# Animated Face Image Generation using GANs





# Research Questions

- To effectively train Generative Adversarial Networks (GANs) to generate animated face images that capture human-like expressions and movements

- To achieve high-quality and diverse generated animated face images in GAN models, by integrating different architectural choices

- To optimize the training of GAN models for generating animated faces using reference images

# Objective & Motivation

- The project aimed to create a dynamic animated facial image using deep learning and reference images as input

- Current face animation technologies in the entertainment and gaming industries are inefficient
  - As a result, animations may appear unrealistic
  - Requiring multiple reference images and unable to capture fine details

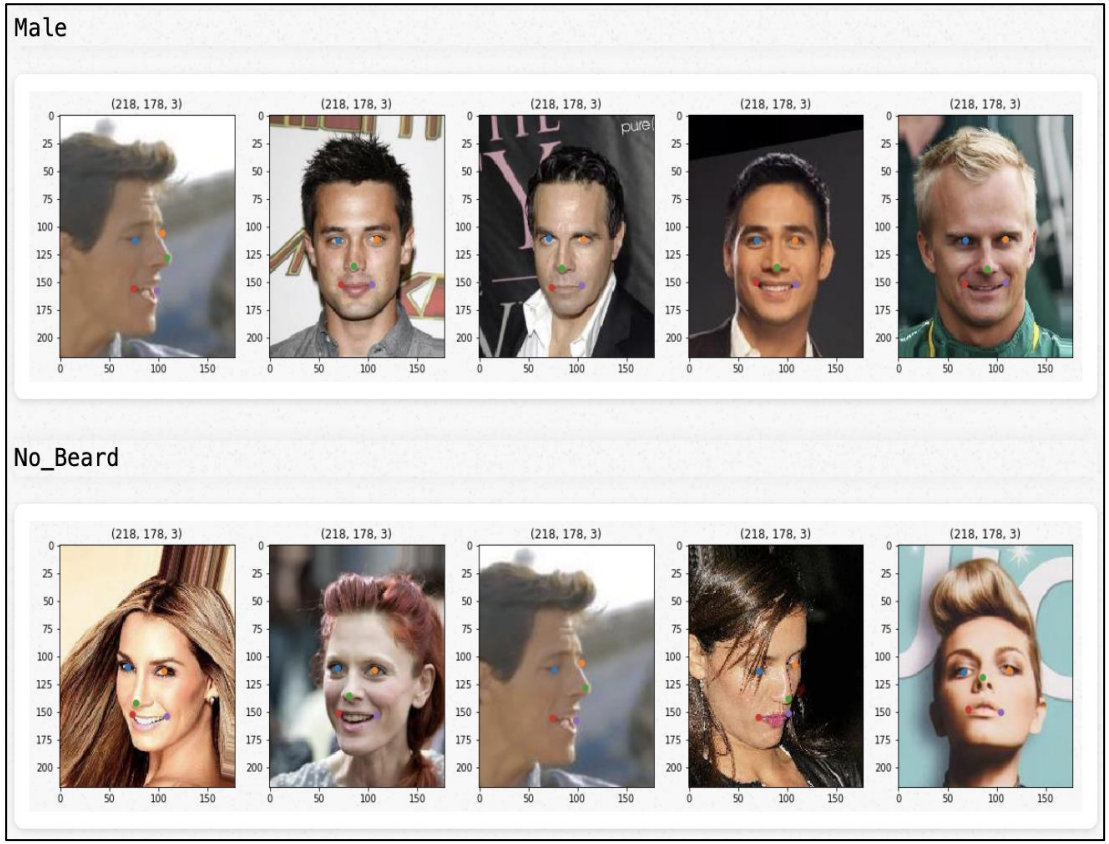
- GAN architecture can address limitations by generating realistic animated face images from reference images
  - Training on a vast dataset to capture intricate details
  - Accurately represent facial style

