Machine Learning for Facial Image Super-resolution

CHEUNG Tsun Hin, 15083269D@connect.polyu.hk

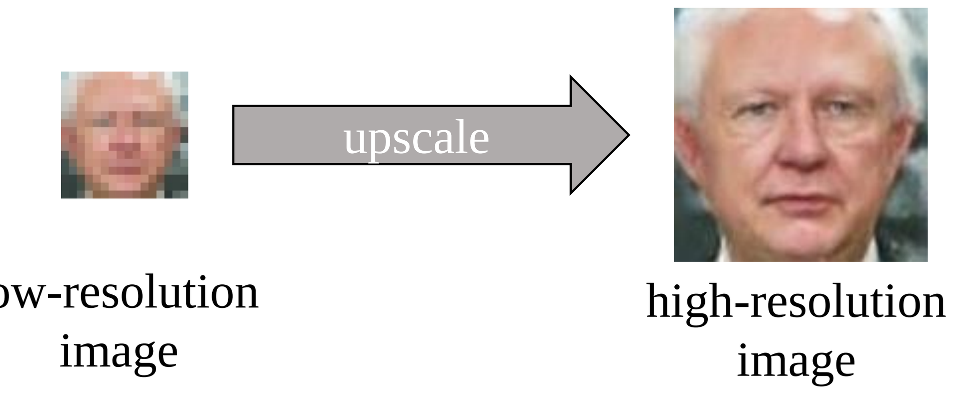
Supervisor: Prof. Kenneth LAM

Department of Electronic and Information Engineering, The Hong Kong Polytechnic University

Introduction

Image Super-resolution

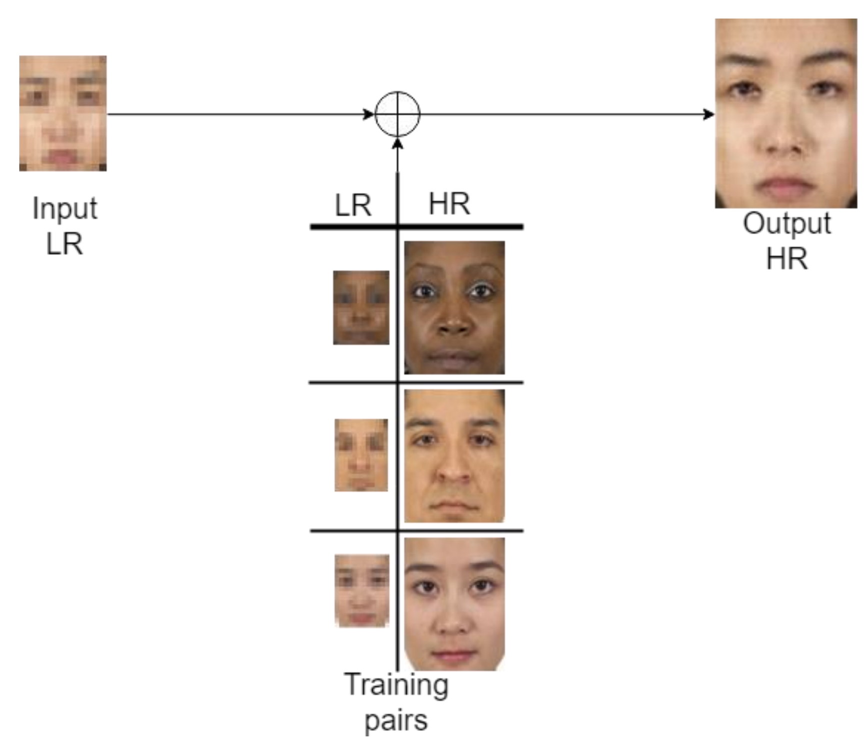
Image super-resolution (SR) is to generate a high resolution (HR) image from a low resolution (LR) one.



- An ill-posed task
- Lost of information during the down-sampling process
- Multiple LR images corresponding to a single HR image

Machine learning approaches

Machine learning method for image super-resolution is to generate a high-resolution image by learning from training pairs of low and high-resolution images.



