HW2·Part1

Xiang Pan 潘翔
1900941520
hover@hust.edu.cn

Q1 The difference between original and reconstruct

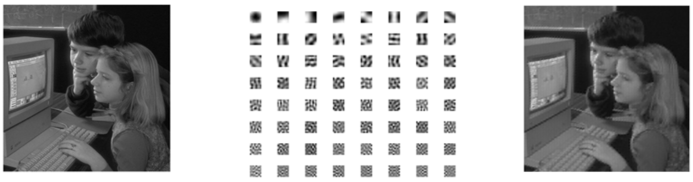

Image 1.1 The original image, patches and the reconstruct picture

Q2 The histogram of principal component's STD

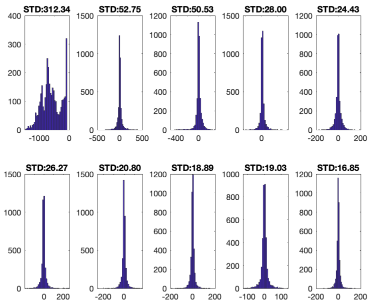

Image 2.1 the hist of 4th principal component

principal component (neurons)'s responses to the images are correlated. And we can observe that the STD's distribution is almost Gaussian distribution, and the larger the eigenvalue is, the more relevant(less STD).

Actually, the result is also a reflection of probabilistic pca(p-pca). A perspective from Bayesian.

$$x = Wz + \varepsilon \text{ ,} \varepsilon \sim N(0, \sigma^2 I)$$

$$f(z;\theta) = Wz + \mu$$

$$p_\theta(z|x = x_i)$$

When we get x, then we can use equation 3 to get z's peak as the result of dimensionality reduction. Then the result is suffering from the Gaussian Noise.

Q3 Number of Principal Components

Image 3.1 Number of Principal Component from 1to 10

As we can see from the process, the distribution of the importance of principal components follows Long-tailed distribution.

HW2·Part2

QC1: PCA & PCoA(change the definition of distance)

We know that the PCA use Euclidean metric as the distance definition to maximize the projection distance and minimize the rebuild cost. While PCoA use other distance definition to reduce the dimension. I try to compare the performance of the two methods.

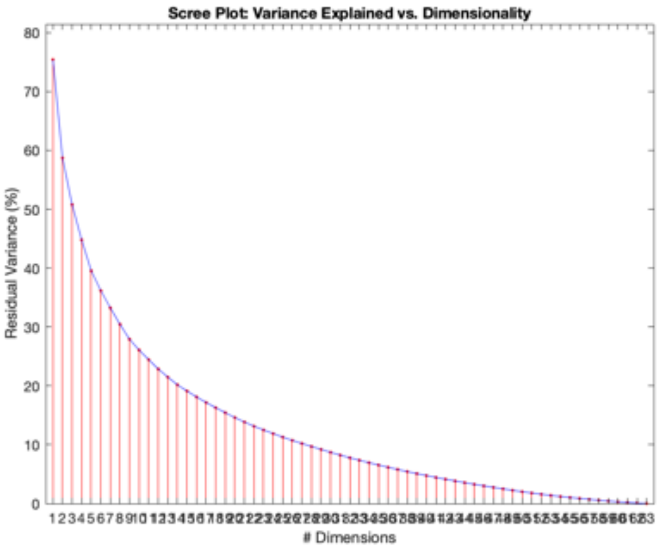

Image C1.1 PCoA Dimention-Risidual curve

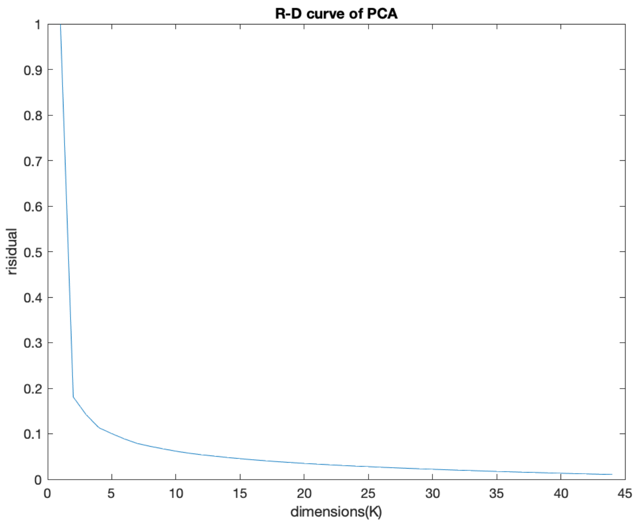

Image C1.1 PCA Dimention-Risidual curve

We can conclude that PCA may be more effective than PCoA on image processing.

