5 Implementation of Mixed-scale Dense Network and Experimentation Analysis

5.1 Training and Testing Dataset Splitting

There are 1000 input and output pairs for each dataset. I distribute 800 of the 1000 pairs for training purpose, 100 of them for validation purpose during the training, and the rest of 100 will be used for testing purpose.

5.2 Evaluation metrics

Since we are solving the image processing problem using a regression model, the final output images (attenuation and phase projection reconstructions) will be evaluated against expected output images (original attenuation and phase projections) in testing. The evaluation metrics are mean square error, normalized mean square error, similarity index measurement and peak signal-to-noise ratio, the most important error to be considered is normalized mean square error. We do the same for regression results by contrast transfer function for comparison.

5.2.1 Mean square error

The mean square error is normally used for a regression problem, it is normally used to measure the distance between two data sets in Euclidean space with norm of 2. It can be expressed as:

$$MSE(x,y) = \frac{\sum_{i=1}^{N}(x_i - y_i)^2}{N} \tag{5.1}$$

5.2.2 Normalized mean square error (NMSE)

The normalized mean square error can be expressed as :

$$NMSE(f(x), g(x)) = 100 \times (\frac{\sum |f(x) - g(x)|^2}{\sum |f(x)|^2})^{1/2} \tag{5.2}$$

a percentage, is used in this case, where $f(x)$ is the ideal image and $g(x)$ is the reconstructed image. It eliminates the influence of the level of magnitude.

5.2.3 Similarity index measurement(SSIM)

The full name of SSIM is structural similarity index measurement, which is an index of the structural similarity between two images. This indicator was first proposed by the Image and Video Engineering Laboratory (Image and Video Engineering Laboratory) at the University of Texas at Austin. The initial SSIM algorithm was used to evaluate the quality of the image compression.

Normally the natural images have structural similarity, and there is a strong correlation between the pixels of the image, especially in the case of spatial similarity. These correlations carry important information about the structure of objects in the visual scene. We assume that the human visual system (HSV) mainly obtains structural information from the visible area. Therefore, by detecting whether the structural information changes, the approximate information of image distortion is perceived.

Some quality evaluation metrics based on error sensitivity (such as MSE, PSNR) use linear transformation to decompose the image signal, which does not involve correlation. The SSIM we want to discuss is used to find a more direct way to compare the structure of the distorted image and of the reference image.

Here x denotes the reference image and y denotes the processed image, the SSIM can be expressed as :

$$SSIM(x,y) = [l(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma} \tag{5.3}$$

where we define three functions $l(x,y)$ for luminance, $c(x,y)$ for contrast and $s(x,y)$ for structure:

$$l(x,y)=\frac{2\mu_x\mu_y+C_1}{\mu_x^2\mu_y^2+C_1} \tag{5.4}$$

here constant C_1 is used to avoid the zeros of the denominator caused by $\mu_x^2\mu_y^2\approx 0$

$$C_1=(K_1L)^2 \tag{5.5}$$

L represents number of gray levels, $L=255$ for 8-bit gray level image; $K_1=0.01$ by default.

and

$$c(x,y)=\frac{2\sigma_x\sigma_y+C_1}{\sigma_x^2\sigma_y^2+C_1} \tag{5.6}$$

here constant C_2 is used to avoid the zeros of the denominator caused by $\sigma_x^2\sigma_y^2\approx 0$

$$C_2=(K_2L)^2 \tag{5.7}$$

L represents number of gray levels, $L=255$ for 8-bit gray level image; $K_1=0.03$ by default.

and

$$s(x,y)=\frac{\sigma_{x_y}+C_3}{\sigma_x\sigma_y+C_3} \tag{5.8}$$

here constant C_3 is used to avoid zeros of denominator caused by $\sigma_x\sigma_y\approx 0$. Normally we set $C_3=\frac{C_2}{2}$, $\alpha=\beta=\gamma=1$.

Let μ represents the average gray level of the image, it can be expressed as:

$$\mu_x=\frac{\sum_{i=1}^{N}x_i}{N} \tag{5.9}$$

and let σ be the standard deviation of the image, which represents the contrast of image, it can be expressed as:

$$\sigma_x=(\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)^2)^2 \tag{5.10}$$

and σ_{x_y} be the co-variance between two image data, it can be expressed as:

$$\sigma_{x_y}=(\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y)) \tag{5.11}$$

In summary, SSIM can be expressed as:

$$SSIM(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_x\sigma_y+C2)}{({\mu_x}^2+{\mu_y}^2+C1)({\sigma_x}^2+{\sigma_y}^2+C2)} \tag{5.12}$$

5.2.4 Peak signal-to-noise ratio(PSNR)

Peak signal-to-noise ratio (often abbreviated as PSNR) is an engineering term that represents the ratio of the maximum possible power of a signal to the destructive noise power that affects its representation accuracy. Since many signals have a very wide dynamic range, the peak signal-to-noise ratio is usually expressed in logarithmic decibel units.

$$PSNR(x,y)=10\log_{10}(\frac{(MAX_I)^2}{MSE(x,y)})=20\log_{10}(\frac{MAX_I}{\sqrt{MSE(x,y)}}) \tag{5.13}$$

where MAX_I is the maximum value that represents the color of an image point. If each sample point is represented by 8-bit gray level, $MAX_I=255$.

Therefore, the smaller the MSE, the larger the PSNR; and the larger PSNR value means better image quality.

Generally speaking, a PSNR value higher than 40dB means that the image quality is very good (ie very close to the original image), value of 30-40dB usually indicates that the image quality is good (that means the distortion is perceptible but acceptable). It indicates that the image quality is poor at 20-30dB. Finally, images with PSNR below 20dB are unacceptable