# Dataset

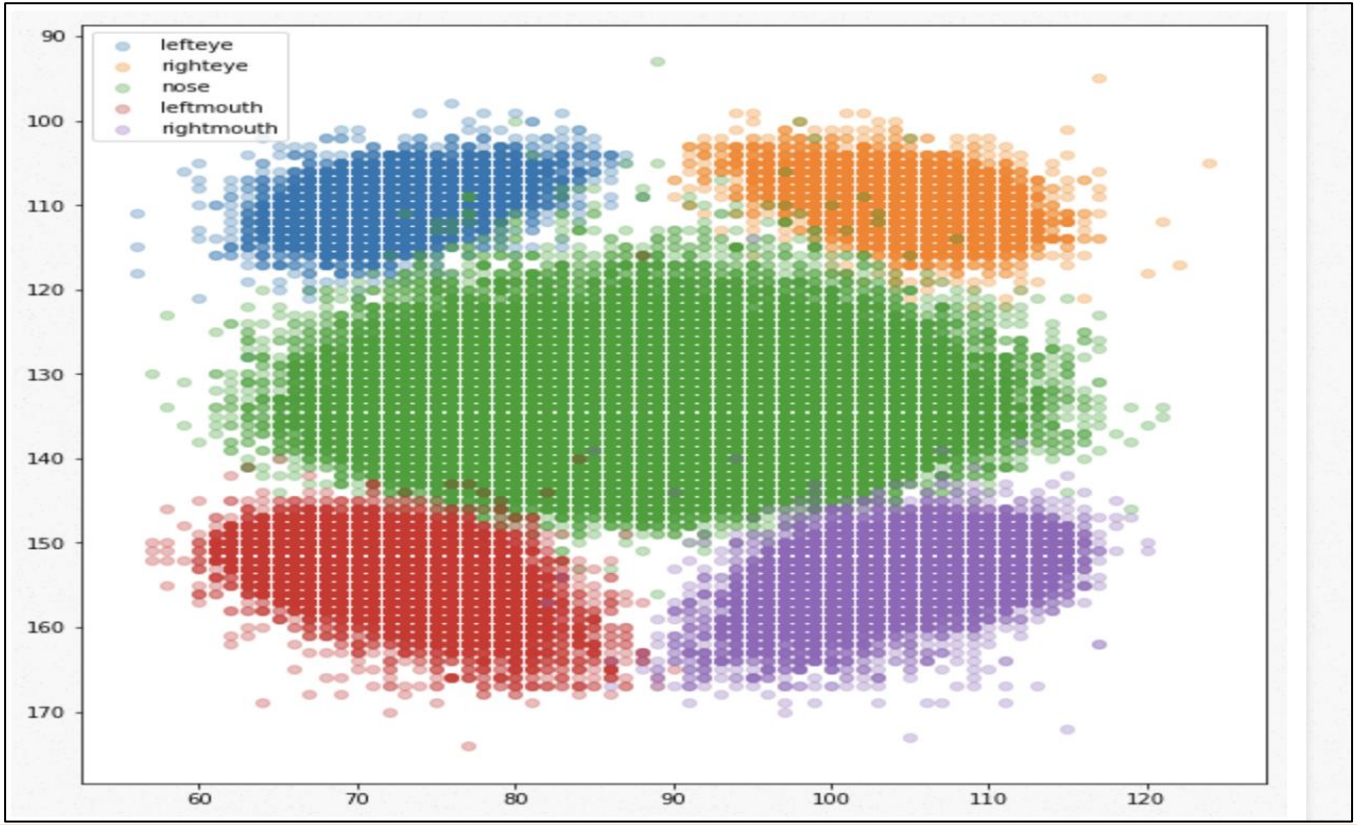
- **CelebFaces Attributes Dataset (CelebA)** is a large-scale face attributes dataset
  - More than **200K** celebrity images, each with **40** attribute annotations
  - **10,177 identities**, **202,599 face images**, and
  - **5 landmark locations**, **40 binary attributes** annotations per image

# Celeb a Dataset EDA

| | |
|---|---|
| Bags_Under_Eyes | 20.46% |
| Bald | 2.24% |
| Bangs | 15.16% |
| Big_Lips | 24.08% |
| Big_Nose | 23.45% |
| Black_Hair | 23.93% |
| Blond_Hair | 14.80% |
| Blurry | 5.09% |
| Brown_Hair | 20.52% |
| Bushy_Eyebrows | 14.22% |
| Chubby | 5.76% |
| Double_Chin | 4.67% |
| Eyeglasses | 6.51% |
| Goatee | 6.28% |
| Gray_Hair | 4.19% |
| Heavy_Makeup | 38.69% |
| High_Cheekbones | 45.50% |
| Male | 41.68% |
| Mouth_Slightly_Open | 48.34% |
| Mustache | 4.15% |
| Narrow_Eyes | 11.51% |
| No_Beard | 83.49% |
| Oval_Face | 28.41% |
| Pale_Skin | 4.29% |
| Pointy_Nose | 27.74% |
| Receding_Hairline | 7.98% |
| Rosy_Cheeks | 6.57% |
| Sideburns | 5.65% |
| Smiling | 48.21% |
| Straight_Hair | 20.84% |
| Wavy_Hair | 31.96% |
| Wearing_Earrings | 18.89% |



# EDA


lefteye
righteye
nose
leftmouth
rightmouth
90
100
110
120
130
140
150
160
170
60
70
80
90
100
110
120

# Cartoon Dataset

- Safebooru cartoon dataset used for training
- Safebooru is a tag-based image
- Maintained by anime enthusiasts
- Includes different unique attributes data
    - Golden eyes
    - Brown eyes
    - Blonde hair
    - Smile
    - Red hair

