Simulating Real-World Challenges: Blind Face Restoration and Upscaling

Rodrigo Golan
Elena Zoppellari

Università degli Studi di Padova

Outline

Introduction

Università degli Studi di Padova

Blind Face Restoration: reconstruct high-quality face images with unknown deterioration.

We use a progressive 3-step GAN architecture with attention mechanism based on facial landmarks.

Two distinct objectives:

1. Restore & Upscale 16×16 blind face images to 128×128
2. Restore 128×128 blind face images to the same resolution

Dataset pre-processing (1)

Labeled Faces in the Wild dataset:

- contains 13,233 face photographies collected from the web, we randomly select 10,586 images for train and 2,650 for test.
- It was originally built for face recognition task, thus poses are natural and unconstrained: presence of sunglasses, hats, multiple faces, emotive expressions, objects that hide part of the face...
- Target images for both objectives are resized to 128×128.

Dataset pre-processing (2)

Università degli Studi di Padova

To create blind images from target images, we apply a combination of degradation techniques:

$$I_{blind} = JPEG_q[(I_{orig} \star k_\sigma) \downarrow_{s \times s} + n_\delta] \uparrow_{u \times u} \quad (1)$$

where q, k, σ, s and δ are randomly chosen for each image from:

$u \times u$	k	σ	s	δ	q
16×16	[1, 5]	[0, 5]	[43, 128]	[0, 10]	[60, 100]
128×128	[1, 11]	[0, 10]	[16, 128]	[0, 30]	[50, 100]

Table: Random parameters ranges

batch size: 32

batch size: 16

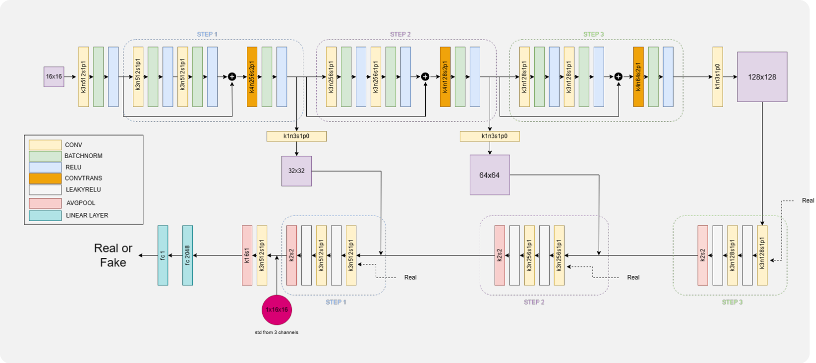
	PSNR↑	SSIM↑	MS-SSIM↑	LPIPS↓	NIQE↓	FID↓
16x16	20.1 ± 0.2	0.59 ± 0.01	0.818 ± 0.002	0.468 ± 0.003	16.78 ± 0.05	240 ± 4
128x128	21.1 ± 0.7	0.57 ± 0.03	0.83 ± 0.02	0.52 ± 0.03	9.8 ± 0.5	151 ± 14

Table: Batch metrics for input blind images

Progressive GAN architecture

Università degli Studi di Padova

For the first objective, we employ a **3-step GAN architecture**, similar to the one proposed by Kim et al.[1] for Super Resolution.

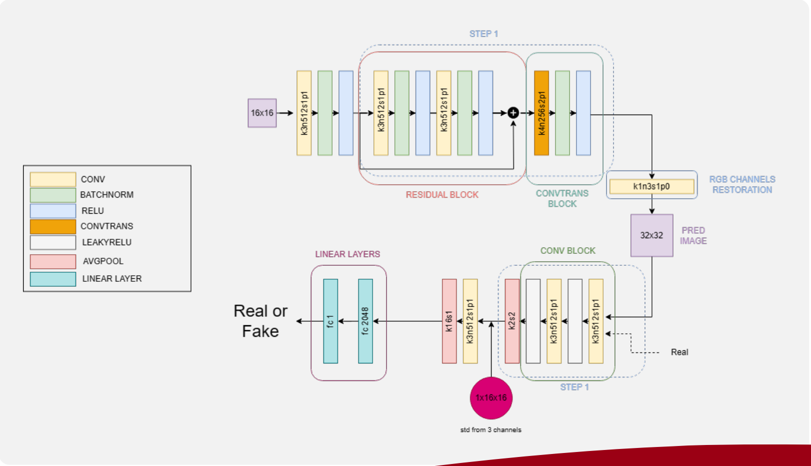


[1] Kim et al., Progressive face super-resolution via attention to facial landmark, 2019

Focus on step 1

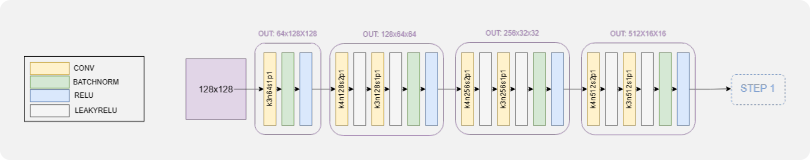
Università degli Studi di Padova

To better understand the progressive architecture, we focus on the blocks in step 1 for both Generator and Discriminator.