Problem Formulation

Image degradation process

The low-resolution images $\vec{I}_l$ is generated from its high-resolution images $\vec{I}_h$ by the following linear model:

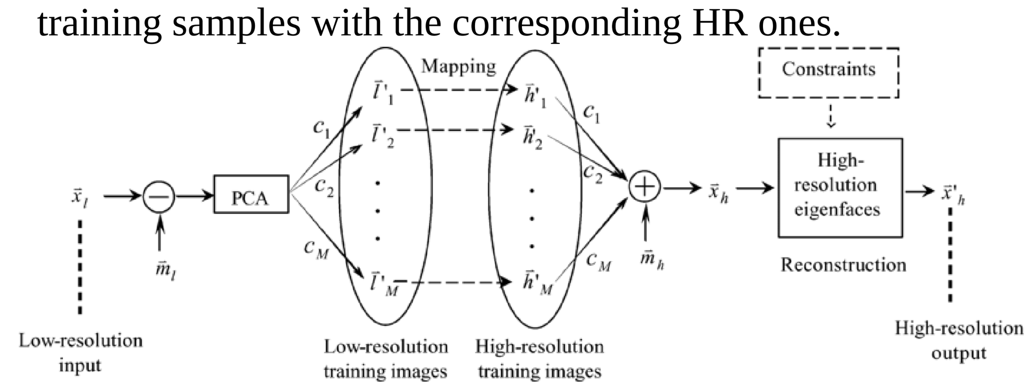
$$\vec{I}_l = H\vec{I}_h + \vec{n}$$

where H is the operation matrix involving blurring and down-sampling and $\vec{n}$ is the random distribution added during image acquisition.

Methodology

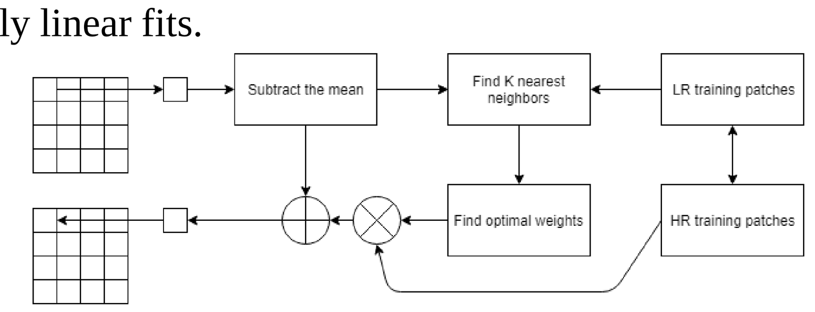
Algorithm 1 – Eigentransformation (PCA)

- Eigentransformation is a transformation based on mapping between LR and HR groups of training samples.
- By Principal Component Analysis (PCA), an input LR image can be represented as a linear combination of LR training samples.
- Keep all the coefficients unchanged, but replace the LR training samples with the corresponding HR ones.



Algorithm 2 – Neighbour Embedding (LLE)

- Locally linear embedding (LLE) is one of the manifold learning methods.
- By neighbourhood-preserving embeddings of high-dimensional inputs, the nonlinear structure is recovered from locally linear fits.



For each input image patch x_t^q, find the optimal weights w_i^q for the k nearest neighbours x_i^q that minimize the reconstruction error ε^q.

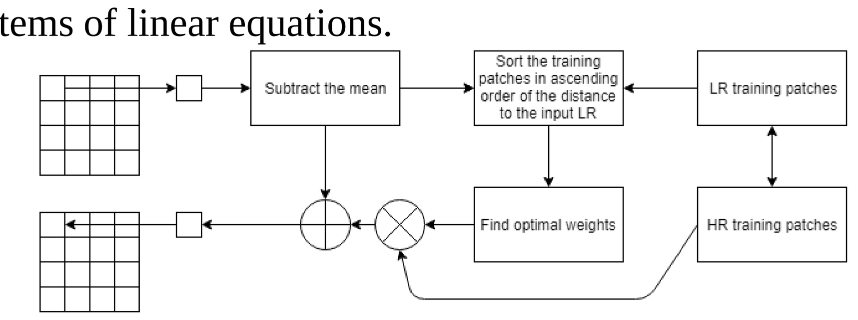
$$\varepsilon^q = \left\| x_t^q - \sum_{i=1}^{k} w_i^q x_i^q \right\|_2^2$$

$$\text{s.t. } \sum_{i=1}^{k} w_i^q = 1$$

Keeping all the weights unchanged, the LR patches are replaced by the corresponding HR ones.

