Semester Project
CS352

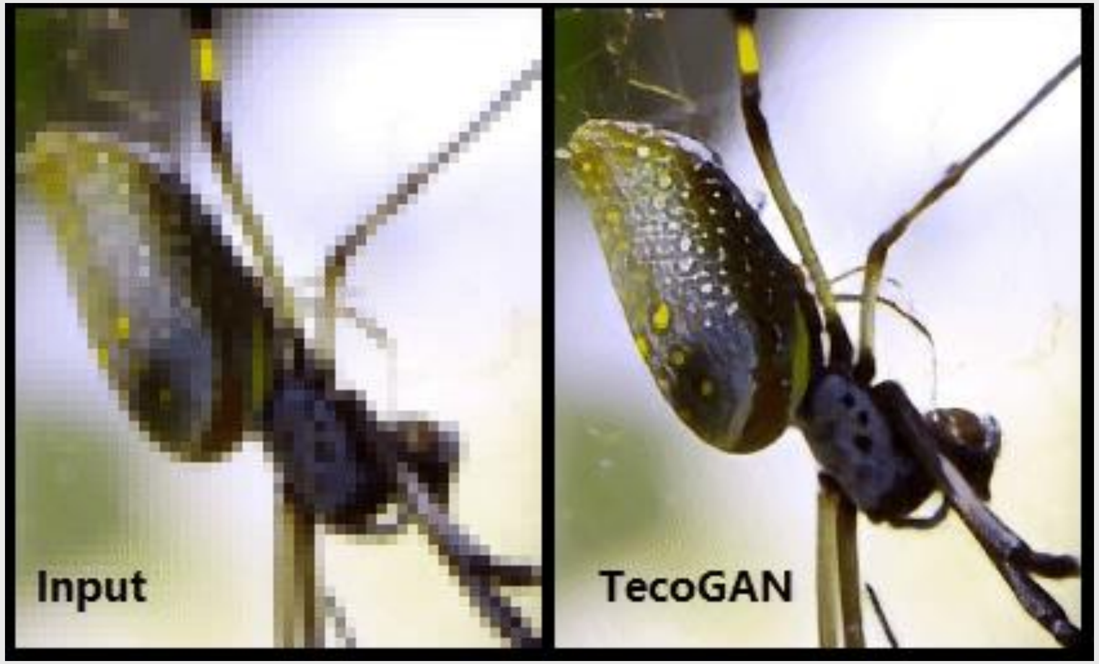
Implementation of SRGAN

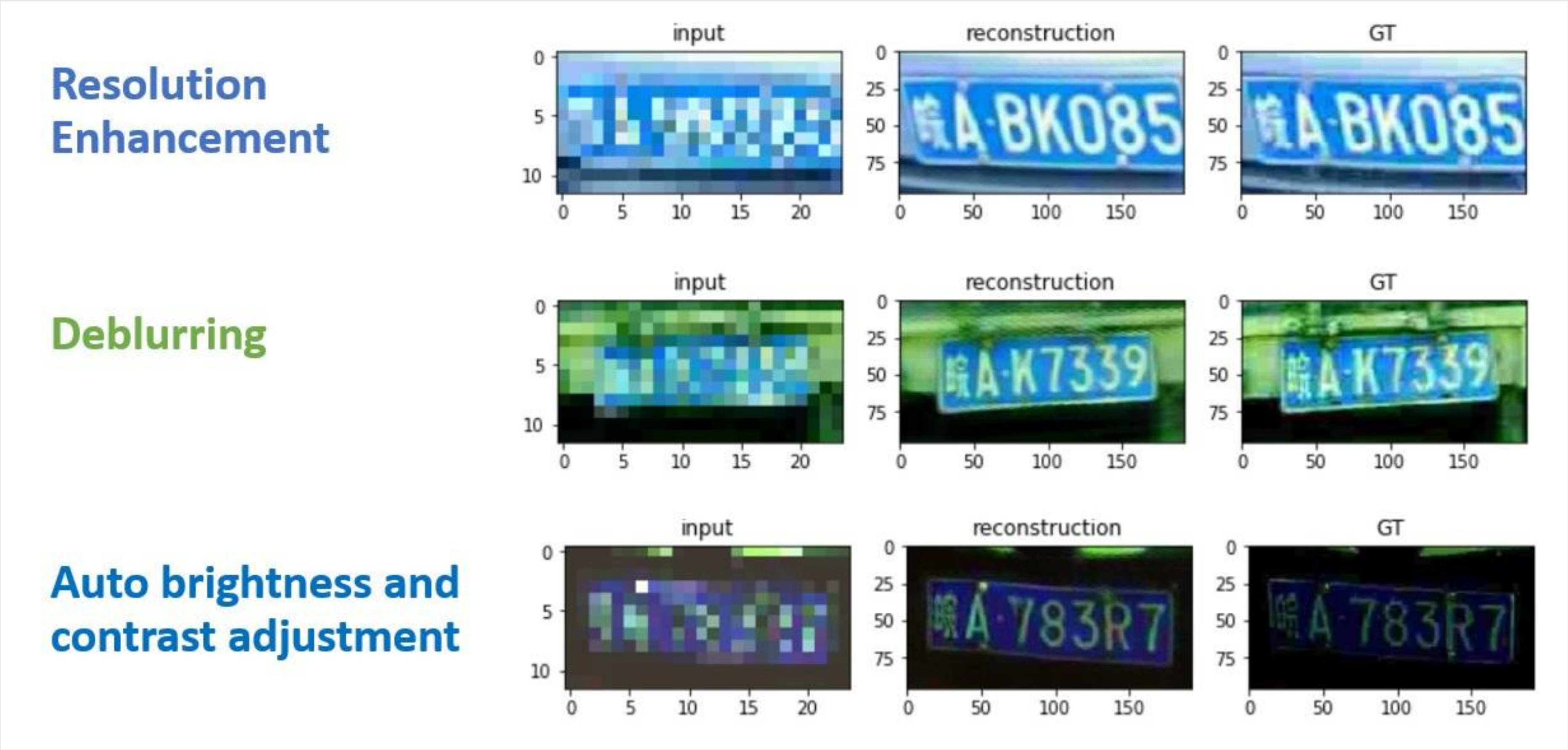
M. Ahabb Sheraz
2021327
BCE

Muhammad Ahmed
2021331
BCE

Super-resolution 101

Super-resolution (SR) involves upsampling a low-resolution image into a higher resolution with minimal information distortion. As our CS352 semester project, we tried to implement the popular Super-Resolution Generative Adversarial Network (SRGAN), an ingenious super-resolution technique that combines the concept of GANs with traditional SR methods.



