

Urban Super-Resolution of Aerial and Satellite Imagery

Resolution Reconstruction of Urban Images (Satellite Images) using AI

It is said that “A picture is worth a thousand words”. Pictures provide a very efficient way of communicating visual information and help us clearly visualize the scenarios. The scientific field of giving the ability to computers, to understand these images or videos is termed as [Computer Vision \(CV\)](#). Computers can understand these images in terms of color values at a pixel level, and the features of the contents. These features can be colors, shapes or sizes, etc. However, dealing with large images is computationally heavy for even an advanced computer. Performing Digital Image Processing algorithms consumes a lot of time and RAM utilization. But with the advancements in high-performance computing technologies and running the Image Processing algorithms on GPUs has further reduced the processing time. The CV has now become a popular topic of academic and industrial research applications. However, the main huddle in this space of CV is the image quality itself. The quality of the image is dependent on the sensors which capture these images, and sensors can be very expensive if extremely high image quality is expected.

With Artificial Intelligence (AI) it is now possible to improve the image quality up to a certain extent, without much loss of data. AI has made significant progress in the processing of large images, using which one can add contents, edit the images, modify the image completely, or even only extract some meta-data from the image.

In remote sensing applications, AI is playing a huge role in enabling automation of analysis of large satellite images/data, which can be used to provide a very detailed insight about a particular location.

Despite many advancements in CV domain and AI applications, there is a major huddle which needs to be effectively crossed for better performance of the automated analysis algorithms, which is the image quality that is produced by various sensors. The quality of images produced by sensors is limited, mostly due to sensor hardware limitations which cannot be much experimented with, as the sensor hardware is a costly piece in the system. Hence “Image Enhancements” are very important as a primary preparation step before analysis of the images. “Image enhancement” is the procedure of improving the quality and the information content of the original image. One such important type of enhancement is resolution improvement, especially in the field of [Satellite Imagery](#).

In satellite images, the image resolution denotes the distance on the ground that is captured by each pixel. Hence it is also known as spatial resolution, or in mathematical terms, it is also called “Ground Sampling Distance” or GSD. In other words, the spatial

resolution signifies how clearly an object can be seen in the satellite image. Higher the resolution (i.e. lower GSD) better will be the details of the contents in the image. A GSD of 0.55m signifies that each pixel in the digital image will show a color value which is the colors on the ground presence in a square box of size 0.55m. To quote Wikipedia, the GSD is defined as “Ground Sample Distance (GSD) in a digital photo (such as an orthophoto) of the ground from air or space is the distance between pixel centers measured on the ground.” In automated remote sensing applications, the GSD plays a very important role as it is the limiting factor that decides how well an analysis can be automated.

AI brings new methods using which the spatial resolution can be improved for an image that originally has low resolution. This is achieved using deep learning techniques, where a Deep Convolutional Neural Network (CNN) is trained on images with pairs of high resolution and corresponding low-resolution image. High-resolution images are resized by a certain factor, so as to create low-resolution images. This method generally gives an image obtained by bilinear or bicubic interpolation. The pairs of high resolution and low-resolution images are fed into the deep learning system where CNN learns to predict the high-resolution image from its paired low-resolution image. The training is continued on a large data set so that the CNN model achieves enough maturity to correctly predict the high-resolution image. Once the CNN is trained sufficiently, the model is further used to predict on images with low resolutions.

There are various approaches suggested by the computer vision research community, and each approach has its own set of pros and cons. Some utilize only fully connected deep convolutional layers while some suggest a non-traditional approach of Generative Adversarial Networks (GAN), while some even suggest a combination of interpolation and deep learning networks.

Our research on comparisons of various models suggests that GANs perform much better than traditional approaches. GAN is a class of machine learning systems in which two neural networks contesting with each other in a game (in the sense of game theory, often but not always in the form of a zero-sum game). Once provided training data, this methodology becomes proficient in new data generation with the same statistical information as the novel training dataset. For instance, once GAN is trained on photographs, it can generate superficially authentic novel photographs having multiple realistic characteristics indistinguishable to human observers. While, in terms of data distribution, the generative network addresses candidate generation by assimilating to map from a latent space to a data distribution of interest, the discriminative network takes care of evaluating for the same by distinguishing candidates obtained by the generative network's generator. The main objective of the training is to increase the discriminative network's error rate.

For creating such a solution, we used a high-resolution satellite imagery data-set which contained images from various satellites and thus had different resolutions (i.e. different GSD). We further filtered out only those images which were within the GSD range of 0.1m to 0.2m. This created our high-resolution data-set. Also, using bicubic

interpolation, we shrunk down the size of these images by a factor of 4. Thus, if the original high-resolution image is (1000 x 1000) pixels, the resized image will be of (250 x 250) pixels. The obtained down-sampled images is the new low-resolution dataset of resolution (or GSD) (0.1 x 4) – (0.2 x 4).

We carried out training of a custom build GAN model using these datasets and used MSE (Mean Square Error) as our evaluation metric (i.e. Loss Function). After sufficient training, the MSE of our model dropped to the order of $10E-2$ on our testing dataset, we terminated the training.

The Resolution Reconstruction solution is now being utilized for many for our object detection solutions, as a pre-processing step, which has lead to an increase in detection performance by up to 20%.