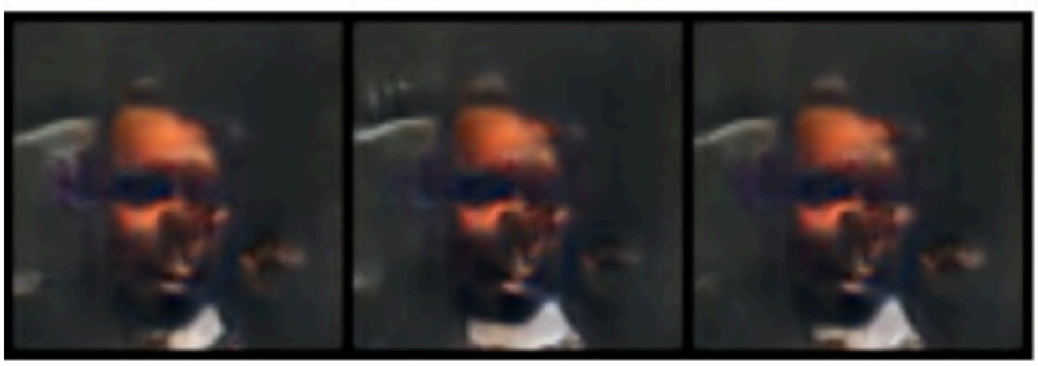


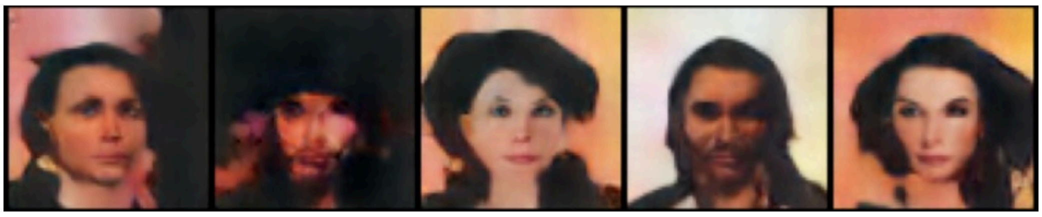


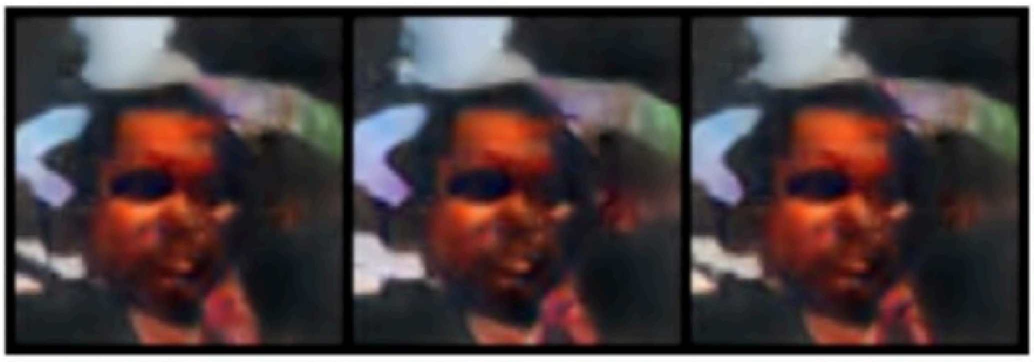


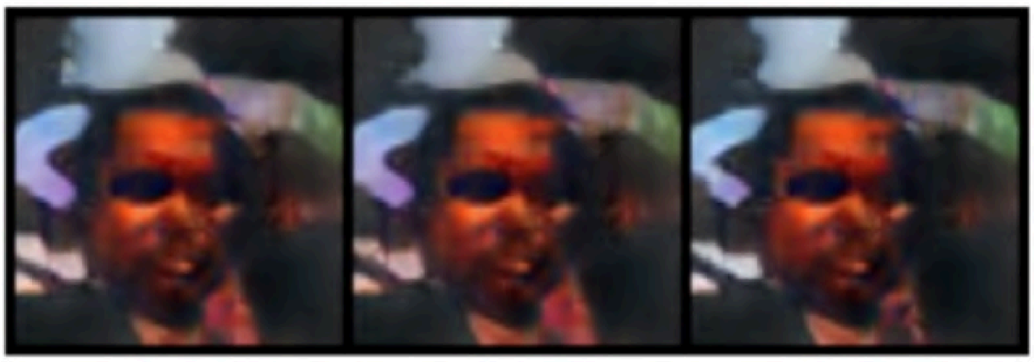