Algorithm 3 – Sparse Representation (SR)

Sparse representation (SR) deals with sparse solutions for systems of linear equations.



For each input image patch x_t^q, find the optimal weights w_i^q for training image patches x_i^q that minimize the reconstruction error ε^q.

$$\varepsilon^q = \left\| x_t^q - \sum_{i=1}^{M} w_i^q x_i^q \right\|_2^2$$

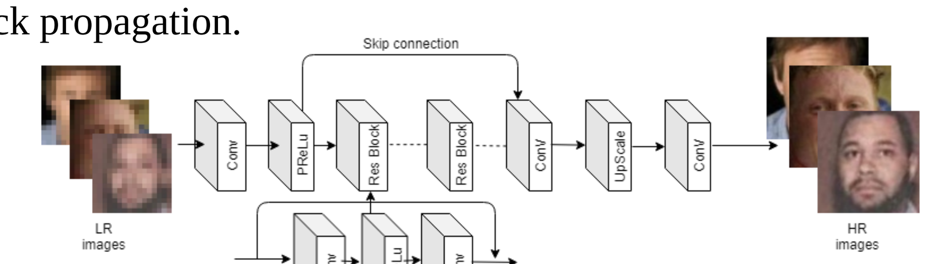
$$\text{s.t. } \sum_{i=1}^{M} w_i^q < \varepsilon_1 \text{ and } \sum_{i=2}^{M} w_i^q - w_{i-1}^q < \varepsilon_2$$

Keeping all the weights unchanged, the LR patches are replaced by the corresponding HR ones.

Methodology (Deep Learning approaches)

Algorithm 4 – Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a deep learning method that learns an end-to-end mapping from LR input to HR output by back propagation.



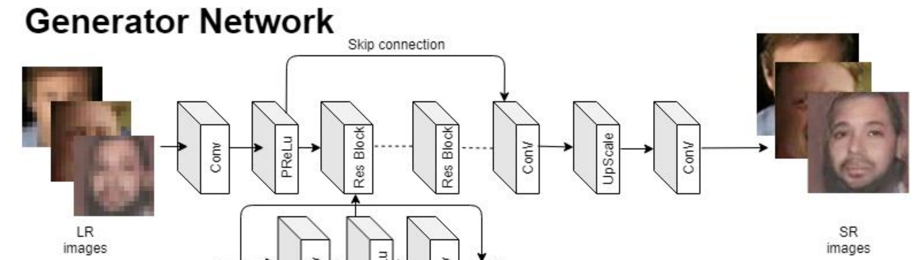
MSE-loss:

$$l_{mse} = \frac{1}{WH}\sum_{x=1}^{W}\sum_{y=1}^{H}\left(I_{x,y}^{HR} - G\left(I_{x,y}^{LR}\right)\right)^2$$

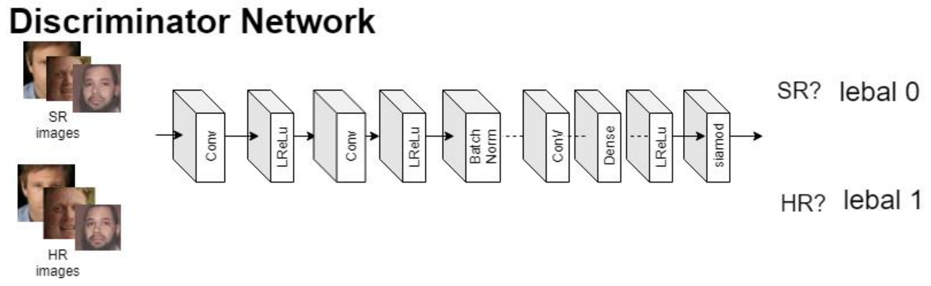
Algorithm 5 – Generative Adversarial Network (GAN)

Generative Adversarial Network (GAN) is another deep learning method that consists of a pair of networks, a generator network and a discriminator network. The generator network fools the discriminator that the generated images are real HR images.