QC2: PCA for Colorful Image Dimension Redcution

We can process the image RGB separately.

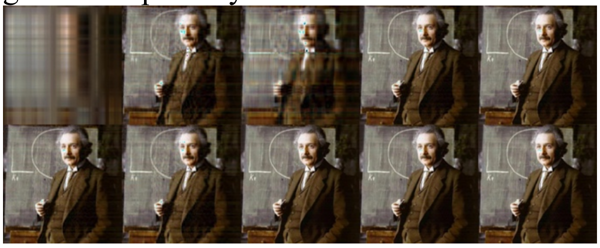

Image C2.1 PCA for Colorful Image Reconstruction from PCNums(1,100,step=10)

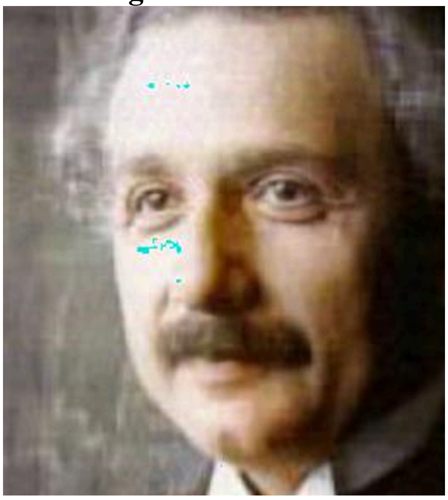

Image C2.2 The reconstruction Error in Image

Actually, the PCA is a non-supervised learning method, and the reconstruction error is relevant to the pixel-level while PCA pay equal attention to each pixel in the picture. There are many methods for solving this problem. For example the Attention Method in VAE to give different weight for each pixel. And we also can use GAN to get a high-level perspective to reduce the low-pixel-loss but insight error.

HW2·Part2

QC3 TCA(transfer component analysis)-PCA thoughts

In transfer learning, we may build the transfer matrix as knowledge transfer function to transfer the NN weights and other models' information. But the transfer matrix may be very big and the amount computation which may be more than train model without the knowledge transfer. Hence, I try to implement the TCA to reduce the dimensions of transfer matrix and maximize the performance.

Process

The process is just like PCA:

1. Target:

$$P(Y_S|\phi(X_S)) \approx P(Y_T|\phi(X_T))$$

2. The target can be formed as finding the function ϕ to minimize

$$dist(X'_{src}, X'_{tar}) = \left\| \frac{1}{n_1}\sum_{i=1}^{n_1}\phi(x_{src_i}) - \frac{1}{n_2}\sum_{i=1}^{n_2}\phi(x_{tar_i}) \right\|_{\mathcal{H}}$$

3. The Kernel Method

$$K = \begin{bmatrix} K_{src,src} & K_{src,tar} \\ K_{tar,src} & K_{tar,tar} \end{bmatrix}$$

$$\widetilde{K} = (KK^{-1/2}\widetilde{W})(\widetilde{W}^{\top}K^{-1/2}K) = KWW^{\top}K$$

4. The definition of distance

$$L_{ij} = \begin{cases} \frac{1}{{n_1}^2} & x_i, x_j \in X_{src}, \\ \frac{1}{{n_2}^2} & x_i, x_j \in X_{tar}, \\ -\frac{1}{n_1 n_2} & \text{otherwise} \end{cases}$$

5. The optimization problem can be formulated as:

$$trace(KL) - \lambda trace(K)$$

$$\begin{aligned} \min_{W} \quad & \operatorname{tr}(W^{\top}KLKW) + \mu \operatorname{tr}(W^{\top}W) \\ \text{s.t.} \quad & W^{\top}KHKW = I_m \end{aligned}$$

Data:

Source: Amazon

Target: Webcam

Results:

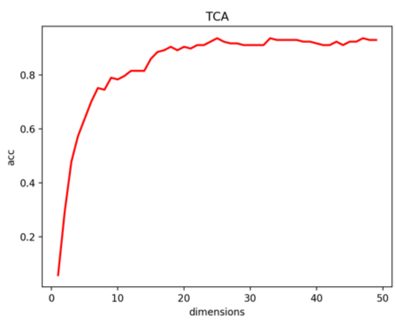

Image C3.1 Transfer Learning with TCA method

Reference

[1] Pan S J, Tsang I W, Kwok J T, et al. Domain adaptation via transfer component analysis[J]. IEEE Transactions on Neural Networks, 2011, 22(2): 199-210.