ECS795P Deep Learning and Computer Vision, 2020

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)

   225 pixels of the low-resolution image are utilized to reconstruct a pixel of the high-resolution image. For each pixel of the high-resolution (output) image, it is constructed by 5×5 filter pixels on the layer 2; for each pixel of the layer 2, it is constructed by 3×3 filter pixels on the layer 1; for each pixel of the layer 1, it is constructed by 9×9 filter pixels on the low-resolution (input) image. Therefore, a pixel of the high-resolution image is reconstructed using pixels of 15×15 (225) filter on the low-resolution image.

2. Why the deep convolutional model is superior to perform image super-resolution? Give one reason to explain it. (10%of CW1)

   a). Unlike traditional methods that handle each component separately, deep convolutional model jointly optimizes all layers. Not all operations have been considered in the optimization in the traditional sparse-coding-based SR methods. On the contrary, in SRCNN, the low-resolution dictionary, high-resolution dictionary, non-linear mapping, together with mean subtraction and averaging, are all involved in the filters to be optimized. That is why deep convolutional model is considered as an end-to-end mapping optimization that consists of all operations.
   b). The information exploited for reconstruction of high-resolution image in SRCNN is comparatively larger than that used in external example-based (traditional) approaches.

3. Please explain the physical meaning of peak signal-to-noise ratio (PSNR) in the context of image super-resolution. PS: place here the ground truth (GT) image, and the high-resolution images by SCRNN (HR-SRCNN) and bicubic interpolation (HR-BI) for reference. Also put the PSNR value below the high-resolution images. (10% of CW1)

GT

HR-BI (PSNR=20.497630173285614)

HR-SRCNN (PSNR=20.74699270372715)

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(I(i) - \hat{I}(i)\right)^2 \quad (1)$$

$$PSNR = 10 \cdot log_{10}\left(\frac{L^2}{MSE}\right) \quad (2)$$

Where L denotes the maximum possible pixel value, MSE is the mean squared error between the ground truth image I and the reconstructed image $\hat{I}$, and both of the two images are with N pixels.

In general cases using 8-bit image representation, L equals to 255 and the typical values for the PSNR vary from 20 to 40, where higher is better. When L is fixed, the PSNR is only related to the pixel-level MSE between images, only caring about the difference between the pixel values at the same positions instead of human visual perception (i.e., how realistic the image looks). This leads to PSNR's poor performance in representing the quality of the super-resolved images in real scenes, in which cases we are usually more concerned with human perception.