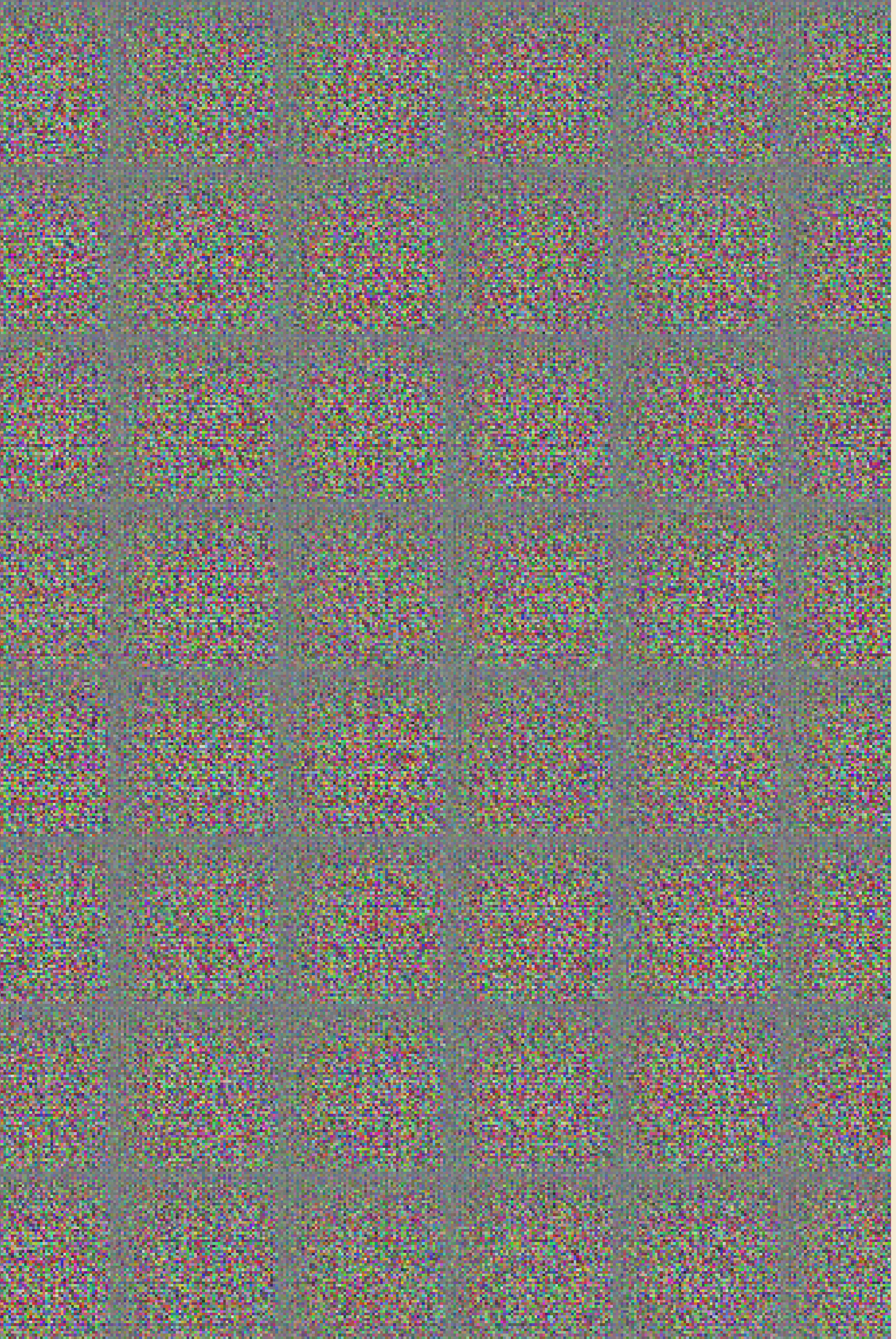


Generation and Completion of Human Face Images

Applied Machine Learning
Final Project

Ananya Sinha	MDS202307
Ankita	MDS202309
Divyanshi Kumari	MDS202322
Rohit Roy	MDS202340

Aim

This project explores image completion using deep learning, focusing on **generating and completing human face images**. Image completion is challenging as it requires inferring missing parts from neighboring pixels and learning content from existing image areas.