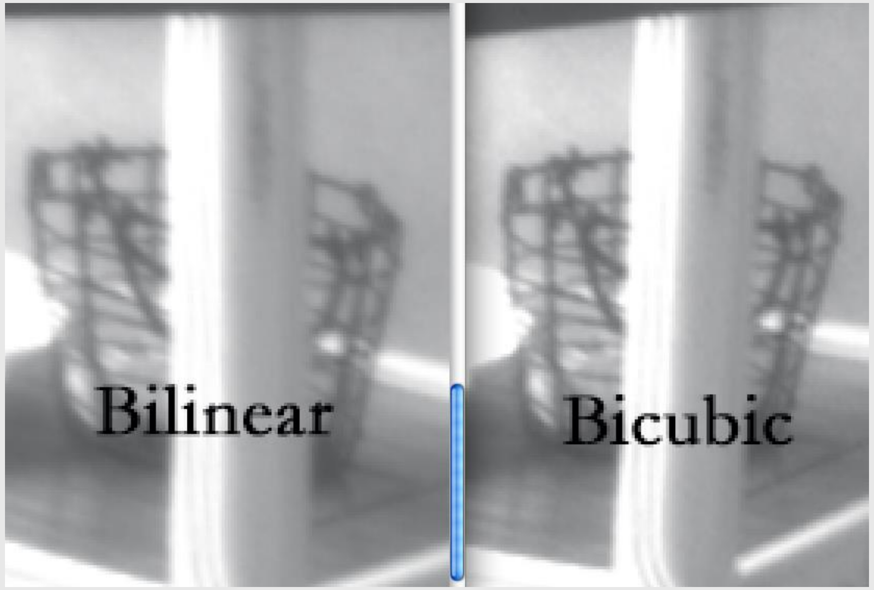


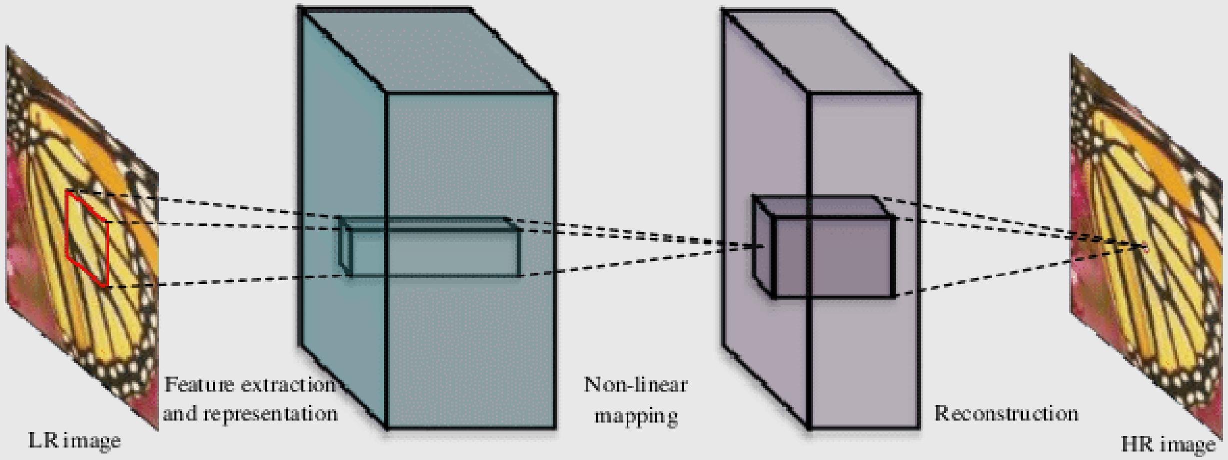
Zixuan Zhang, Chengxuan Cai: License Plate Enhancement - From TV shows to reality (GitHub)

Background

Single Image Super-Resolution (SISR) Techniques (Prior to 2017)

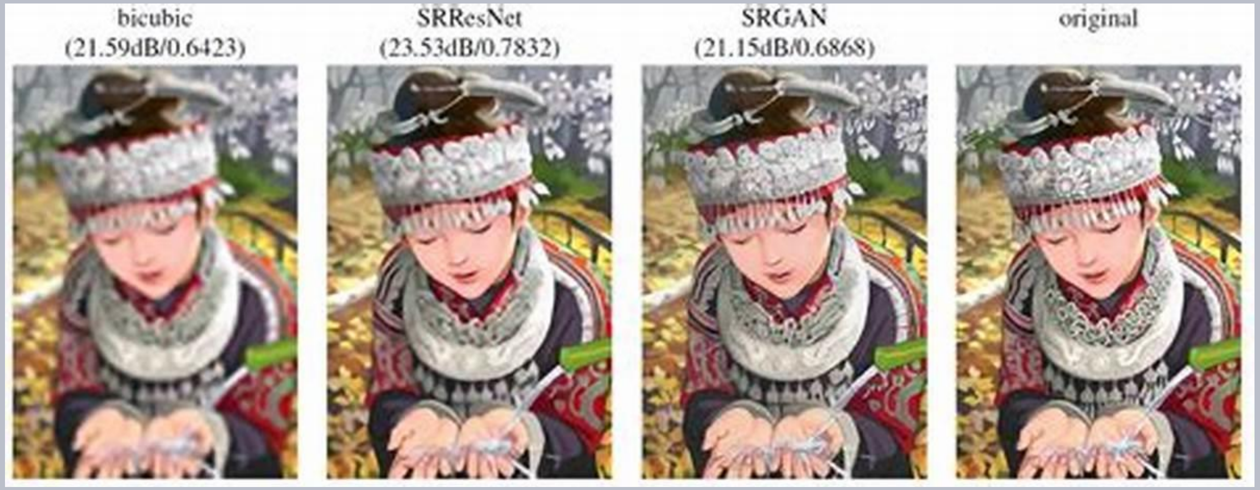
Prior to the introduction of SRGANs, traditional single-image super-resolution (SISR) techniques mainly relied on interpolation methods such as bicubic interpolation. These methods often resulted in blurry images and did not capture fine details. Researchers then began exploring deep learning techniques, particularly convolutional neural networks (CNNs), for image super-resolution. Models such as Super-Resolution Convolutional Neural Network (SRCNN) showed promising results in enhancing the resolution of images. However, they tend to produce artifacts and over-smoothed regions in the super-resolved images, especially in areas containing high-frequency details

