CS663 Assignment-5

Saksham Rathi, Kavya Gupta, Shravan Srinivasa Raghavan
Department of Computer Science,
Indian Institute of Technology Bombay

Question 3

Solution

The paper titled *Towards Real-World Blind Face Restoration with Generative Facial Prior* addresses the problem of **real-world blind face restoration** in images with unknown and complex degradation, such as low resolution, noise, blur, and compression artifacts. Blind restoration approaches aim to handle complex real-world images that have mixed degradation types.
Venue and Publication Year: Presented on arXiv, 2021.
Cost Function Optimized: The GFP-GAN model optimizes a multi-term loss function to balance restoration fidelity, realism, and perceptual quality. The total loss L_{total} is given by:

$$L_{\text{total}} = L_{\text{rec}} + L_{\text{adv}} + L_{\text{comp}} + L_{\text{id}}$$

where each term is detailed below.
1. *Reconstruction Loss* L_{rec}: This loss encourages the restored image $\hat{y}$ to match the ground-truth image y, using both an L_1-norm loss and a perceptual loss to maintain pixel-level and feature-level similarity:

$$L_{\text{rec}} = \lambda_{\text{L1}} \|\hat{y} - y\|_1 + \lambda_{\text{per}} \|\phi(\hat{y}) - \phi(y)\|_1$$

where:
- $\|\hat{y} - y\|_1$ is the pixel-wise L_1 loss.
- ϕ represents a pretrained feature extractor (e.g., VGG-19) for computing perceptual similarity.
- λ_{L1} and λ_{per} are weights for the L_1 and perceptual losses, respectively.

2. *Adversarial Loss* L_{adv}: This loss encourages the restored image to appear realistic by guiding it towards the natural image distribution using a discriminator:

$$L_{\text{adv}} = -\lambda_{\text{adv}} \mathbb{E}_{\hat{y}} \left[\text{softplus}(D(\hat{y})) \right]$$

where:
- D is the discriminator function.
- λ_{adv} is the adversarial loss weight.

3. *Facial Component Loss* L_{comp}: To enhance perceptually significant facial components (e.g., eyes, mouth), this loss incorporates local discriminators and a feature style loss based on the Gram matrix:

$$L_{\text{comp}} = \sum_{\text{ROI}} \left(\lambda_{\text{local}} \mathbb{E}_{\hat{y}_{\text{ROI}}} [\log(1 - D_{\text{ROI}}(\hat{y}_{\text{ROI}}))] + \lambda_{\text{fs}} \|\text{Gram}(\psi(\hat{y}_{\text{ROI}})) - \text{Gram}(\psi(y_{\text{ROI}}))\|_1 \right)$$

where:
- ROI denotes regions of interest (e.g., left eye, right eye, mouth).
- D_{ROI} is the local discriminator for each region.
- ψ represents multi-resolution features from the local discriminators.
- $\text{Gram}(\cdot)$ computes Gram matrix statistics, capturing texture information.
- λ_{local} and λ_{fs} are weights for the local discriminative and feature style losses, respectively.

4. *Identity Preserving Loss* L_{id}: This loss helps maintain the identity of the face in the restored image by measuring the similarity between the restored image $\hat{y}$ and the ground-truth y in a feature space:

$$L_{\text{id}} = \lambda_{\text{id}} \|\eta(\hat{y}) - \eta(y)\|_1$$

where:
- η is a face feature extractor (e.g., ArcFace) that captures identity-relevant features.
- λ_{id} is the weight for the identity preserving loss.

Each term in the total loss function L_{total} thus contributes to the balance of high fidelity, perceptual quality, and realistic facial restoration in degraded images.