## Data Preprocessing

**Image Resizing**

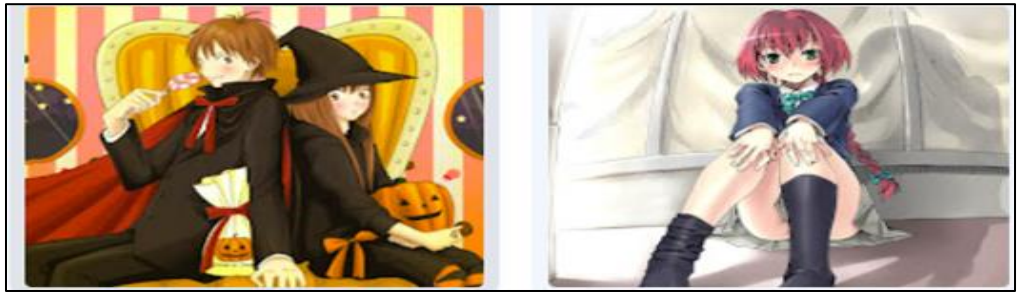


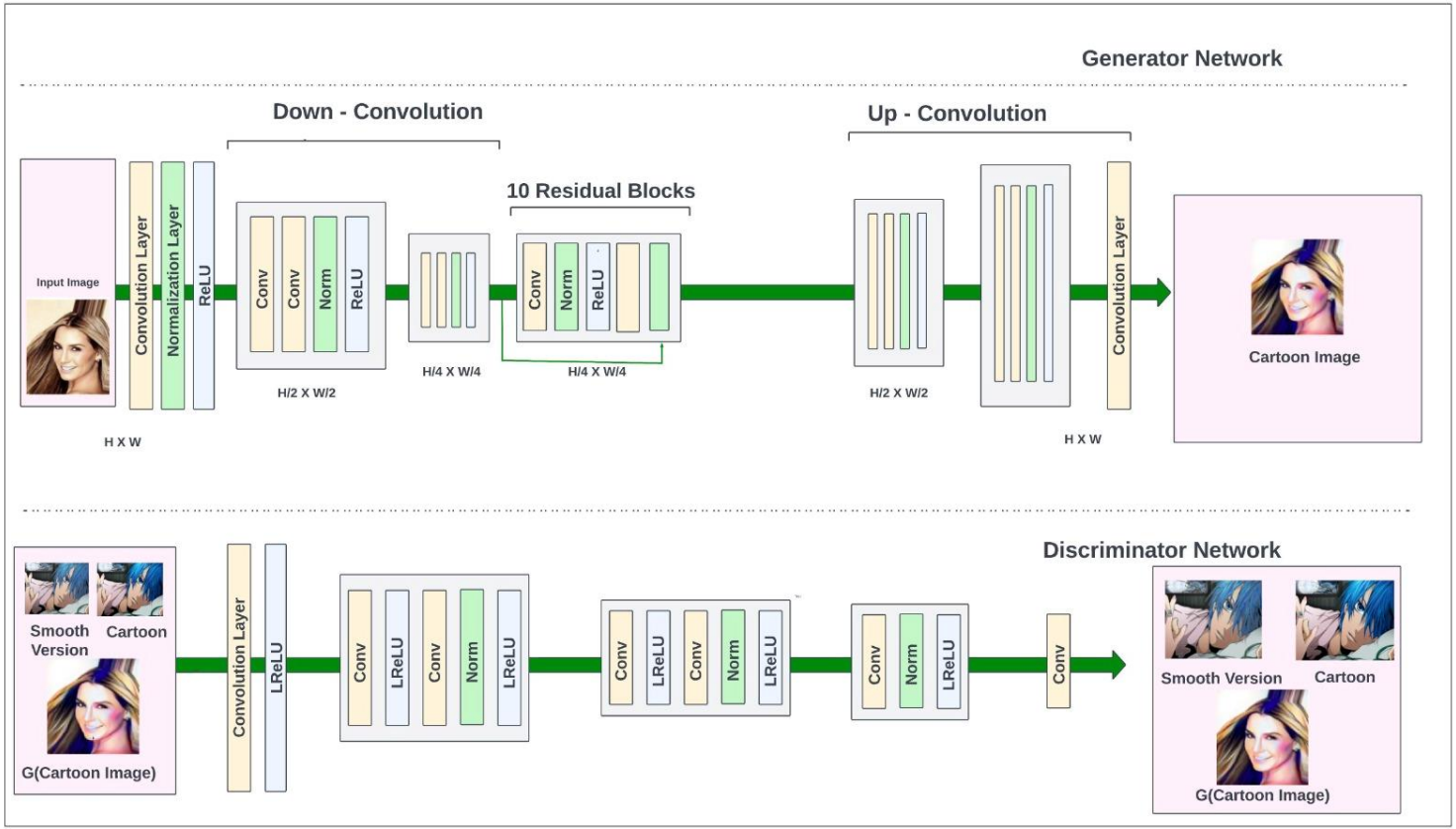
**Image Smoothing**

# Literature Review

| Model | Year | Type | Dataset | Training Approach | Output Style | Complexity | Advantages |
|---|---|---|---|---|---|---|---|
| AnimeGAN | 2020 | GAN | Content Images-Real World Photos, Style Images- Anime Images | Unsupervised | Anime | Relatively Low Complexity | Low Computational Cost, High Quality Results, Flexibility |
| JOJOGAN | 2022 | GAN | Gan Inversion Produced | Unsupervised | Cartoon | High Complexity | No external dataset required, Produce output using Reference style |
| Cartoon StyleGAN-2 | 2021 | Style-based GAN | Webtoon, Disney, Metfaces images | Unsupervised | Cartoon | High Complexity | Can generate high-quality, diverse cartoonish images. Allows control over various image attributes through latent space manipulation. |
| CartoonGAN | 2018 | cGAN | Real World Photos from Coco Dataset, Cartoon Images, Smoothed version of Cartoon Images | Unsupervised | Cartoon | Moderate Complexity | Preserves content while transforming images into cartoon style, Good balance between computational cost and quality results, Flexibility in generating diverse cartoon-like images. |

