ECS795P Deep Learning and Computer Vision, 2022

Course Work 1: Image Super-resolution Using Deep Learning

1. Suppose the settings of a SRCNN as: f1=9, f2=3, f3=5, how many pixels of the low-resolution image are utilized to reconstruct a pixel of the high-resolution image with the SRCNN? (10% of CW1)
   The computation for estimation of a high resolution pixel utilizes the information is as shown below:
   $= (f1 + f2 - 1 + f3 - 1)^2$
   $= (9 + 3 - 1 + 5 - 1)^2$
   = **225**

2. Why the deep convolutional neural network is superior to perform image super-resolution? Give one reason to explain it. (10% of CW1)
   Deep convolution modelling, also known as Super resolution convolution neural network modelling (SRCNN), provides end-to-end mapping between low and high resolution images. This is possible because of the hidden CNN layers traditional example-based methods. Patch extraction and aggregation are performed entirely within the neural network layers, requiring less pre and post processing of the input and output images and working relatively quickly. SRCNN works well for multiple channels, has good performance metrics, and is lightweight. This is why deep convolution models outperform other models in terms of performance.

3. Please explain the meaning of peak signal-to-noise ratio (PSNR) in the context of image super-resolution. PS: give the ground truth (GT) image, and the high-resolution images by SCRNN (HR-SRCNN) and interpolation (HR-Base) for reference. Also put the PSNR value below the high-resolution images. (10% of CW1)
   Peak signal-to-noise ratio (PSNR) is an engineering term that refers to the ratio of a signal's maximum possible power to the power of corrupting noise that affects the fidelity of its representation. PSNR is usually expressed as a logarithmic quantity using the decibel scale because many signals have a very wide dynamic range.

   PSNR is commonly used to quantify reconstruction quality for lossy-compressed images and videos. PSNR is most easily defined via the mean squared error

(MSE). Given a noise-free m×n monochrome image I and its noisy approximation K, MSE is defined as below:

$$MSE = \frac{1}{m\,n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2.$$

The PSNR (in dB) is defined as

$$\begin{aligned} PSNR &= 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \\ &= 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \\ &= 20 \cdot \log_{10}(MAX_I) - 10 \cdot \log_{10}(MSE). \end{aligned}$$

Note: MAXI is the maximum possible pixel value of the image

GT

HR-Base (PSNR=20.497630181368823)

HR-SRCNN (PSNR=22.92269645238934)

References:

https://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio

https://medium.com/coinmonks/review-srcnn-super-resolution-3cb3a4f67a7c

https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/#:~:text=In%20deep%20learning%2C%20a%20convolutional,a%20special%20technique%20called%20Convolution.