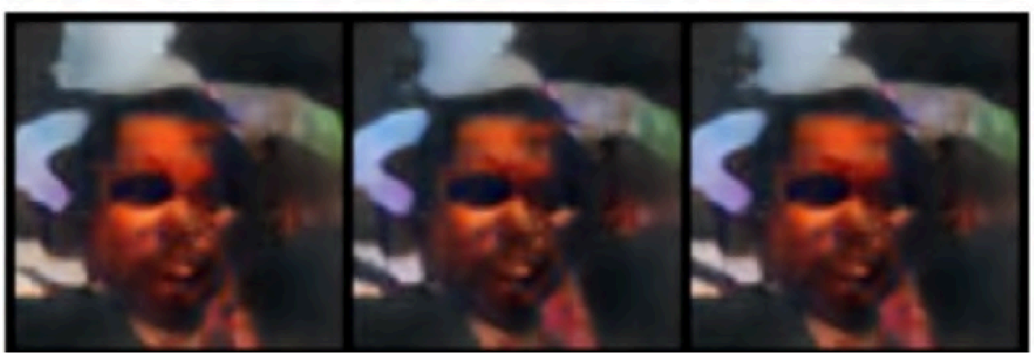
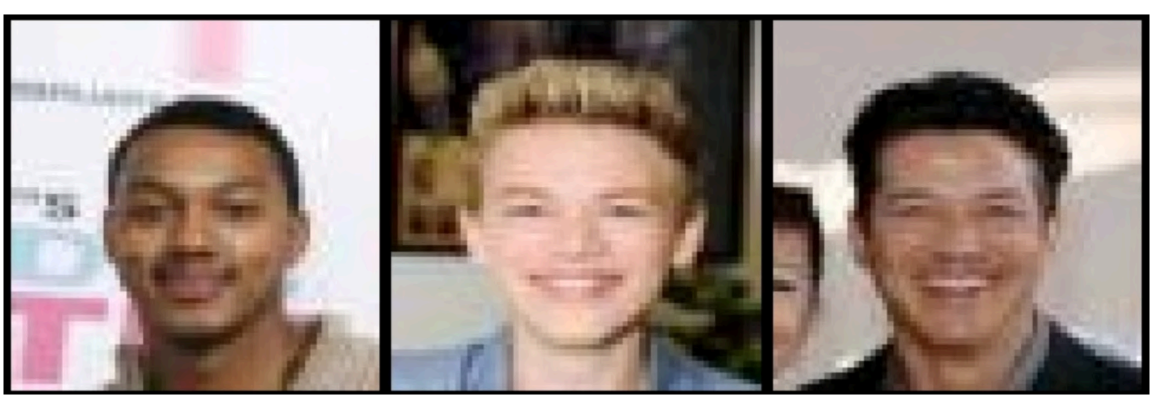
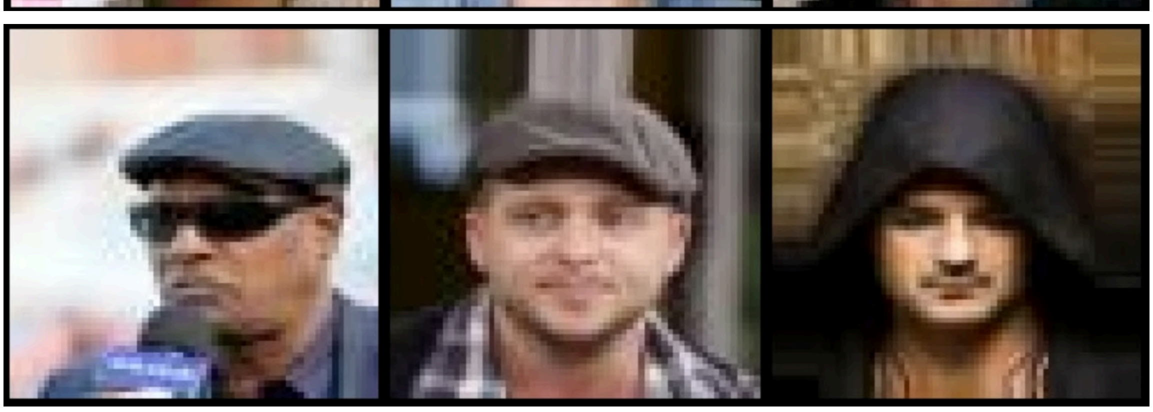
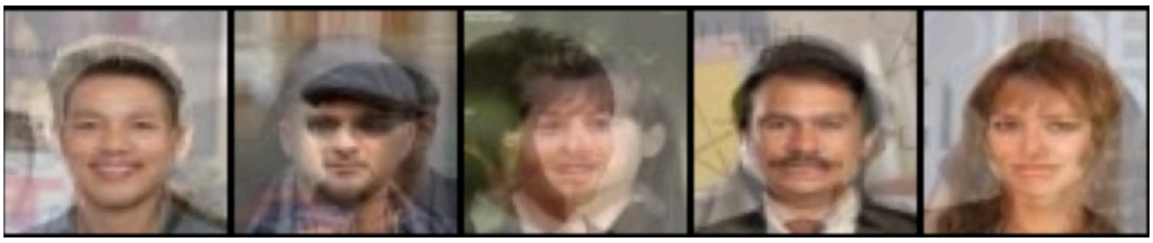