Università degli Studi di Padova

To deal with the second objective, 128 × 128 blind image restoration, we incorporate an encoder into the previous model.



The output of the encoder is a tensor of size $512 \times 16 \times 16$ which is passed to step 1 of the progressive GAN architecture.

Progressive training

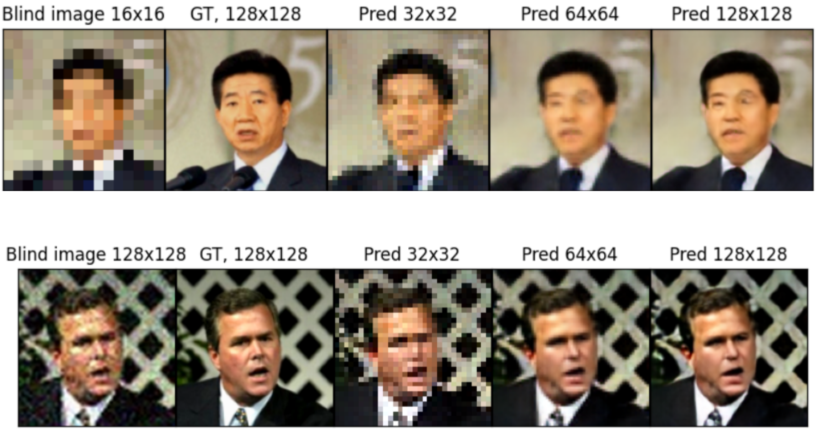
Given an epoch, each batch of the train set:

1. passes through STEP 1 of both generator and discriminator and parameters are updated alternately;
2. passes through STEP 1 and 2 of both generator and discriminator and parameters are updated alternately;
3. passes through the whole architecture and parameters are updated alternately;

In this way the model is trained to restore gradually the final output image with resolution 128x128.

Progressive images visualization

Università degli Studi di Padova



Losses overview

Multiple losses are used in training, aiming at stability and visually pleasing results.

- **Adversarial loss:**

$$L_{adv} = \mathop{\mathbb{E}}_{I_{orig} \sim P_r}[D(I_{orig})] - \mathop{\mathbb{E}}_{I_{blind} \sim P_g}[D(I_{blind})] + \lambda \mathop{\mathbb{E}}_{\hat{I} \sim P_{I_{blind}}}[||\nabla_{\hat{I}} D(\hat{I})_2 - 1||^2] \tag{2}$$

- **Mean-Absolute-Error loss:**

$$L_{pixel} = \frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} \left| (I_{orig})_{x,y} - (\, G(I_{blind})\,)_{x,y} \right| \tag{3}$$

- **Perceptual loss:**

$$L_{percep} = \sum_{i} \frac{1}{W_i H_i} \sum_{x=1}^{W_i} \sum_{y=1}^{H_i} \left| \phi_i(I_{orig})_{x,y} - \phi_i(\, G(I_{blind})\,)_{x,y} \right| \tag{4}$$

FAN-based attention loss

A pre-trained FAN model is employed to locate facial landmarks which are extracted during the pre-processing phase from the original and 2× downsampled images.

- **Attention loss:**

$$L_{atten} = \sum_{\substack{(x_l, y_l),\\ l \in \Lambda}} \left(\left| I_{orig} - G(I_{blind}) \right| * h_{5x5} \right) [x_l, y_l] \tag{5}$$

We apply attention around each landmark's coordinates using a gaussian kernel h, independetly for steps 2 & 3.

Dynamic weighting and overall training loss

Università degli Studi di Padova

The loss strategy must balance between stability and visual performance.

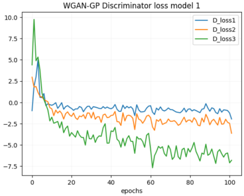
In the primary phase of training the adversarial loss is prioritized, while the weigths of the other losses are a function of the epoch, increasing at most linearly.