We used a **deep-convolutional generative adversarial network (DCGAN)** to generate natural human faces and designed an image completer to fill missing pixels contextually and perceptually.

Motivation

- Restore missing or damaged regions in face images (e.g., occlusions, noise).
- Useful for applications like photo editing, facial recognition, and image restoration.
- Human faces require high semantic consistency — small errors are easily noticeable.
- Masking parts (e.g., mouth) trains the model to infer realistic and context-aware details.
- Enables generation of visually plausible and structurally coherent face completions.

Dataset Preparation and Selection

Source

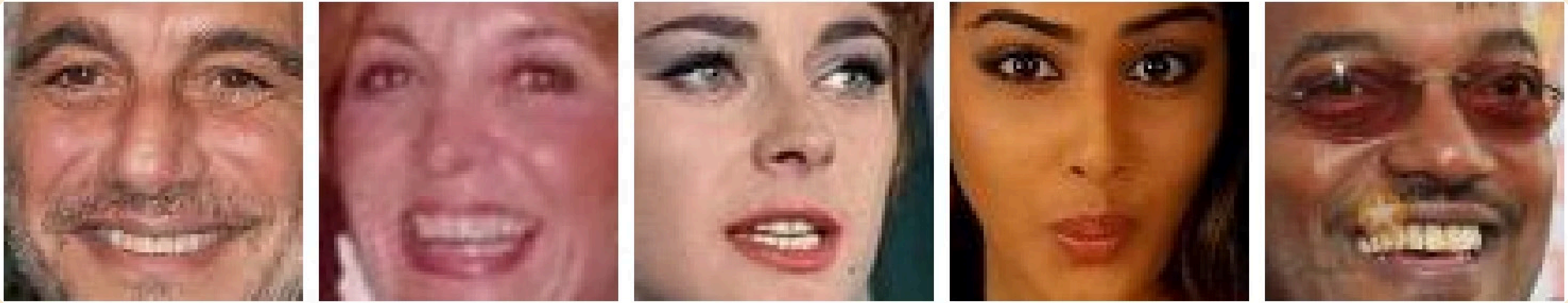
Used the **CelebA** dataset with 202,599 RGB face images of 10,177 identities, cropped to 80×80 pixels centered on faces.

Straight-facing Images

Developed a metric to shortlist 91,895 straight-facing facing images, utilising symmetry around the nose, vertical alignment of eyes and eye angle

Preprocessing

Pixel values rescaled from [0,255] to [-1,1] before training; masked 30×30 pixel area near the nose/mouth for completion tasks



DCGAN Architecture for Image Generation

Generator

Input: 100-length noise vector; passes through dense, upsampling, and convolutional layers with tanh activation to output 80×80×3 images.

Layers: Dense → Upsampling + Convs

Discriminator

Processes 80×80×3 images through convolutional and max-pooling layers to output probability of real image using sigmoid activation.