Generator Network



Discriminator Network



Perceptual loss:

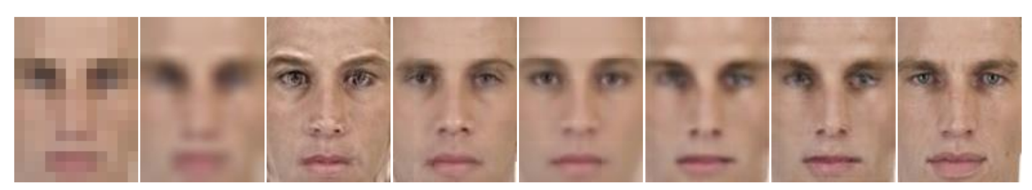
$$l_{total} = l_{vgg} + 0.006\, l_{adversairal}$$

$$l_{vgg} = \frac{1}{WH}\sum_{x=1}^{W}\sum_{y=1}^{H}\left(V\left(I_{x,y}^{HR}\right) - V\left(G\left(I_{x,y}^{LR}\right)\right)\right)^2$$

$$l_{adversairal} = \sum_{n=1}^{N} -\log D\left(G\left(I^{LR}\right)\right)$$

Results

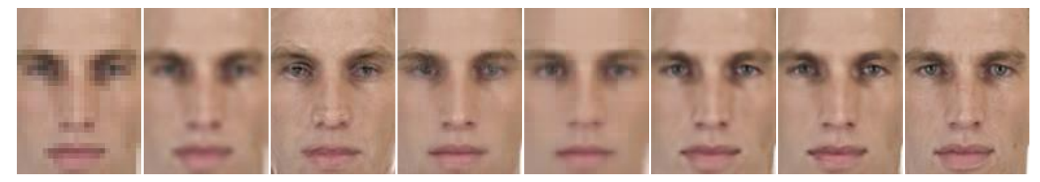
1. Upscaling factor 8



	BC	PCA	LLE	SR	CNN	GAN	HR
PSNR (dB)	24.57	25.33	26.24	26.34	27.96	27.36	INF
SSIM	0.712	0.744	0.784	0.794	0.829	0.803	1

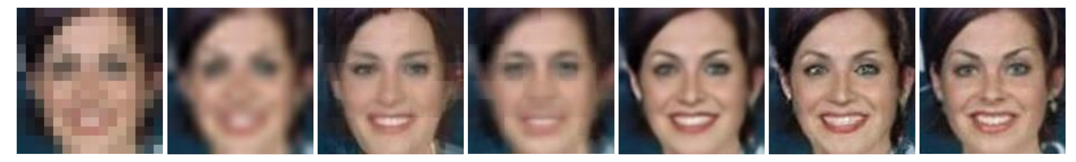
Results

2. Upscaling factor 4



	BC	PCA	LLE	SR	CNN	GAN	HR
PSNR (dB)	28.49	27.96	29.28	29.37	32.07	30.73	INF
SSIM	0.853	0.808	0.856	0.860	0.908	0.890	1

3. Unconstrained dataset



	BC	LLE	SR	CNN	GAN	HR
PSNR (dB)	24.75	24.77	25.22	28.32	27.06	INF
SSIM	0.689	0.691	0.711	0.820	0.773	1

4. Noise Performance

	BC	PCA	LLE	SR	CNN	GAN
PSNR (dB) ($\sigma = 0.01$)	24.51	24.59	25.77	26.10	27.94	27.30
PSNR (dB) ($\sigma = 0.05$)	23.65	24.02	25.18	25.77	25.83	25.56
dB drops	0.86	0.57	0.59	0.33	2.11	1.74

Conclusion

In this project, some machine learning algorithms for facial image super-resolution are implemented and their performances are measured.

The deep learning methods outperform all the conventional machine learning methods in terms of PSNR and SSIM. These methods are efficient because no optimization is needed.

The conventional machine learning methods, such as sparse representation, enjoy the noise-robust characteristics for face hallucination.