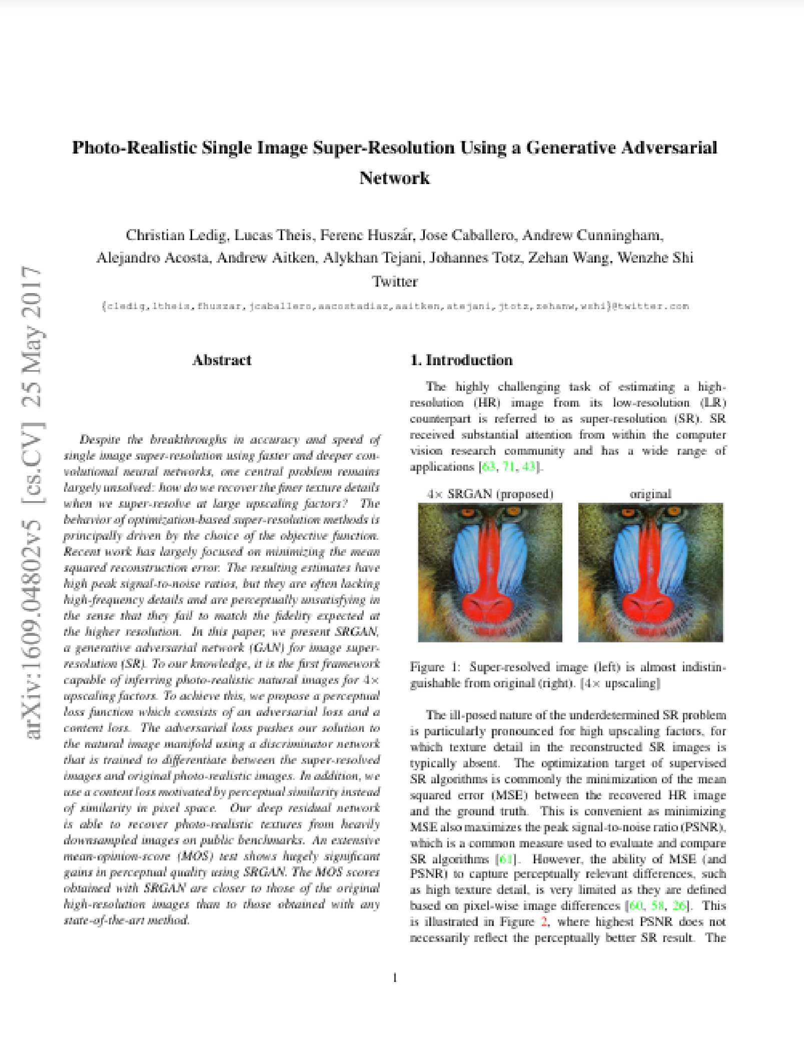




Introduction of SRGAN (2017)

The SRGAN model as proposed in C. Ledig et al., introduced a novel approach to super-resolution using the adversarial training framework of GANs. SRGAN offers superior perceptual quality, sharper edge and texture recovery, and better realism compared to SRCNN, making it a preferred choice for single-image super-resolution tasks in various applications. This is a paper in 2017 CVPR with over 3000 citations.





arXiv:1609.04802v5 [cs.CV] 25 May 2017

Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network

Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham,
Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi
Twitter
{cledig,ltheis,fhuszar,jcaballero,aacostadiaz,aaitken,atejani,jtotz,zehanw,wshi}@twitter.com

Abstract

Despite the breakthroughs in accuracy and speed of single image super-resolution using faster and deeper convolutional neural networks, one central problem remains largely unsolved: how do we recover the finer texture details when we super-resolve at large upscaling factors? The behavior of optimization-based super-resolution methods is principally driven by the choice of the objective function. Recent work has largely focused on minimizing the mean squared reconstruction error. The resulting estimates have high peak signal-to-noise ratios, but they are often lacking high-frequency details and are perceptually unsatisfying in the sense that they fail to match the fidelity expected at the higher resolution. In this paper, we present SRGAN, a generative adversarial network (GAN) for image super-resolution (SR). To our knowledge, it is the first framework capable of inferring photo-realistic natural images for 4× upscaling factors. To achieve this, we propose a perceptual loss function which consists of an adversarial loss and a content loss. The adversarial loss pushes our solution to the natural image manifold using a discriminator network that is trained to differentiate between the super-resolved images and original photo-realistic images. In addition, we use a content loss motivated by perceptual similarity instead of similarity in pixel space. Our deep residual network is able to recover photo-realistic textures from heavily downsampled images on public benchmarks. An extensive mean-opinion-score (MOS) test shows hugely significant gains in perceptual quality using SRGAN. The MOS scores obtained with SRGAN are closer to those of the original high-resolution images than to those obtained with any state-of-the-art method.