Conv + MaxPooling → Dense → Sigmoid

Model Summary

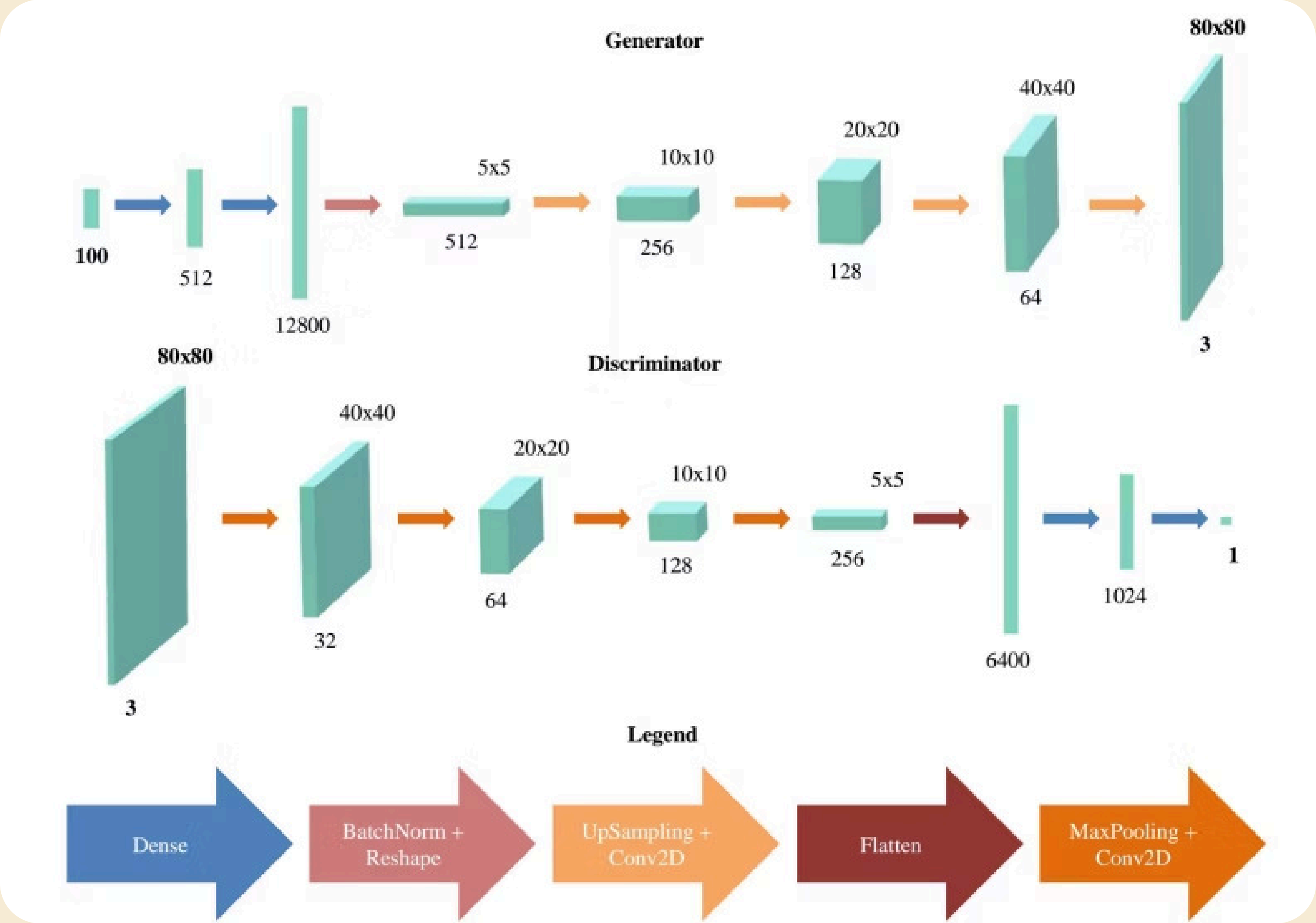
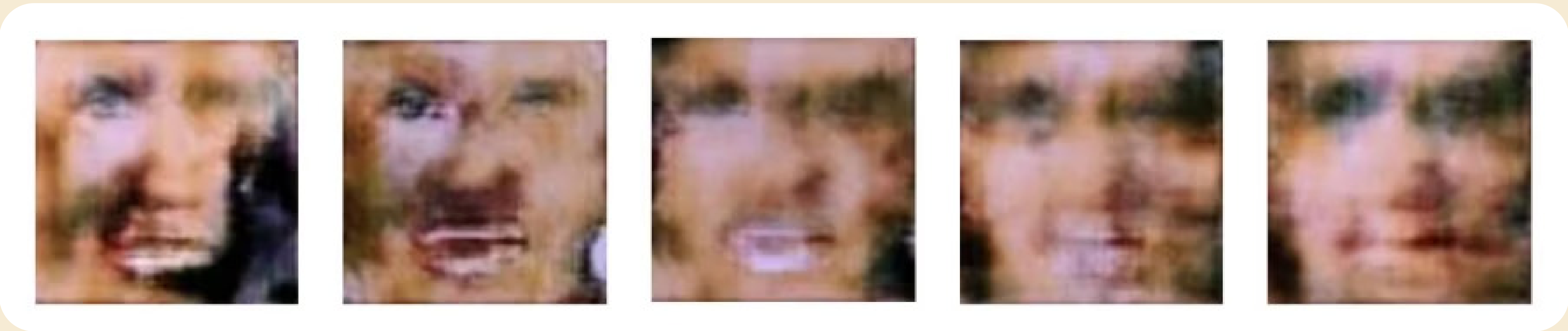