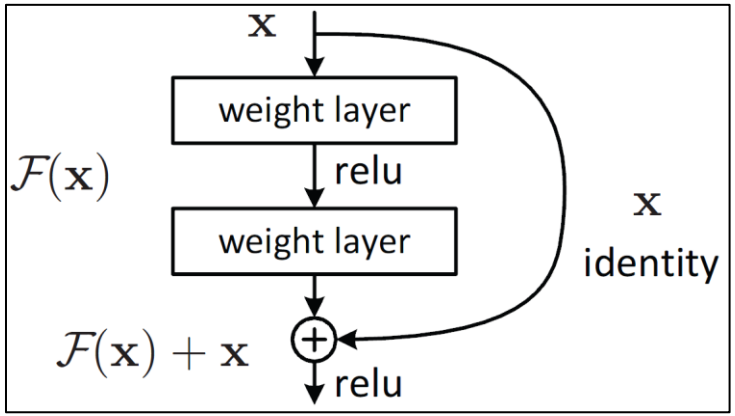
# ArtifyGAN

## Model Architecture



# ArtifyGAN

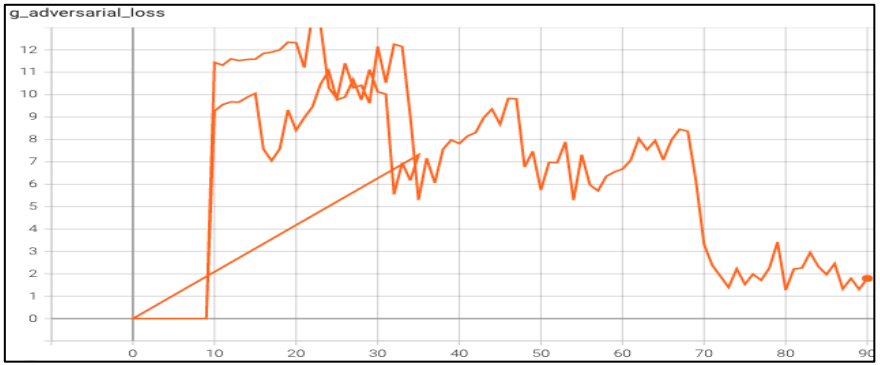
- Inspired from CartoonGAN architecture
- Increased the ResNet blocks from 8 to 10 to
  - Improve feature extraction
  - Better representation learning
  - Increase Model capacity
- Removed the Convolution filters after the ResNet blocks to improve the computational efficiency
- Coco Dataset and celeb a was used instead of Flickr data
- Safebooru Cartoon dataset was used instead of extracting images from anime movie “Spirited Away”

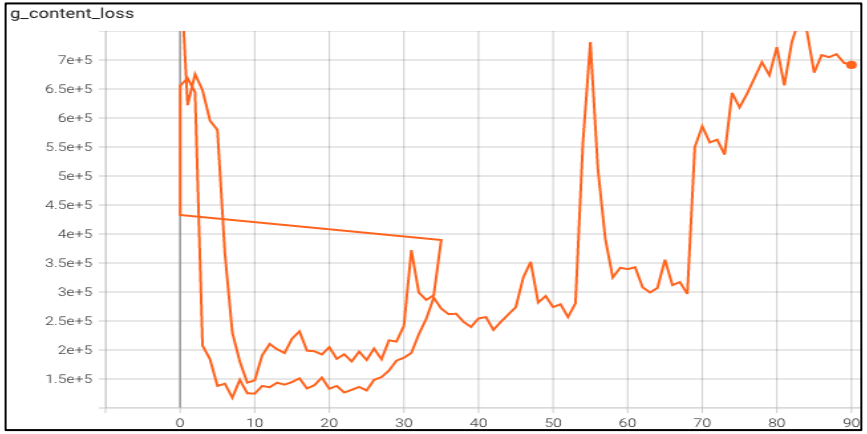


# Experimental Settings

| Category | Requirement |
|---|---|
| Hardware | GPU (Google Colab) |
| Software | Google Colab (Jupyter notebook) |
| Libraries | PyTorch, NumPy, Matplotlib, PIL, OpenCV, tqdm, TensorBoard |

| Hyper-parameters | |
|---|---|
| Optimizer | Adam<br>Learning rate: 0.001 |
| w | 0.00001 |
| Epochs | 240 |
| Batch Size | 16 |
| Loss Function | Adversarial Loss |