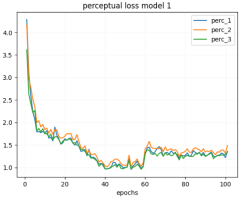
The overall training loss at step k is given by:

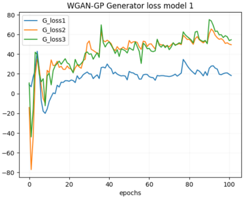
$$L_{train} = L_{adv} + \alpha_k L_{pixel} + \beta_k L_{percep} + \gamma_k L_{atten} \tag{6}$$

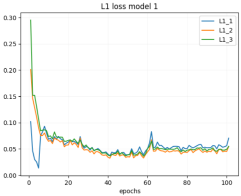
Training convergence for model 1

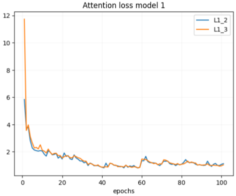
Training for 102 epochs on NVIDIA T4 Tensor Core GPU





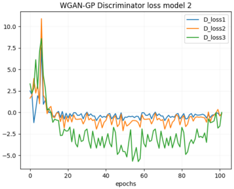


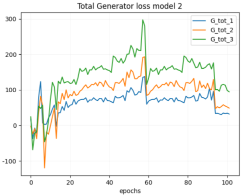


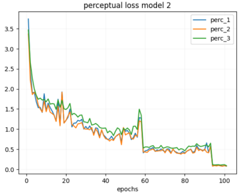


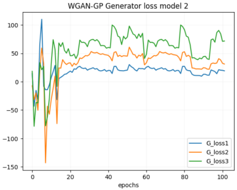
Training convergence for model 2

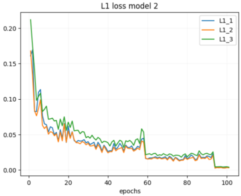
Training for 102 epochs on NVIDIA T4 Tensor Core GPU

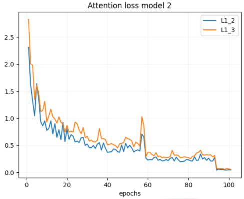












Results for objective 1

The assessment on the test set is shown. While pixel-wise metrics do not exhibit great improvement, visual quality is enhanced.

pred img	PSNR↑	SSIM↑	MS-SSIM↑	LPIPS↓	NIQE↓	FID↓
16x16	20.6 ± 0.3	0.65 ± 0.01	0.862 ± 0.004	0.295 ± 0.006	7.9 ± 0.3	40 ± 2

Table: Batch metrics for generated images (1)

Results for objective 2

We find similar improvements also for objective 2:

pred img	PSNR↑	SSIM↑	MS-SSIM↑	LPIPS↓	NIQE↓	FID↓
128x128	22.0 ± 0.4	0.64 ± 0.01	0.862 ± 0.009	0.31 ± 0.01	6.8 ± 0.2	35 ± 2

Table: Batch metrics for generated images (2)

We investigate some common causes for visually poor generated images.

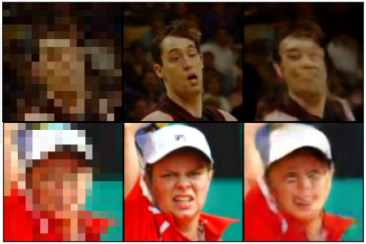
Causes of systematic underperformance

a) Multiple faces

We investigate some common causes for visually poor generated images.

Causes of systematic underperformance

b) Difficult poses and rare expressions



Critical issues

We investigate some common causes for visually poor generated images.

Causes of systematic underperformance

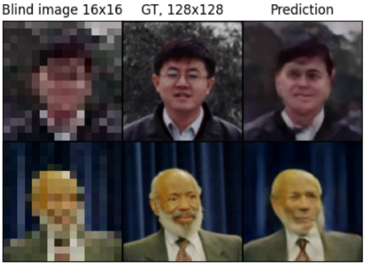
c) Face interference with other objects

Critical issues

We investigate some common causes for visually poor generated images.

Causes of systematic underperformance

d) Ethnic groups underrepresentation



We investigate some common causes for visually poor generated images.

Causes of systematic underperformance

e) Gender assessment

We investigate some common causes for visually poor generated images.

Causes of systematic underperformance

f) Fine details smoothing, e.g. wrinkles

Results combining models

We investigate the possibility of employing both models sequentially. Such possibility could be further researched in the future to enhance small resolution crops inside images with blind-type deterioration.

Conclusions

This work has explored the Blind Face Restoration problem from two different angles.

As demonstrated by the visual metrics, our models improve the photo-realism of the images.

Possible future enhancements:

- a combination of the two models both sequentially and in parallel
- an extended training time, a more robust and precise prior, a larger dataset
- a more complex architecture, i.e. residual blocks in the encoder.