Image Completion Methodology

1. Apply a 30×30 mask over the mouth region to create a partially visible image and a binary mask.
2. Define a loss combining contextual loss (visible region match) and perceptual loss (realism via discriminator).
3. Optimize a random latent vector z using gradient descent to minimize total loss.
4. Reconstruct the final image by blending real pixels (unmasked) with generated ones (masked).



Training Losses

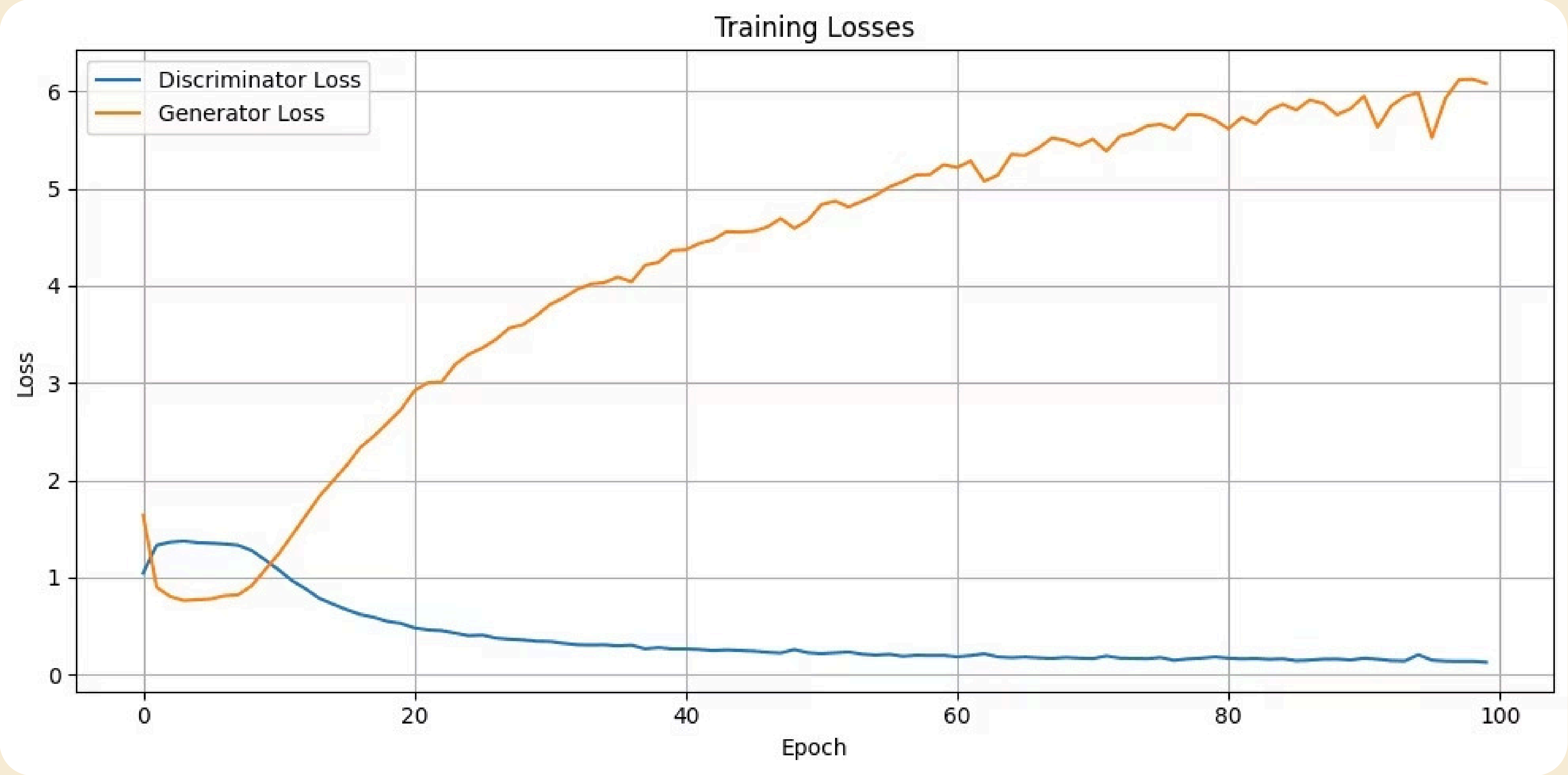


Image Quality: Generated faces show moderate detail; skin tones less even than real images; further training degraded quality, thus we used results from the 9th epoch

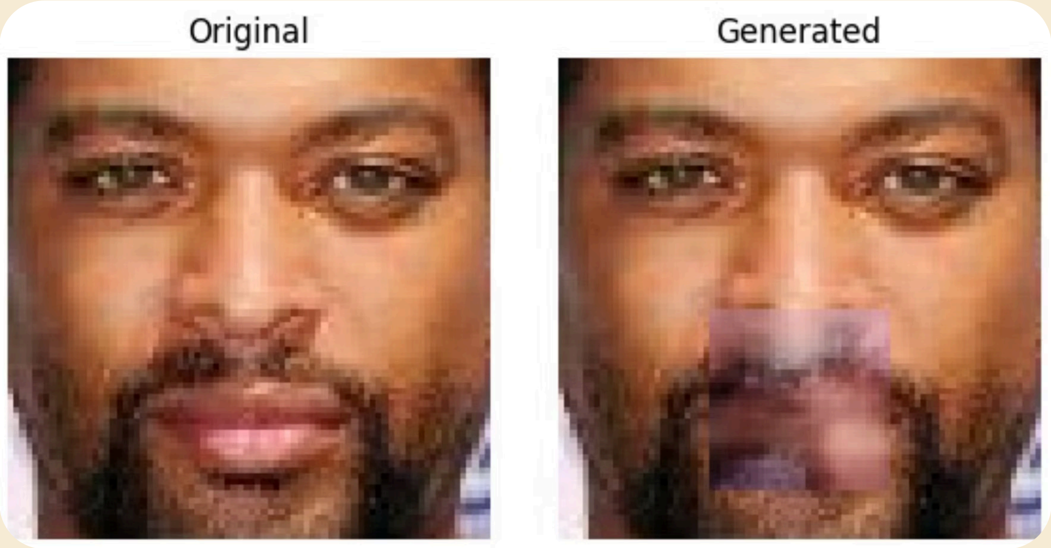
Image Completion Results and Efficiency

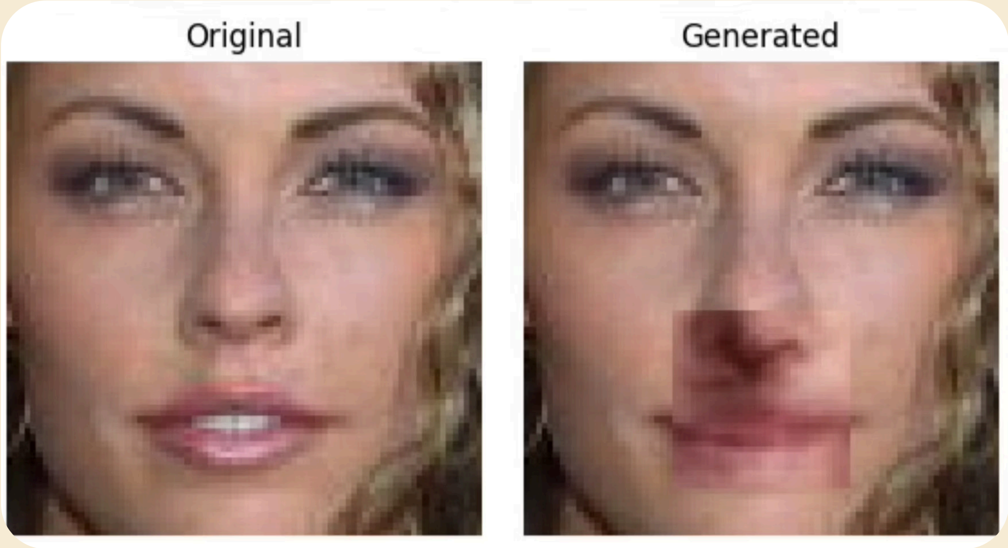
Completion Quality

Completed images are perceptually correct and consistent but lack fine details, with blurry inpainted pixels.