Image-Level Arithmetics





Average:



Question 2: CycleGAN

CycleGAN is a type of generative adversarial network (GAN) designed for unsupervised image-to-image translation tasks. Unlike traditional GANs, CycleGAN introduces cycle consistency loss, which encourages the generated images to preserve key characteristics from the original domain after translation. This enables the model to learn mappings between two domains without requiring paired training data. CycleGAN has been widely used for various applications, such as style transfer, artistic rendering, and domain adaptation, due to its ability to generate high-quality images with consistent domain transformations.

Results for CycleGAN

Men without glasses to men with glasses and vice versa (5 images)

1. 

2. 

3. 

4. 

5. 

Men with glasses to women with glasses and vice versa (5 images)

1. 

2. 

3. 

4. 

5. 

Inferences from Part B

DCGAN vs CycleGAN

Image Quality:

- DCGAN:
  - Demonstrates the ability to generate high-quality images with realistic details.
- CycleGAN:
  - Produces images with consistent transformations between different domains, but may exhibit artifacts or distortions.

Training Complexity:

- DCGAN:
  - Requires paired training data for supervised learning tasks.
  - Training process involves optimizing adversarial and reconstruction losses.
- CycleGAN:
  - Operates on unpaired datasets, simplifying data collection and labeling.
  - Training involves optimizing adversarial losses and cycle consistency losses.

Application Flexibility:

- DCGAN:
  - Primarily used for generating images based on learned latent representations.
  - Suitable for tasks such as image synthesis and data augmentation.
- CycleGAN:
  - Specifically designed for image-to-image translation tasks without paired data.
  - Enables domain adaptation, style transfer, and artistic rendering.

Performance Trade-offs:

- DCGAN:
  - Excels in generating visually appealing images with fine details.
  - May struggle with preserving semantic content and consistency across transformations.
- CycleGAN:
  - Focuses on maintaining consistency in domain translations but may sacrifice image fidelity.
  - Offers versatility in handling diverse image translation tasks but may require tuning for optimal results.

Conclusion:

- Both DCGAN and CycleGAN offer distinct advantages and trade-offs in terms of image quality, training complexity, and application flexibility.
- The choice between DCGAN and CycleGAN depends on the specific requirements of the task, dataset characteristics, and desired output quality.