Super-Resolution

Super-Resolution in image processing means upsampling and therefore interpolation between pixels of an image. It can be interpreted as the opposite of downsampling. To make images larger in the image dimensions it is necessary to predict the values of the additional pixels between the original pixels. One of the easiest ways and also a traditional method to do this is applying a bicubic interpolation. New methods have evolved in the recent years and the use of neural networks is outperforming all other methods developed so far.

**Visual Example for better Understanding**

![Figure 1: Low resolution image.](image1.png) ![Figure 2: Interpolated low resolution image.](image2.png) ![Figure 3: High resolution image](image3.png)

Figure 1 shows a picture with 25x25 pixels, whereas on the upper right you can see the same image in original resolution 100x100 pixels. Figure 2 shows the upscaled left image with bicubic interpolation. One can see that the image is blurred compared to the original resolution image on the right (cf. figure 3). This effect is caused by incorrect prediction of the new pixel values. The aim of a Super-Resolution neural network is learning the missing pixel values for the upscaled image as good as possible.

**Metrics**

In order to describe the quality of the upscaling method it is necessary to define a metric which describes the similarity between the predicted (upscaled) image and the ground truth (full resolution) image. In this section, some of the commonly used metrics, that are employed for problems of this nature, are described.

**PSNR**

Peak Signal to Noise Ratio (PSNR) is a commonly used metric to define the similarity between two images. It is calculated using the Mean-Square-Error (MSE) of the pixels and the maximum possible pixel value (MAX) as follows:

$$PSNR = 10 \cdot \log\left(\frac{MAX^2}{MSE}\right)$$
A high PSNR value corresponds to a high similarity between two images and a low value corresponds to a low similarity respectively. (4)

**SSIM**

The structural similarity index is developed in order to improve traditional methods such as PSNR, which have been proven to be inconsistent with human visual perception. It takes luminance, contrast and structure of both images into account.

The SSIM index is calculated on various windows of an image. The measure between two windows \( \mathbf{x} \) and \( \mathbf{y} \) of common size \( N \times N \) is:

\[
\text{SSIM}(x, y) = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}
\]

with:

- \( \mu_x \) the average of \( \mathbf{x} \);
- \( \mu_y \) the average of \( \mathbf{y} \);
- \( \sigma_x^2 \) the variance of \( \mathbf{x} \);
- \( \sigma_y^2 \) the variance of \( \mathbf{y} \);
- \( \sigma_{xy} \) the covariance of \( \mathbf{x} \) and \( \mathbf{y} \);
- \( c_1 = (k_1L)^3 \) and \( c_2 = (k_2L)^3 \) two variables to stabilize the division with weak denominator;
- \( L \) the dynamic range of the pixel-values (typically this is \( 255 \) or \( 0 \) to \( 1 \));

\[ k_1 = 0.01, \quad k_2 = 0.03 \]

by default. (3)

More metrics can be found in literature:

- IFC (Information Fidelity Criterion)
- NQM
- WPSNR (Weighted Peak Signal to Noise Ratio)
- MSSSIM (Multi Scale Structural Similarity)

**Classical Approaches**

Three classical approaches are briefly described below:

**Bicubic Interpolation**

Bicubic Interpolation considers 16 surrounding pixels to predict new pixel values.

**Bilinear Interpolation**

Bilinear Interpolation considers 4 surrounding Pixels to predict new pixel values.

**Nearest Neighbor**

The Nearest Neigboor method simply predicts the pixel values from the value of the nearest neighbor pixel.
Figure 4: Architecture of SRCNN. Three convolutional layers in total. (2)

SRCNN consists of 3 convolutional layers with filter sizes 9x9 for the first layer, 1x1 for the second layer and 3x3 for the last layer. The first layer generates 64 feature maps, the second 32 and the last one generates the output. Figure 4 shows color images, but in practice only grayscale images are used to train and apply the network. The first layer filters can be interpreted as feature detectors, such as corners, lines, etc. They are visualized in the figure 5.
VDSR (Very Deep Super Resolution)

The VDSR network consists of 20 convolutional layers. The input and output image share the same size. This is achieved by padding with zeros in every convolution. The key element here is the residual learning, which is applied by adding the input image to the output from the last convolutional layer. In this way only the difference between low and high resolution is learned by the network. It makes sense because both images are sharing the same low frequencies and thus do not need to be considered in the training process.

The filter size of each convolution except the first one is 3x3x64. The receptive field of the network is therefore 41x41 pixels. Each convolution except the last one generates 64 feature maps, some of which visualized in the graphic above. Data augmentation via rotation and flipping is used for training. In order to gain speed and reduce the size of the network the training data is decomposed into patches with size 41x41. This also helps to increase the amount of training data. Out of 291 images approximately 140,000 Patches can be generated.

In order to aid the network to converge, gradient clipping and L2-regularization is used. The learning rate is decreased every 20 epochs by a factor of 0.1, which improves performance.

Comparison of different Methods and State of the Art Performance

The table below shows a few methods of super resolution approaches. The datasets can be found as standard in today’s literature. All networks are trained with Set291, a set of images containing 291 natural images.
Applications

Image Super-Resolution is used in many areas such as:

- Surveillance
- Remote Sensing
- Medical Imaging (e.g., ultrasonic images, x-ray-images)
- Video Standard Conversion (i.e., SD to HD)
- Photocameras (i.e., postprocessing of images)
- Printing (i.e., enhance print quality on paper of low resolution images)
- Biometrics (i.e., fingerprint/face recognition)
- Commercial (barcode reading)
- Military (tracking and detecting)
- Satellite Imaging (i.e., weather forecasting)
Literature


Weblinks

https://github.com/huangzehao/caffe-vdsr (Implementation of VDSR in Caffe)