Runtime

Completion process is time-consuming; smaller DCGAN used for faster results with moderate quality.

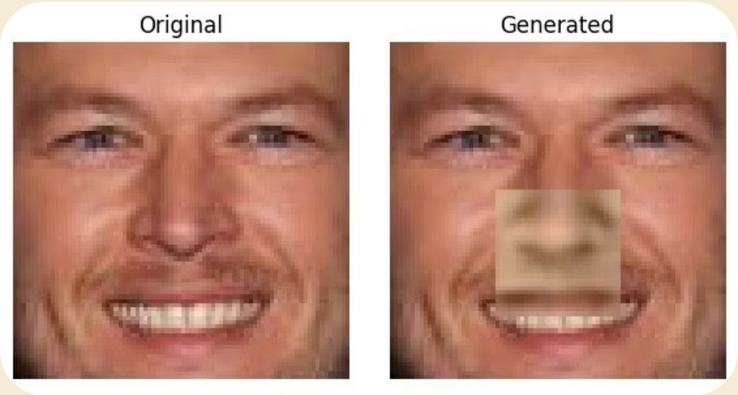


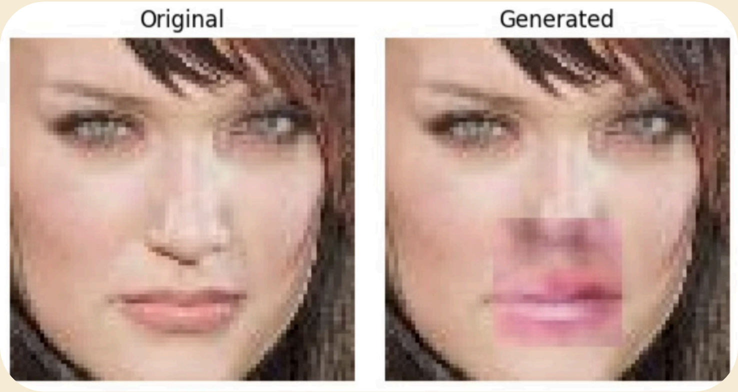


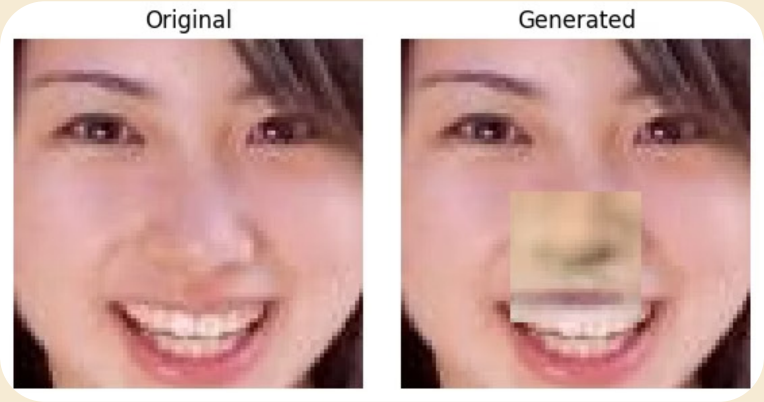
Results

Default Parameters

By default, we are taking 30×30 pixel mask size, and iterating 500 times for image completion.