1. Introduction

The highly challenging task of estimating a high-resolution (HR) image from its low-resolution (LR) counterpart is referred to as super-resolution (SR). SR received substantial attention from within the computer vision research community and has a wide range of applications [63, 71, 43].

4× SRGAN (proposed) original

Figure 1: Super-resolved image (left) is almost indistinguishable from original (right). [4× upscaling]

The ill-posed nature of the underdetermined SR problem is particularly pronounced for high upscaling factors, for which texture detail in the reconstructed SR images is typically absent. The optimization target of supervised SR algorithms is commonly the minimization of the mean squared error (MSE) between the recovered HR image and the ground truth. This is convenient as minimizing MSE also maximizes the peak signal-to-noise ratio (PSNR), which is a common measure used to evaluate and compare SR algorithms [61]. However, the ability of MSE (and PSNR) to capture perceptually relevant differences, such as high texture detail, is very limited as they are defined based on pixel-wise image differences [60, 58, 26]. This is illustrated in Figure 2, where highest PSNR does not necessarily reflect the perceptually better SR result. The

1

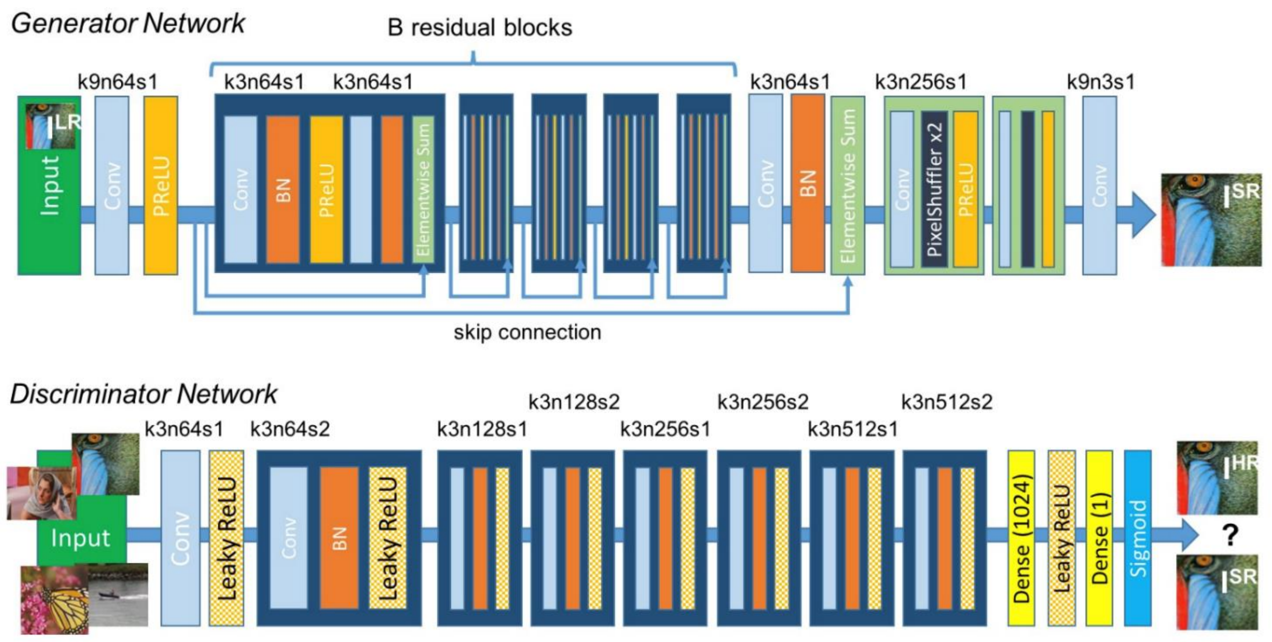


Figure 4: Architecture of Generator and Discriminator Network with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer.

