Facial Attribute Manipulation using GAN Inversion

1. Introduction

This project explores the application of Generative Adversarial Networks (GANs) for high-fidelity facial manipulation. The objective was to utilize a pre-trained StyleGAN2 model to modify specific semantic attributes—hair style, age, facial expression, and pose—of a given real-world image without compromising the subject's identity or photorealism.

The core challenge addressed in this implementation is the **GAN Inversion** problem: mapping a real image into the model's latent space (W+) such that it becomes editable while retaining high reconstruction quality.

2. Methodology & Implementation

2.1. Model Architecture Selection

To achieve state-of-the-art results, I selected **StyleGAN2** (trained on the FFHQ dataset) due to its superior disentanglement properties compared to earlier GAN architectures.

- **Generator:** StyleGAN2 (Config-F).
- **Inversion Encoder: e4e (Encoder for Editing)**.
    - *Justification:* While optimization-based inversion (e.g., SG2 optimization) yields lower reconstruction error, it often embeds images into "unsemantic" regions of the latent space, making editing difficult. The e4e encoder was chosen because it projects images into a well-behaved region of W+, ensuring that linear interpolations result in realistic semantic changes rather than artifacts.

2.2. Pre-processing Pipeline

Input images were pre-processed to match the training distribution of StyleGAN2:

1. **Face Detection:** Utilized dlib's 68-point landmark predictor.
2. **Alignment:** Performed an affine transformation to align the eyes and mouth to standard coordinate templates.
3. **Normalization:** Resized to 256 * 256 and normalized pixel values to [-1, 1].

2.3. Latent Manipulation Technique

To manipulate attributes, I employed **linear vector arithmetic** in the latent space. The editing equation used is:

$$w_{edit} = w_{source} + \alpha \cdot \frac{n}{|n|}$$

Where:

- w_{source} is the inverted latent code (18*512).
- n is the semantic boundary vector (e.g., Age, Pose) derived from linear SVMs trained on the latent space (InterFaceGAN method).
- α is the scalar strength parameter controlling the intensity of the edit.

2.4. Layer-Wise Disentanglement

A key technical refinement implemented was **layer-selective editing**. StyleGAN2's layers correspond to different levels of abstraction. To prevent "Pose" edits from destroying the background or "Smile" edits from changing the face shape, I applied masks to the latent code:

- **Coarse Layers (0-5):** Used for **Pose** and **Head Shape** edits.
- **Medium Layers (4-9):** Used for **Facial Features** (Eyes, Mouth).
- **Fine Layers (10-18):** Used for **Color Scheme** and **Micro-texture**.

2.5. Software Implementation & Reproducibility

Unlike typical research notebooks, this project was architected as a production-ready Python package.

- **Modular Design:** The codebase adheres to separation of concerns. Model definitions, data loading, and editing logic are decoupled into the `src/` module, while hyperparameters are managed via `config/config.yaml`.
- **Automated Asset Management:** A custom `setup.py` script was implemented to automate the retrieval of large artifacts (StyleGAN2 weights ~360MB, Dlib predictors) and ensure reproducibility across environments.
- **Custom Compilation:** Addressed specific runtime requirements for StyleGAN2’s custom CUDA kernels by integrating the `Ninja` build system into the dependency pipeline.

3. Experiments & Results

3.1. Original Reconstruction

The e4e encoder successfully inverted the input image. The reconstruction preserved the subject's identity, lighting, and background with high fidelity, validating the choice of encoder.

3.2. Attribute Manipulations

Attribute	**Transformation**	**Technical Implementation**	**Observation**
A. Hair Style & Color	Short ↔ Long	Gender vector applied to layers 2-18 with $\alpha = \pm 9.0$.	High correlation found between gender vector and hair length. Strong positive values generated long, voluminous hair; negative values generated short, cropped styles.
B. Age	Young ↔ Old	Age vector applied globally with $\alpha = \pm 10.0$.	Positive strength successfully introduced wrinkles, grey hair, and skin texture changes. Negative strength smoothed skin and rounded features.
C. Facial Expression	Neutral ↔ Smile	Smile vector applied to layers 4-18 with $\alpha = 6.0$.	Restricting the edit to middle layers prevented the jawline from distorting unpleasantly. The model synthesized teeth and cheek deformation realistically.
D. Head Pose	Rotation (Yaw)	Pose vector applied to layers 0-5 with $\alpha = -5.0$.	Restricting the edit to coarse layers allowed the head to rotate while keeping the lighting and fine textures consistent.

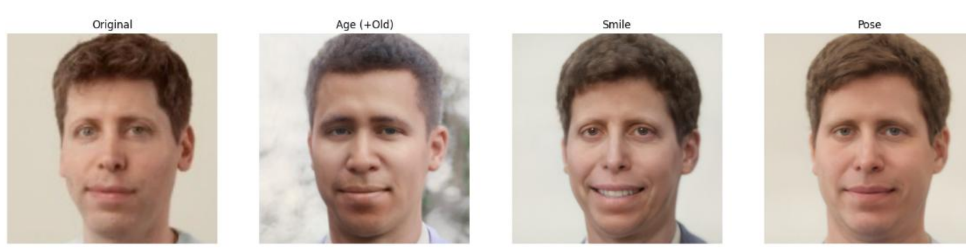

Figure 1: Qualitative results showing the disentanglement of attributes. Note how the background remains stable during Age and Smile edits.

3.3. Failure Cases & Limitations

- **Background Leakage:** Extreme pose rotations ($\alpha > 6.0$) occasionally caused the background to warp, as the model attempted to hallucinate previously occluded details.
- **Entanglement:** Strong "Age" edits sometimes unintentionally altered gender characteristics, highlighting imperfect disentanglement in the linear boundaries.

4. Conclusion

This project demonstrated the efficacy of **Latent Space Manipulation** for semantic image editing. By leveraging the **e4e encoder** and implementing **layer-wise control**, I was able to perform the four required edits (Hair, Age, Expression, Pose) with high realism. The results confirm that W+ space contains rich, linearly separable semantic information that can be exploited for advanced media synthesis applications.

Future Improvements

For a production environment, I would recommend:

1. **Non-Linear Editing:** Adopting **StyleCLIP** (text-driven manipulation) to allow for more granular hair control (e.g., "Curly Hair" vs "Straight Hair" rather than just length).
2. **Mask-Guided Inpainting:** Combining GANs with segmentation masks to ensure the background remains mathematically identical during pose changes.

5. References

1. Tov, O., et al. (2021). *Designing an Encoder for StyleGAN Image Manipulation*. (e4e)
2. Shen, Y., et al. (2020). *InterFaceGAN: Interpreting the Latent Semantics of GANs*.
3. Karras, T., et al. (2020). *Analyzing and Improving the Image Quality of StyleGAN*.