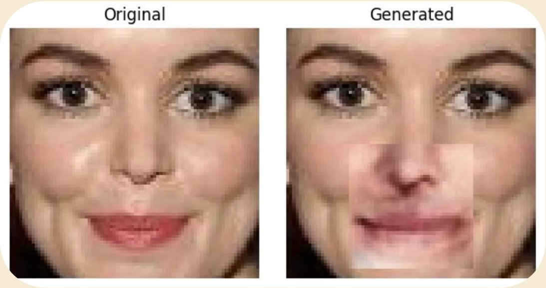


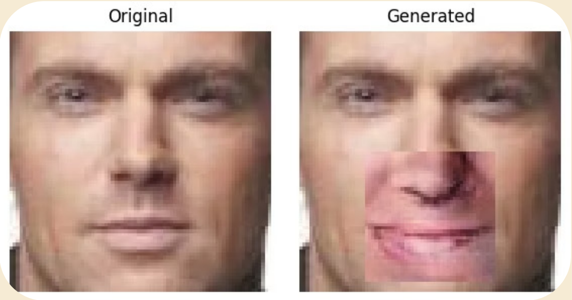
The nose and lip structures are generated properly, however, there is some inconsistency with the skin tone of the person.

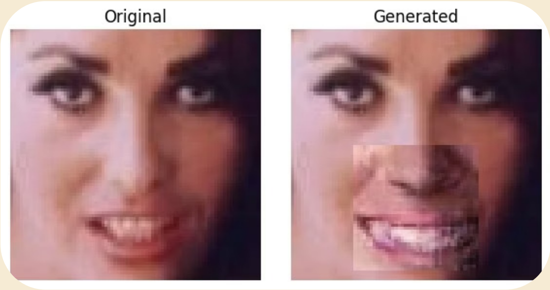
Results

Bigger Mask

Here, we have increased the mask size to 40×40 pixels.





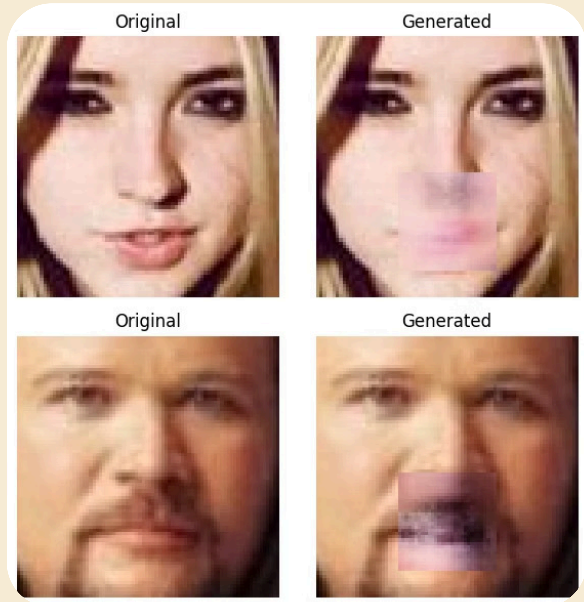


The issue with a bigger mask is that the generator does not have a reference (for example the edges of the lips) to correctly fill the mask.

Results

Higher Iterations

We go back to the 30×30 pixel mask, but now we are iterating 50,000 and 100,000 times respectively.



With higher iterations, while the issue with skin tone is mostly solved, the nose and lip structure is lost.

Architecture and Training Challenges

Architecture Sensitivity

Substituting max-pooling with strided convolutions and adding Batch Normalization layers often worsened results.

Activation Functions

tanh outperformed ReLU and leaky ReLU in hidden layers for this DCGAN implementation.

Training Stability

GAN training is unstable; balancing generator and discriminator is difficult; best model achieved after 9 epochs.

Conclusions

- The model can generate **realistic facial structures**, especially in small missing regions, but struggles with larger masks and fine details.
- **Contextual and perceptual losses** enabled semantically meaningful completions, though at the cost of visual sharpness.
- Observed clear trade-offs between **mask size**, **iteration count**, and **image realism**.
- **GAN training instability** and architectural sensitivity posed challenges, requiring careful tuning.

Future Scope

- **Enhanced Preprocessing:** Select images with uniform lighting and without occlusions (e.g., glasses); Improve straight-face detection using facial landmarks and pose estimation.
- **Flexible Masking Techniques:** Move beyond fixed square masks to free-form and irregular masking.
- **Leveraging Advanced Architectures:** Fine-tune pre-trained models like **StyleGAN** or **Conditional GANs (CGANs)** for sharper and more detailed results; Explore transfer learning to reduce training time and improve generalization.