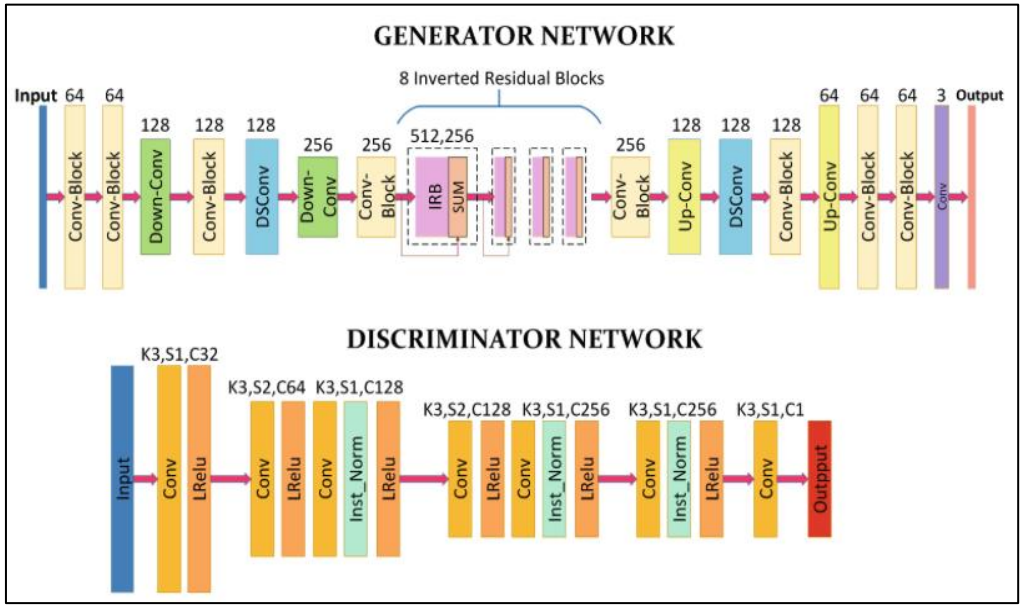
# AnimeGAN

## Pre-trained Model Experimental Setting

Fine Tuning:

- Fine-tuned and trained on preprocessed animated images (maintained the ratio of Anime and celebA)
- Reduced the training time by decreasing the size of dataset

| AnimeGAN – Pretrained Model | Hyperparameter |
|---|---|
| Generator – Learning Rate | 0.00002 |
| Discriminator – Learning Rate | 0.00004 |
| Training Epochs | 40 |
| Batch Size | 26 |
| Optimizer | Adam |
| Freezing Generator Layer | Until_layer 5 |
| Freezing Discriminator Layer | Until_layer 5 |
| Real Images : Style Images Ratio | 6000:2000 |



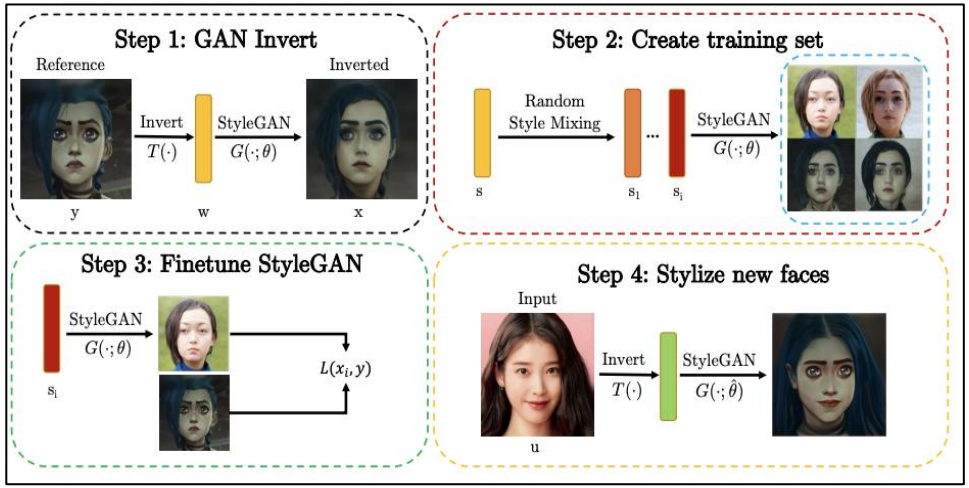
# JOJOGAN

## Pre-trained Model Experimental Setting

Fine Tuning:

- Fine-tuned and trained on preprocessed 1 cartoon style image
- Reduced the training time(ms) and the memory consumption by using the decreased discriminator channel size.

| JOJOGAN – Pretrained Model | Hyperparameter |
|---|---|
| Generator – Learning Rate | .0002 |
| Discriminator – Learning Rate | 0.0003 |
| Training Epochs | 50 |
| Batch Size | input.shape[0](training with 1 image at 1 time) |
| Optimizer | Adam |
| Added Generator Layer | 3 |
| Added Discriminator Layer | 3 |
| Channel Size | 512 |



# JOJOGAN output



DATA255 Project Presentation

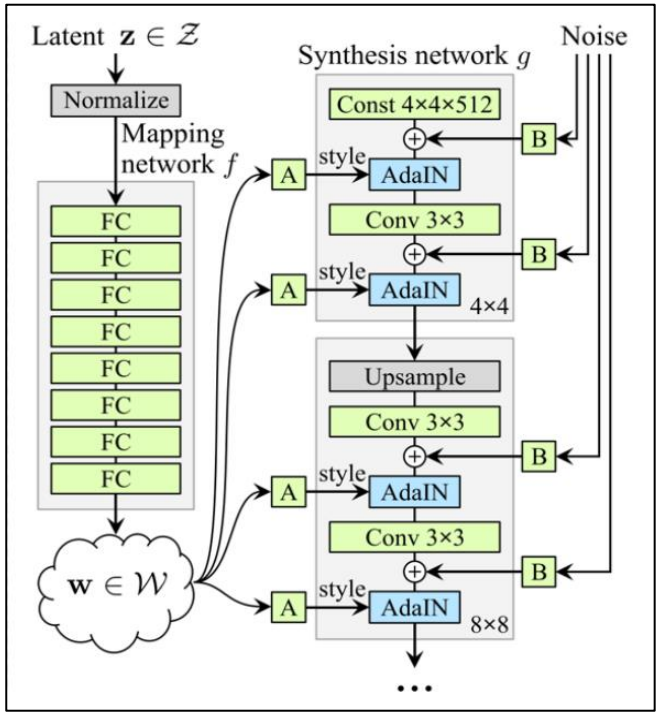
# StyleGAN-2

## Pre-trained Model Experimental Setting

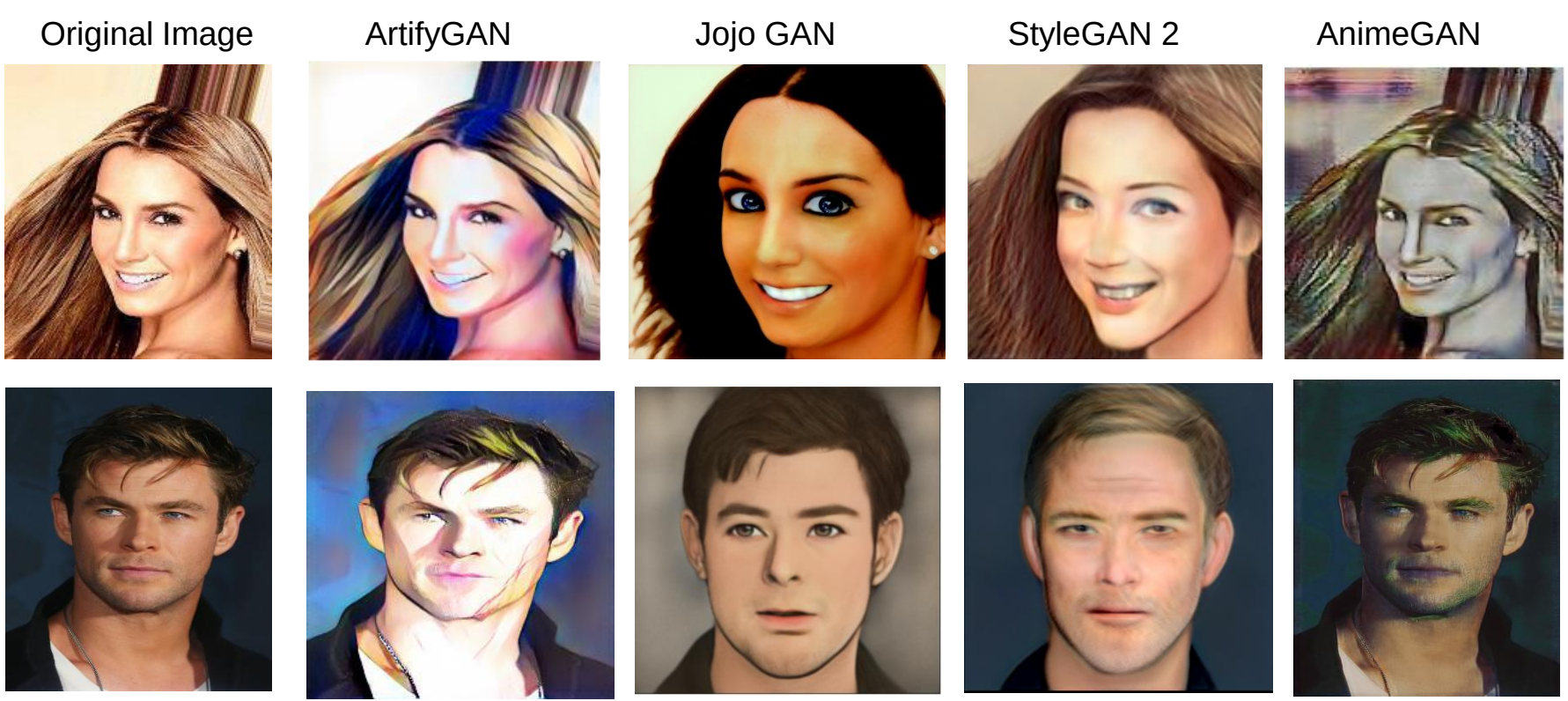
Fine Tuning:

- Fine-tuned and trained on preprocessed data (Image Upsizing, Face Alignment, Converting 32-bit images to 24-bit images)

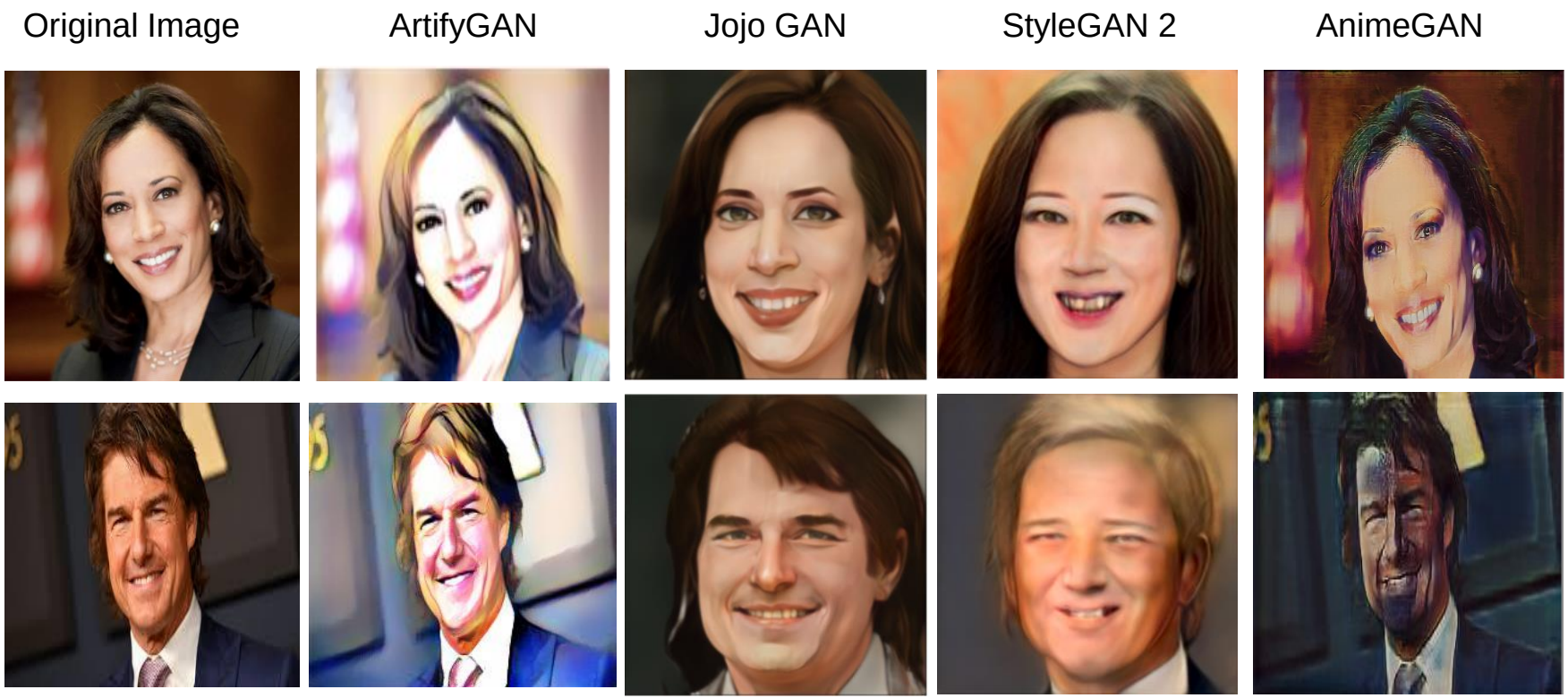
| StyleGAN-2 ADA | Hyper-parameter tuning |
|---|---|
| Number of Layers Freezed | 4 |
| Learning rate, optimizer | 0.002, Adam |
| Style mixing Probability (gamma) | 50 |
| Kimg (Number of images the model will be trained on) | 1200 |



