Unpaired Face Restoration via Latent Space Exploration

Presented By:
Shubham Harsh
MML2023004
Under the supervision of:
Dr. Shivram Dubey

Introduction

- • Objective: Restore HQ face images from LQ images with unknown degradations.
- • Motivation:
- - Real-world face restoration is challenging due to unseen degradations.
- - Paired data is often unavailable for supervision.
- • Key Innovation:
- - Leverages StyleGAN latent space for unpaired restoration via a learnable cross-quality shift.
- - Proposes a two-branch framework to improve fidelity and handle unpaired datasets.

Key Concepts

- • Paired vs. Unpaired Face Restoration:
- - Paired: Requires aligned HQ and LQ images, limited to known degradation types.
- - Unpaired: No need for paired data, making it robust to unseen degradations.
- • StyleGAN Latent Space:
- - HQ and LQ images naturally form separate subspaces.
- - Transforms LQ → HQ via simple vector arithmetic using a learned cross-quality shift.

Methodology

- • Learnable Cross-Quality Shift:
- - A trainable latent-space transformation applied in StyleGAN's W space.
- - Enables direct conversion between LQ and HQ representations.
- - Provides adjustable restoration levels by scaling the shift.
- • Two-Branch Framework:
- - HQ Branch: Encodes HQ images and generates reconstructed HQ images.
- - LQ Branch: Encodes LQ images with degradation estimation to map to HQ space.
- - Cycle-Consistency: Ensures input and reconstructed outputs retain the same content.
- • Degradation Modeling:
- - Estimates degradations such as blur, noise, and compression directly in latent space.
- - Generates synthetic degraded versions of HQ images to train robustly.

