IMAGE RESOLUTION USING SRGAN

PROJECT REPORT

Kaushik Pattipati
Department of Computer Science
University of Nebraska,Omaha
kpattipati@unomaha.edu

1 Introduction

The task of image super-resolution, where a low-resolution image is transformed into a high-resolution counterpart, has gained significant attention due to its wide-ranging applications in medical imaging, satellite surveillance, video streaming, and more[Add Citation for applications]. Traditional methods, such as bicubic interpolation, fail to recover fine textures, often resulting in blurry outputs that lack realistic detail. With the advent of deep learning, approaches like Convolution neural networks (CNNs) have shown promising results in improving the structural quality of super-resolved images.

Super-Resolution Generative Adversarial Network (SRGAN) has emerged as a groundbreaking framework for perceptual image super-resolution. Unlike methods that focus solely on minimizing pixel-wise errors. SRGAN combines deep residual learning with adversarial training to generate images that are accurate and realistic. The adversarial training mechanism enables SRGAN to achieve a balance between pixel-wise accuracy and perceptual realism, addressing critical challenges in image super-resolution. The generated images not only exhibit sharper details but also appear more natural and visually appealing.

1.1 Problems to Be Solved

Despite significant advancements in deep learning, several challenges persist in image super-resolution:

- **Detail Recovery:** Traditional methods struggle to reconstruct fine textures and intricate details, leading to visually unappealing results.
- **Perceptual Realism:** Many approaches optimize for pixel-wise accuracy, resulting in images that appear smooth and unnatural to the human eye.
- **Generalization:** Ensuring that the model performs well across diverse datasets and image types without overfitting remains a challenge.

SRGAN addresses these issues by introducing adversarial training, which emphasizes perceptual realism alongside structural accuracy. The framework aims to bridge the gap between pixel-level fidelity and human-perceived quality.

1.2 Applications of the Method

SRGAN can provide high-quality image super-resolution:

- **Medical Imaging:** Enhancing MRI, CT, and ultrasound scans for improved diagnostic accuracy and treatment planning.
- **Satellite Surveillance:** Improving the resolution of satellite imagery to aid in urban planning, environmental monitoring, and disaster management.
- **Forensic Analysis:** Refining surveillance footage to reveal critical details for law enforcement and investigative purposes.

The paper explores the architecture, training methodology, and performance of SRGAN, showcasing its ability to push the boundaries of image super-resolution.

2 Related work

Multiple methods have been proposed for Image Super-resolution such as bicubic and nearest-neighbor interpolation Keys [1981] were computationally efficient but often resulted in blurry outputs and failed to reconstruct high-frequency details.

2.1 Key Contributions of Related Work

When Deep learning was introduced, multiple approaches to super-resolution have been proposed. Including SRCNN (Super-Resolution Convolutional Neural Network), proposed a shallow three-layer CNN to learn the mapping between low-resolution and high-resolution images Dong et al. [2014]. While effective, SRCNN was limited by its depth, which restricted its ability to capture complex features.

To address this, VDSR (Very Deep Super-Resolution) employed deeper architectures with residual learning to enable efficient training and improved reconstruction quality [Kim et al., 2016]. Further advancements included ESPCN (Efficient Sub-Pixel Convolutional Neural Network), which introduced sub-pixel convolution layers for efficient upsampling [Shi et al., 2016].

At later stage, Generative Adversarial Networks (GANs) revolutionized super-resolution by focusing on perceptual realism. SRGAN (Super-Resolution Generative Adversarial Network) combined a generator network with a discriminator to produce visually realistic high-resolution images Ledig et al. [2017]. SRGAN employed a perceptual loss based on a pre-trained VGG network, enabling it to generate textures that closely resemble natural images.

Despite its success, SRGAN faced challenges such as training instability and a tendency to overemphasize textures at the cost of structural accuracy Blau and Michaeli [2018]. Subsequent works like ESRGAN (Enhanced Super-Resolution GAN) refined SRGAN by introducing a relativistic discriminator and residual-in-residual blocks, further improving perceptual quality Wang et al. [2018].

2.2 Advantages of SRResNet and SRGAN

Super-resolution reconstruction model (SRResNet), a precursor to SRGAN, focused on minimizing pixel-wise error using residual blocks and sub-pixel convolution for accurate high-resolution reconstruction Ledig et al. [2017]. While SRResNet excelled in structural accuracy, it often produced smooth outputs lacking realistic textures. SRGAN extended SRResNet by introducing adversarial training, which balances structural fidelity with perceptual realism.

Compared to traditional methods, SRGAN demonstrates several advantages:

- **Perceptual Quality:** By incorporating adversarial and perceptual losses, SRGAN generates outputs with realistic textures and finer details.
- **Balanced Reconstruction:** The combination of content and adversarial loss ensures that the outputs are visually appealing without compromising structural accuracy.
- **Versatility:** SRGAN has shown success across diverse datasets, making it suitable for a wide range of applications such as medical imaging, satellite analysis, and video enhancement.