SRGAN

Architecture

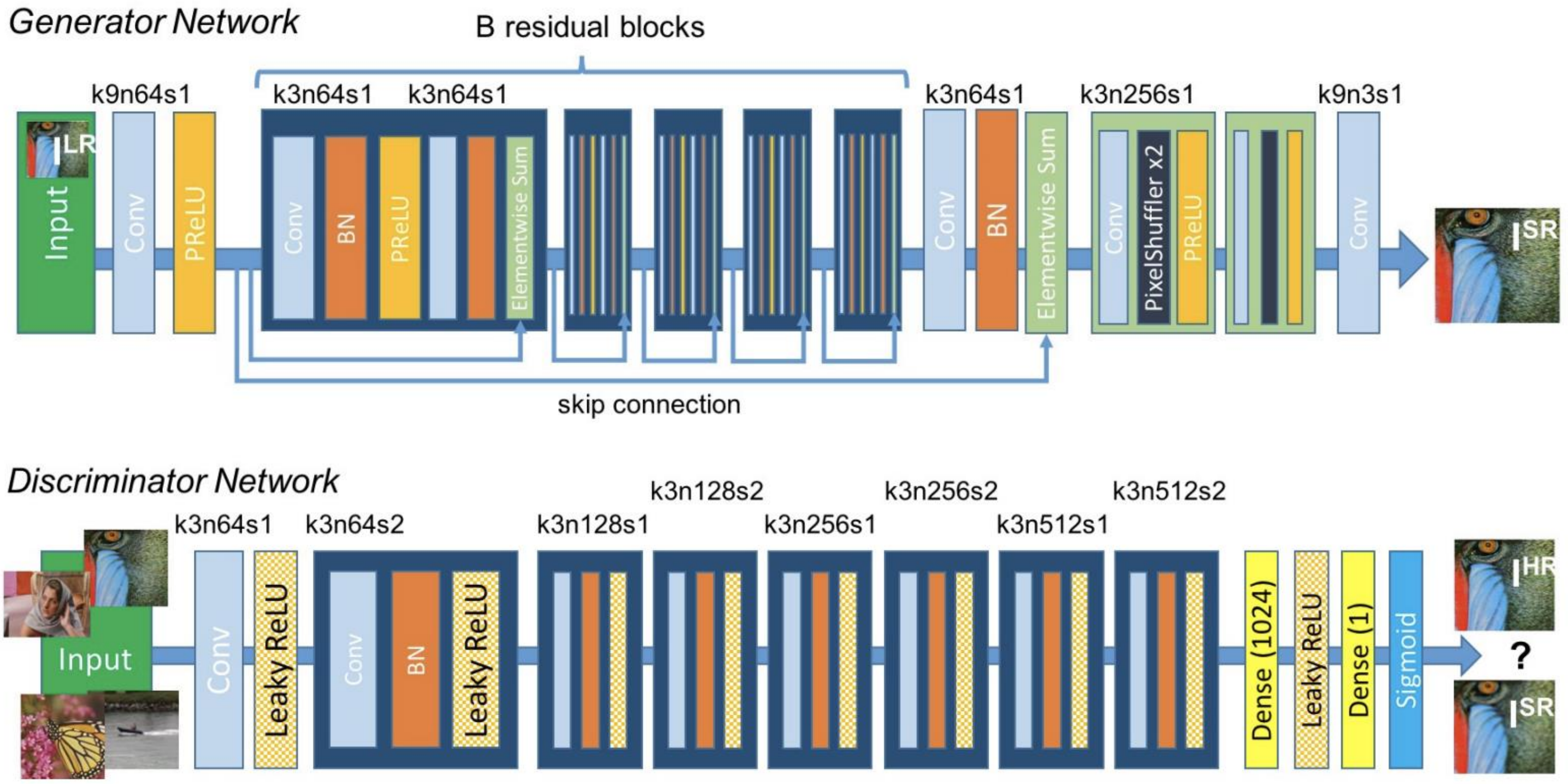


Figure 4: Architecture of Generator and Discriminator Network with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer.

Generator and Discriminator

The generator network employs residual blocks to keep information from previous layers alive and allow the network to choose from more features adaptively. Instead of adding random noise as the generator input, we pass the low-resolution image. The discriminator network is standard and is like the ones used in other GAN models.

Perceptual Loss

The perceptual loss is the combination of both adversarial loss (discriminator loss) and content loss (generator loss). The perceptual loss is formulated as the weighted sum of a content loss and an adversarial loss component.

$$l^{SR} = \underbrace{\underbrace{l^{SR}_{\mathrm{X}}}_{\text{content loss}} + \underbrace{10^{-3} l^{SR}_{Gen}}_{\text{adversarial loss}}}_{\text{perceptual loss (for VGG based content losses)}}$$

Content loss

The content loss evaluates the image quality based on its perceptual quality. **VGG** loss based on the ReLU activation layers of the pre-trained 19-layer VGG network is usually defined as the content loss.