# Result Analysis



# Result Analysis



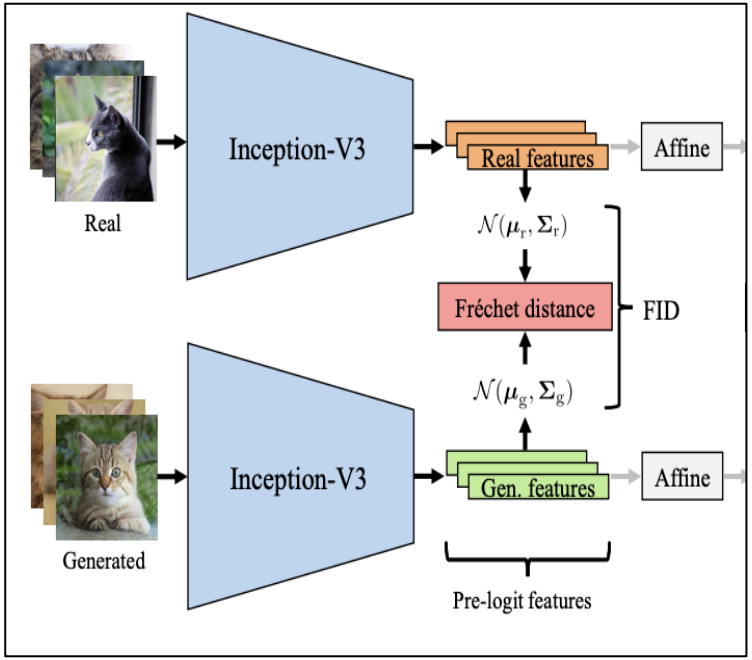
# Evaluation Metrics FID (Frechet Inception distance)

The key idea behind in FID is to separately embed real and generated images to a vision-relevant feature space, and compute a distance between the two distributions

Step 1. Images passed through a pre-trained classifier network (eg. Inception-V3) to produce feature vectors

Step 2. Both distributions of features are approximated to calculate the distance

Lower is better !!



# Evaluation Metrics

FID-Metric

| Model | Image1 | Image2 | Image3 | Image4 | Image5 | Image6 | Mean | Total |
|---|---|---|---|---|---|---|---|---|
| AnimeGAN | 201.03 | 212.64 | 187.43 | 147.10 | 210.36 | 155.08 | 185.60 | 1113.64 |
| StyleGAN2 | 93.14 | 100.46 | 137.99 | 77.84 | 140.91 | 137.6262 | 114.65 | 687.99 |
| **ArtifyGAN** | **107.10** | **89.81** | **96.70** | **103.16** | **179.13** | **135.91** | **118.63** | **711.82** |
| JOJOGAN | 183 | 177 | 170 | 180 | 190 | 173 | 178.8 | 1073 |

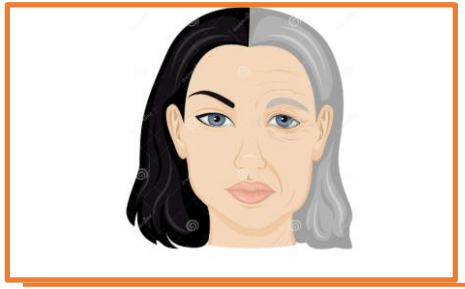
# Conclusion

ArtifyGAN is less complex as compared to other existing models for generating animated images

Designing the model from scratch allowed us to define the architecture and modify hyperparameters to better suit our specific dataset and target style

Although ArtifyGAN was trained on a smaller dataset and for less training time as compared to other pre-trained models, it is still producing high-quality animated results

# Future Works



- **Expression transfer:** To transfer expressions between different faces in the dataset
- **Facial feature modification:** To modify other facial features like hairstyle, adding accessories like glasses or hat
- **Training on a larger and diverse dataset**: To improve its generalization capabilities
- **Age progression:** To create animated images of faces that look older or younger than the original input.

# References


Back, J. (2021). Fine-Tuning StyleGAN2 For Cartoon Face Generation. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2106.12445

Chen, J., Liu, G., & Chen, X. (2019). AnimeGAN: A Novel Lightweight GAN for Photo Animation. In *Communications in computer and information science* (pp. 242–256). Springer Science+Business Media. https://doi.org/10.1007/978-981-15-5577-0_18

Chong, M. J., & Forsyth, D. (2022, October). Jojogan: One shot face stylization. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVI* (pp. 128-152). Cham: Springer Nature Switzerland.

Chen, Y., Lai, Y., & Liu, Y. (2018). *CartoonGAN: Generative Adversarial Networks for Photo Cartoonization*. https://doi.org/10.1109/cvpr.2018.00986

# Thank you