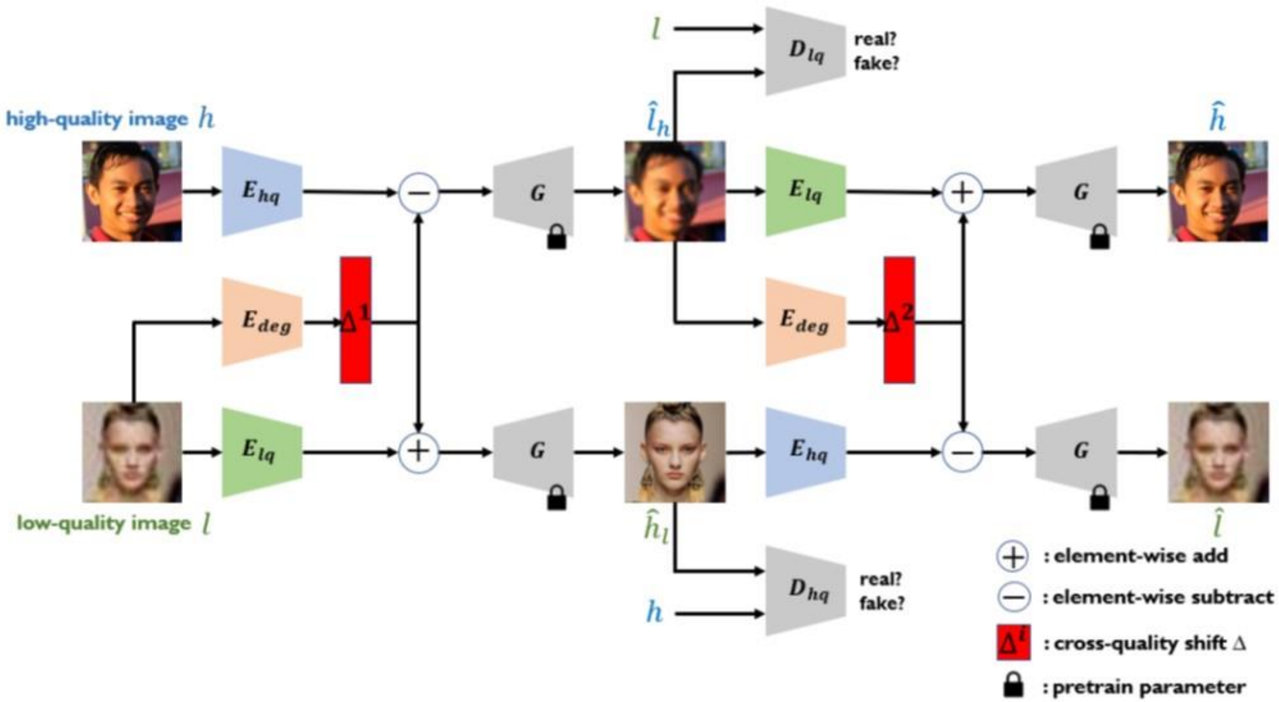
Dataset

- • HQ Dataset:
- - FFHQ (Flickr-Faces-HQ): 70,000 high-quality images at 1024×1024 resolution.
- - Preprocessed to align faces and resized to 256×256.
- • LQ Dataset:
- - Derived from CelebAHQ.
- - Simulated degradations:
- - Mild: Downsample with Bicubic (factor 8) + JPEG compression (quality 90–95).
- - Moderate: Downsample + Add Gaussian Noise (σ=20–25).
- - Severe: Blur with Gaussian Kernel + Downsample + Gaussian Noise (σ=10–15).
- - Ensures no overlap with FFHQ for unpaired training.


Original Image
Mild Degradation
Moderate Degradation

Architecture

- • Core Components:
- - Encoders: Extract latent codes.
- - StyleGAN: Maps latent codes back to the image space.
- - Discriminators: Ensure outputs resemble LQ and HQ distributions.
- • Two-Branch Framework:
- - Upper Branch: Encodes HQ images.
- - Lower Branch: Encodes LQ images and estimates degradations with Edeg.
- - Cycle Consistency: Adds constraints to maintain fidelity.



Dong, Yangyi, et al. "Unpaired Face Restoration via Learnable Cross-Quality Shift." Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023.

Training:

- Used pre-trained encoder for high quality images.
- Trained a resnet-34 backed encoder for synthetic low-quality images.
- Pre- training of encoder done for 500 epochs.
- Training of the whole architecture done for 250 epochs.

```
(490) E_h2l: 0.052, E_l2h: 0.074
(491) E_h2l: 0.059, E_l2h: 0.188
(492) E_h2l: 0.062, E_l2h: 0.115
(493) E_h2l: 0.127, E_l2h: 0.260
(494) E_h2l: 0.065, E_l2h: 0.043
(495) E_h2l: 0.072, E_l2h: 0.139
(496) E_h2l: 0.057, E_l2h: 0.117
(497) E_h2l: 0.089, E_l2h: 0.087
(498) E_h2l: 0.047, E_l2h: 0.111
(499) E_h2l: 0.074, E_l2h: 0.070
(500) E_h2l: 0.079, E_l2h: 0.121
```

Training of encoders

```
(240) D_h2l: 2.110, D_l2h: 1.367
(241) D_h2l: 2.344, D_l2h: 0.512
(242) D_h2l: 1.972, D_l2h: 2.285
(243) D_h2l: 2.033, D_l2h: 2.418
(244) D_h2l: 2.047, D_l2h: 1.134
(245) D_h2l: 1.820, D_l2h: 1.460
(246) D_h2l: 2.628, D_l2h: 1.985
(247) D_h2l: 2.174, D_l2h: 1.216
(248) D_h2l: 1.923, D_l2h: 0.786
(249) D_h2l: 1.805, D_l2h: 1.598
(250) D_h2l: 2.290, D_l2h: 2.099
```

Training of whole architecture

Test Instance:

Experimental Results

Quantitative Results:

- Metrics: MSE, PSNR, LPIPS

Method	**MSE (↓)**	**PSNR (↑)**	**LPIPS (↓)**
Model Paper	30.45	N/A	0.13
This implementation	295.3	24.22	-0.26

The LPIPS score of -0.26 indicates that the model is functioning as intended for the task; however, there is room for improvement to achieve the desired performance levels for optimal results.

Conclusion

Future Work:

- Plan 1: Will to try to make this paper work, complete the implementation
- Plan 2: Will move to a new model, more apt. for the problem statement.

References:

- “Dong, Yangyi, et al. "Unpaired Face Restoration via Learnable Cross-Quality Shift." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.”
- Karras, Tero, Samuli Laine, and Timo Aila. "A style-based generator architecture for generative adversarial networks." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.
- Shen, Yujun, et al. "Interpreting the latent space of gans for semantic face editing." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
- Karras, Tero. "Progressive Growing of GANs for Improved Quality, Stability, and Variation." arXiv preprint arXiv:1710.10196 (2017).