The pixel-wise **MSE loss** is calculated as:

$$l_{MSE}^{SR} = \frac{1}{r^2WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{x,y}^{HR} - G_{\theta_G}(I^{LR})_{x,y})^2$$

$$l_{VGG/i.j}^{SR} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2$$

Adversarial loss

In addition to the content losses described so far, SRGAN also adds the generative component of GANs to the perceptual loss. This encourages the network to favor solutions that reside on the manifold of natural images, by trying to fool the discriminator network.

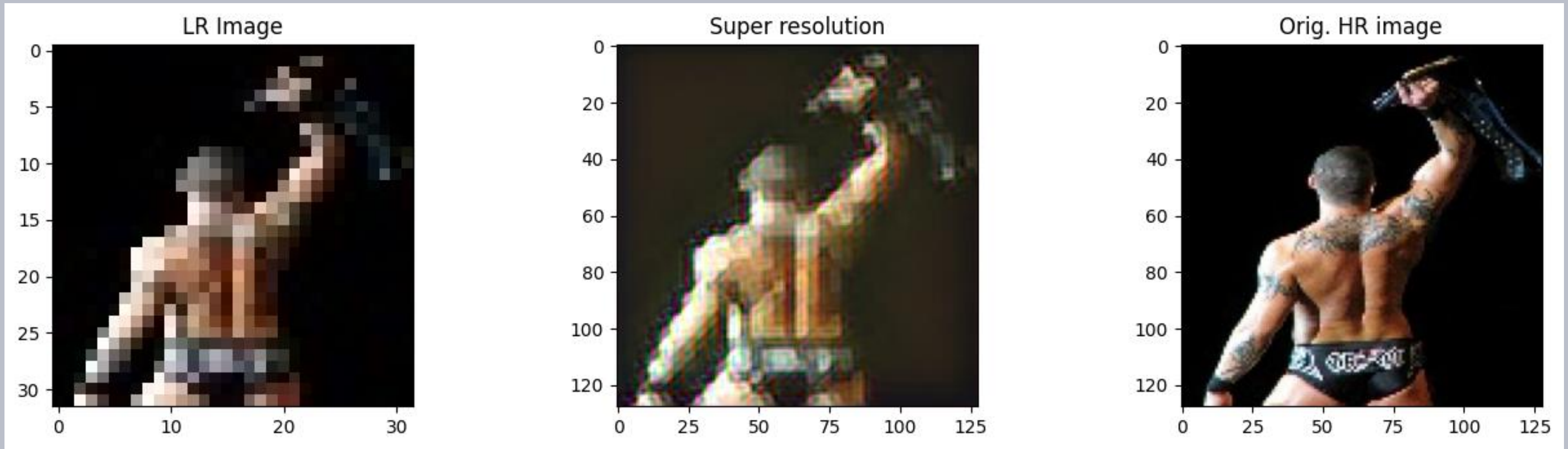
$$l_{Gen}^{SR} = \sum_{n=1}^{N} -\log D_{\theta_D}(G_{\theta_G}(I^{LR})) \tag{6}$$

Here, $D_{\theta_D}(G_{\theta_G}(I^{LR}))$ is the probability that the reconstructed image $G_{\theta_G}(I^{LR})$ is a natural HR image. For better gradient behavior we minimize $-\log D_{\theta_D}(G_{\theta_G}(I^{LR}))$ instead of $\log[1 - D_{\theta_D}(G_{\theta_G}(I^{LR}))]$ [22].

Training and Testing

- Nvidia GTX 1660Ti 6GB
- 5k images from MIRFLICKR25k dataset
- 10 epochs
- Learning rate = 0.001

