Super-Resolution landscape

Ajinkya Indulkar
UMass Amherst
140 Governors Drive
Amherst, MA 01003
aindulkar@cs.umass.edu

Sarvesh Upadhyay
UMass Amherst
140 Governors Drive
Amherst, MA 01003
supadhyay@cs.umass.edu

Abstract

Super-resolution refers to the task of increasing the dimensions of an image from a Low resolution image to a High resolution image. This problem has been an active research problem for the past couple of decades. Earlier techniques used interpolation in the form of bicubic interpolation, bilinear interpolation etc. Recently numerous techniques based on Deep Learning have come up. People have used CNNs, GANs and Autoencoders to do image superresolution with convincing results. Recently researchers used randomly-initialized deep neural nets as priors to achieve promising results. In this project, we explore different superresolution techniques and compare their performance.

1. Introduction

Super-resolution task refers to converting a Low resolution image to a High resolution image using upscaling. It is a standard inverse problem and is comparable to denoising in that sense. Super-resolution is ill-posed because there are multiple High resolution images that map to a single Low resolution image.

Traditionally, super-resolution has been performed by using standard interpolation techniques such as Bicubic Interpolation [3]. Such techniques have a downside that the upscaling performed by them is general to all images and is not dataset specific. Because of that, it fails to preserve the original statistics of the image. Additionally, the upscaled images look blurry and have little aesthetic value to humans. Newer Deep Learning based techniques result in aesthetically better looking images with sharp edges and better pixel signal to noise ratio.

Dont et al. [2] introduced a Superresolution Convolutional Neural Network that learns a mapping from a LR image to a HR image using a Deep Convolutional Neural Network. One downside of this paper was that they used traditional upscaling methods such as bicubic interpolation in the prediction step. This results in artificially smooth images and result in unnatural results. Newer techniques use a transposed convolution or a deconvolution to learn a filter that maps from a downscaled image to a upscaled image.

Some recent techniques [4] used Generative Adversarial Network to learn an Adversarial network to do superresolution of low-res images.

Another technique proposed by Ulyanov et al. [6] used image priors or regularizers to learn image superresolution as opposed to network weights.

As it is apparent that there are numerous techniques in the current superresolution landscape that use different ideas for superresolultion, we compare different modern Deep Learning based techniques to some traditional techniques such as interpolation.

2. Problem Statement

In this project we concentrate on the problem of super-resolution. It refers to the task of increasing the resolution of an image from a Low Resolution (LR) image to a High Resolution image (HR).

To solve the problem of super-resolution, we are using BSD 100 image dataset [5]. This dataset contains three subsets corresponding to low-resolution and high-resolution image pairs for 2x, 3x and 4x super resolution factors respectively. All the images have been cropped accordingly to accommodate for super resolution factor. This avoids misalignment of the ground truth high resolution image and super-resolved images.

In order to quantitatively, measure the performance of the model we are using PSNR (Peak Signal to Noise Ratio) score whcih is defined as:

$$PSNR = 20\log_{10}(255) - 10\log_{10}(MSE)$$

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}(I(i,j) - F(i,j))^2$$

where $F(i,j)$ is the upscaled image and $I(i,j)$ is the ground truth high resolution image. Moreover, during calculation of PSNR we convert color image to YCrCb color space to take into account the different level of eye sensitivity for different image color channels.

There have been numerous super-resolution methods proposed in the past few years. In this project, we focus on some of the recent techniques:

Laplacian Pyramid Super-Resolution Network (LapSRN). LapSRN tries to fix a few drawbacks with traditional CNN-based techniques. Traditional methods like SRCNN[2] use interpolation for the upscaling step before prediction. This results in blurry images which don't look natural. Furthermore, they cannot learn complex mappings. LapSRN fixes this by using deconvolutions instead of traditional interpolation based mehtods. This allows the network to learn reverse filters that are better at upscaling. Secondly, old methods use ℓ_2 loss which results in blurry predictions because in reality a single Low Res patch can be mapped to many different High Res patches which the ℓ_2 loss cannot model. LapSRN fixes this by using a different loss function named Charbonnier loss function that doesn't suffer from this problem. Thirdly, older methods only do a single super-resolution step e.g. 8x. Such large factors are difficult to map in one step. One solution would be to do multiple upscaling steps but that is expensive to do. LapSRN fixes this by doing multiple SE prediction in a single forward pass.

Super Resolution Generative Adversarial Network (SRGAN). Ledig et al. [4] proposed a GAN for 4X super-resolution of single images. It was one of the first successful techniques using GANs for super-resolution. The result was sharp high res images. Older CNN based methods were able to get images with a low PSNR but resulted in relatively smooth images which were aesthetically unsatisfying.

Deep Image Prior. This paper by Ulyanov et al. [6] takes a very different approach to doing SR. Instead of learning a distribution that maps from LR images to HR images, they focus on the structure of the network to encode information about the statistics of an image.

3. Technical Approach

As described in the last approach, we plan on using three different approaches to solve the problem of Super-resolution. We will be doing 4X super-resolution for all approaches.

3.1. Baseline

We use simple bicubic interpolation as a baseline. Although, this baseline will just serve as a comparison for older methods and is not representative of today's landscape. Any recent model will surely achieve better PSNR.