3 Methodology

The paper leverages SRGAN framework to enhance low-resolution images into high-resolution counterparts. SRGAN combines a generator network(refer Figure 1), based on the **Super-Resolution Residual Network (SRResNet)**, and a discriminator network trained in an adversarial setting. This section details the architecture, loss functions, and training process.

3.1 Generator Network (SRResNet)

The generator is designed to reconstruct high-resolution images from low-resolution inputs using residual learning and efficient upsampling techniques. Its architecture includes the following components:

- **Input Layer:** Accepts a low-resolution image of size $n \times n \times 3$.

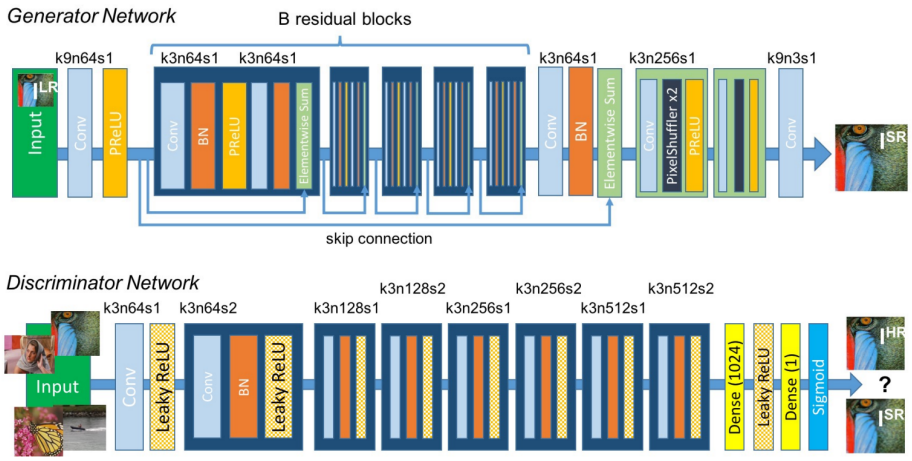

Figure 1: SRGAN Architecture. Source [Ledig et al., 2017]

- **Residual Blocks:** The generator consists of B residual blocks, each incorporating:
  - A convolutional layer with kernel size 3×3, stride 1, and 64 filters.
  - Batch normalization to stabilize training and improve convergence.
  - PReLU activation function to introduce non-linearity.
  - Element-wise summation for skip connections, facilitating gradient flow and feature reuse.

  The residual block operation is given as:

$$F_{residual} = F_{input} + Conv2D(BN(PReLU(Conv2D(F_{input}))))$$

- **Upsampling Layers:** Two sub-pixel convolution layers are used for efficient upscaling, implemented using pixel shuffling. This avoids artifacts and ensures computational efficiency:

$$F_{upsampled} = DepthToSpace(Conv2D(F_{residual}, k = 3))$$

- **Output Layer:** A final convolutional layer with kernel size 9×9 and a `tanh` activation function generates the super-resolved image.

3.2 Discriminator Network

The discriminator is a convolutional neural network that evaluates the realism of the generated images by classifying them as real or fake. Its architecture includes:

- **Input Layer:** Accepts an image of size $128 \times 128 \times 3$ (real or generated).
- **Convolutional Layers:** Sequential layers with increasing filter sizes (64, 128, 256, and 512) and strides, capturing hierarchical features. Each convolutional layer includes batch normalization and LeakyReLU activation:

$$F_{conv} = LeakyReLU(BN(Conv2D(F_{input})))$$

- **Fully Connected Layers:** The flattened output is passed through dense layers with 1024 nodes and LeakyReLU activation.
- **Output Layer:** A single node with a `sigmoid` activation function outputs the probability of the input being real.

3.3 Loss Functions

The SRGAN framework combines two loss functions to optimize the generator and discriminator:

1. **Content Loss:** Based on a perceptual loss derived from a pre-trained VGG19 network. This ensures the generated image matches the high-level semantic features of the ground truth:

$$\mathcal{L}_{content} = \frac{1}{N} \sum_{i=1}^{N} \|\phi(I_{HR}) - \phi(I_{SR})\|^2$$

   where ϕ represents the feature maps from VGG19, I_{HR} is the high-resolution ground truth, and I_{SR} is the generated image.

2. **Adversarial Loss:** Encourages the generator to produce realistic images that can fool the discriminator:

$$\mathcal{L}_{adv} = -\mathbb{E}[\log D(I_{SR})]$$

3. **Discriminator Loss:** The discriminator minimizes the following loss:

$$\mathcal{L}_D = -\mathbb{E}[\log D(I_{HR})] - \mathbb{E}[\log(1 - D(I_{SR}))]$$

The total generator loss is given by:

$$\mathcal{L}_G = \mathcal{L}_{content} + \lambda \mathcal{L}_{adv}$$

where λ is a weighting factor for adversarial loss.