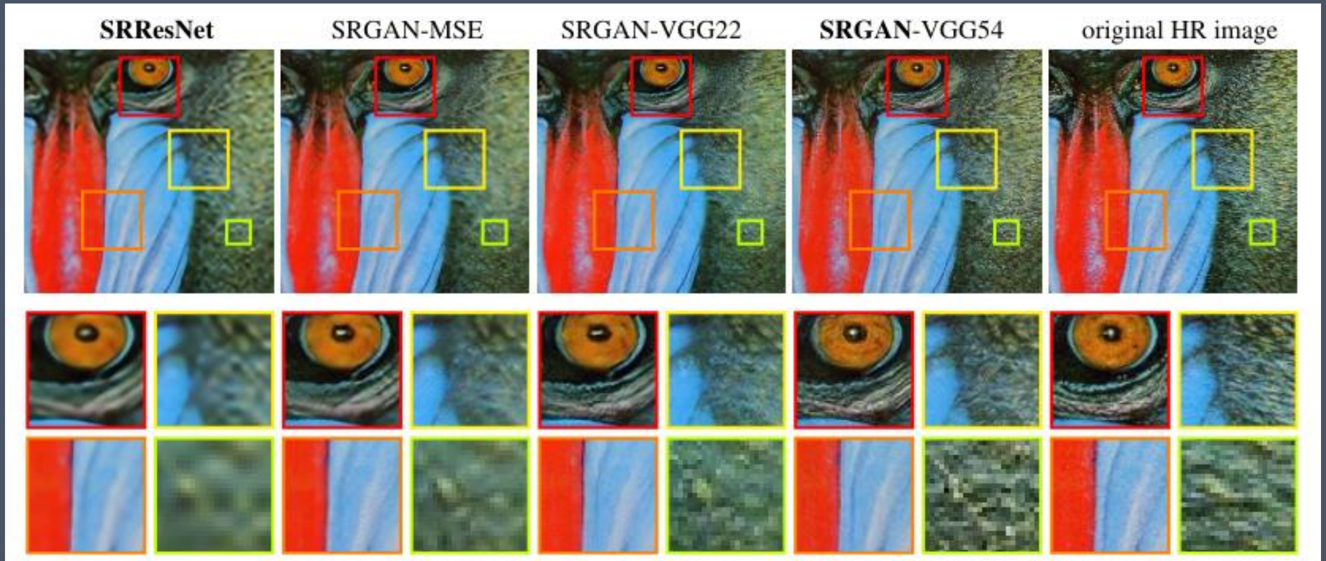
Our results

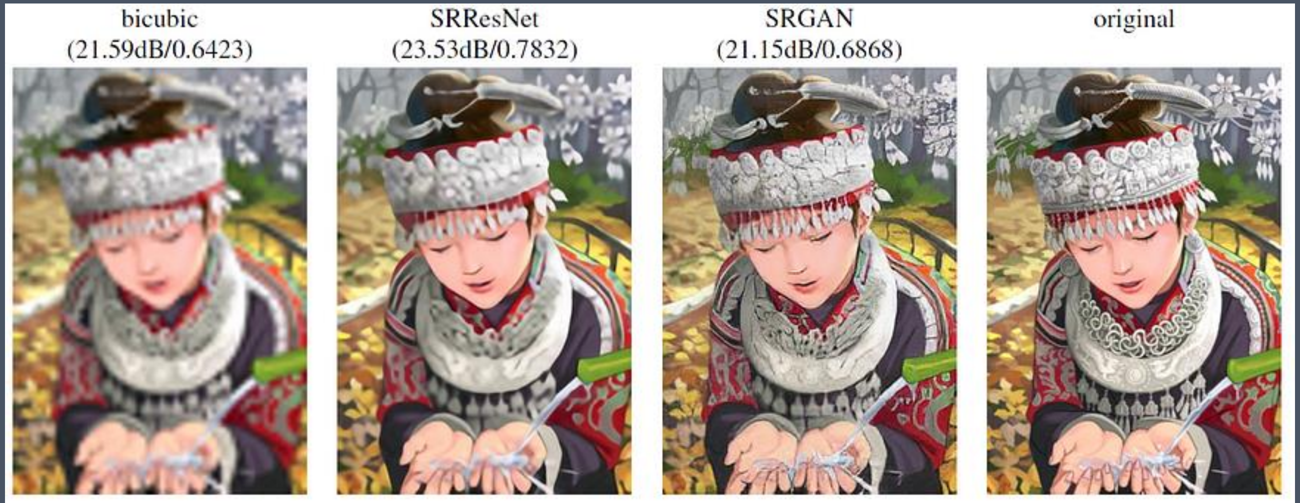

LR Image
Super resolution
Orig. HR image


LR Image
Super resolution
Orig. HR image

Results from the original paper

The original authors trained the model on an NVIDIA Tesla M40 GPU using a random sample of 350 thousand images from the ImageNet database. Their SRGAN variants were trained with 105 update iterations at a learning rate of 0.0001 and another 105 iterations at a lower rate of 0.00001.





Conclusion

After gaining a brief knowledge of the concepts of image and video resolutions, we understood the concept of SRGANs in further detail. We then explored the architecture of this network in detail by looking at the generator and discriminator blocks accordingly. Despite our small training set and low number of epochs, the model was able to show its effectiveness by producing better-quality images when compared to LR images.

Thank

You!