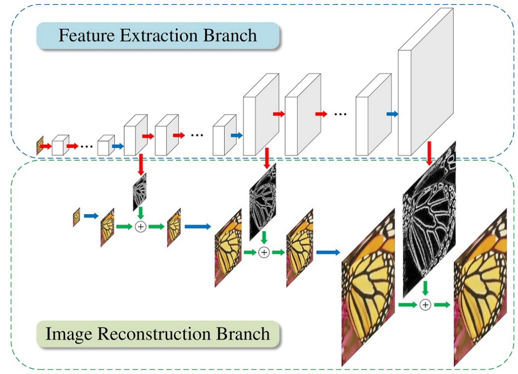

Figure 1. LapSRN architecture. Red arrows=Conv, Blue arrows=Deconv, Green=Elemwise Sum.

3.2. LapSRN

LapSRN architecture is shown in Figure 1. It consists of two branches.

Feature extraction Each level s consists of d conv layers and one deconv layer. The output of this deconv layer is used as input to the next level and is also used to construct a residual image for the Image reconstruction branch. Each level learns features at a finer level.

Image reconstruction. In each level, the input image is upscaled by 2X with a deconv layer. This is then summed elementwise to the residual image from the feature extraction branch.

We also use the robust Charbonnier Loss function. [1]

3.3. SRGAN

The SRGAN network architecture is shown in Figure 2. It is fairly self explanatory and is described in detail in the original SRGAN paper [4].

Our aim is to estimate I^{SR} from a low res I^{LR}. We have access to I^{HR} during training. We want to solve:

$$\hat{\theta_G} = arg \min_{\theta_G} \frac{1}{N} \sum_{n=1}^{N} l^{SR}(G_{\theta_G}(I_n^{LR}), I_n^{HR})$$

A GAN consists of a discriminator D_{θ_D} and a Generator G_{θ_G} and we want to find the solution the minmax problem:

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{I^{HR} \sim p_{\text{train}}(I^{HR})}[\log D_{\theta_D}(I^{HR})] + \mathbb{E}_{I^{LR} \sim p_G(I^{LR})}[\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR}))]$$

The loss function used here l^{SR} is called the Perceptual loss function and is described in detail in the SRGAN paper [4].

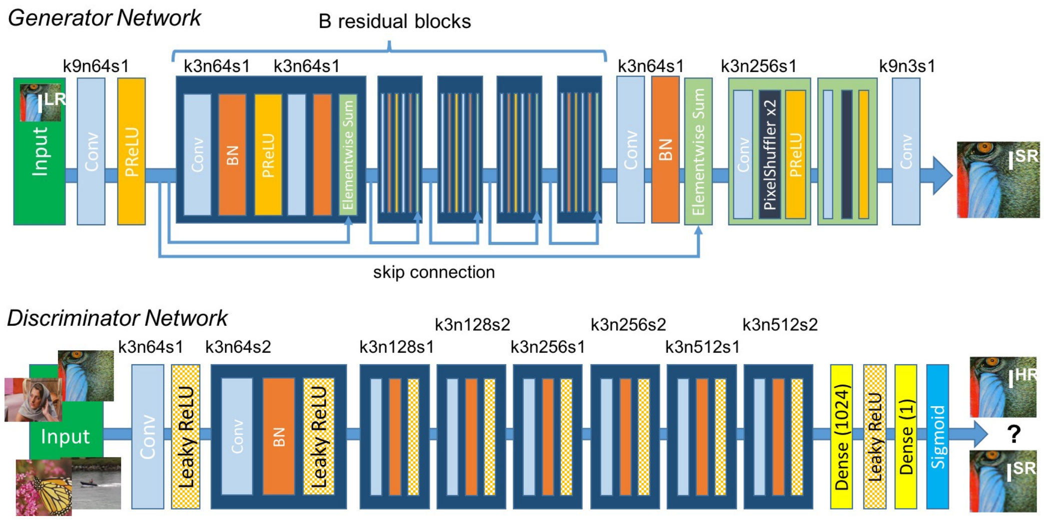

Figure 2. SRGAN Network Architecture: Generator and Discriminator

3.4. Deep Image Prior

Image super-resolution is an optimization task of the form

$$x^* = \min_x E(x; x_0) + R(x)$$

where x is the HR and x_0 is the LR image. The first term is the data term and the second term can be seen as a prior or a regularization term.

The Deep Image Prior paper uses a generator network $x = f_\theta(z)$ to generate x and uses gradient descent to optimize θ as follows:

$$\theta^* = arg \min_\theta E(f_\theta(z); x_0), x^* = f_{\theta^*}(z)$$

Here, z is the latent vector. The authors used a uniformly distributed code vector.

It might seem that the generator should overfit on the noisy image but the authors of the paper showed that the network is "reluctant" to learn LR and it first learns HR and then learns LR images. So, stopping the training process early should result in the generator producing HR images. For the purposes of super-resolution, we set the data term to:

$$E(x; x0) = ||d(x) - x_0||^2$$

where $d(x)$ is a downsampling operator that maps from the HR to LR.

4. Preliminary Results

We calculated PSNR score for all three super resolution factor subsets using Bilinear, Bucubic and Nearest Neighbour interpolation algorithm. PSNR scores were found to be comparable for all three algorithms. We except our Neural Network based models to outperform all these three tradional interpolation algorithms.

	Bicubic	**Bilinear**	**Nearest**
2x	37.5640	37.1617	37.1329
3x	36.4831	36.2783	36.0357
4x	35.8503	35.6117	35.4552

Table 1. Baseline scores

References

[1] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on image processing*, 6(2):298–311, 1997.

[2] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2016.

[3] R. Keys. Cubic convolution interpolation for digital image processing. *IEEE transactions on acoustics, speech, and signal processing*, 29(6):1153–1160, 1981.

[4] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint*, 2016.

[5] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. 2:416–423, July 2001.

[6] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. *arXiv preprint arXiv:1711.10925*, 2017.