3.4 Use of VGG16 for Perceptual Loss

The VGG16 model, pre-trained on the ImageNet dataset, was employed to calculate perceptual loss. Specifically, the feature maps from intermediate layers of VGG16 were extracted to compare high-level semantic similarities between the generated and ground truth images. The following layers were used:

- **Conv2_2:** Focuses on low-level details such as textures and edges.
- **Conv5_4:** Captures high-level semantic information, ensuring perceptual realism.

3.5 Training Process

The training process was divided into two phases:

- **Phase 1: Pretraining the Generator:** The generator was pretrained using pixel loss (MSE) to ensure structural accuracy in the generated images.
- **Phase 2: Adversarial Training:** The generator and discriminator were trained in an adversarial setting. Perceptual loss (using VGG16) and adversarial loss were combined to optimize the generator:

$$\mathcal{L}_G = \mathcal{L}_{perceptual} + \lambda \mathcal{L}_{adv}$$

  where λ is a weighting factor for the adversarial loss.

The model was created using TensorFlow and trained on the TensorFlow Flowers dataset. The training uses a batch size of 64 and an SGD optimizer with a learning rate of 0.0001.

4 Experiment

The experiment is conducted using the **TensorFlow Flowers Dataset**, the dataset contains a diverse collection of flower images with varying textures, colors, and patterns. In addition, its limited in quantity making it easy to train the model. The data preparation process involved:

- **Training Set:** The majority of images were allocated to the training set.
- **Testing Set:** A subset of 600 images was reserved for testing and evaluation.
- **Preprocessing:** High-resolution images were resized to 128×128 pixels, and low-resolution images were created by downscaling these to 32×32 pixels using bilinear interpolation.

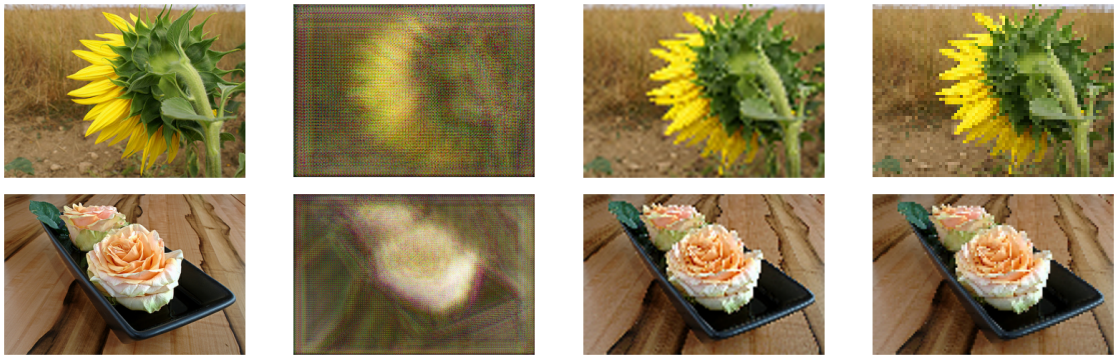

Figure 2: Image - Ground Truth, Low-Resolution (Bicubic) , SRGAN Output, Bicubic Interpolation (From left to right)

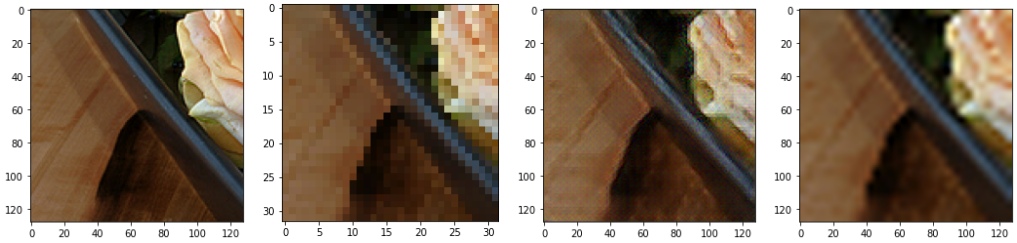

Figure 3: Ground Truth Image,Low resolution (Bicubic), SRGAN Output, Bicubic Interpolation

4.1 Experiments Conducted

The evaluation of SRGAN was performed based on qualitative analysis, focusing on visual comparisons to assess the perceptual quality of the generated super-resolution images. The following observations were made:

1. **Perceptual Quality:** The outputs of SRGAN were visually compared to bicubic interpolation and the ground truth high-resolution images. The SRGAN-generated images exhibited sharper textures, enhanced edges, and more realistic details compared to bicubic interpolation. Figures 2 and 3 highlight these improvements, showcasing the ability of SRGAN to recover high-frequency details.
2. **Visual Comparison:** A side-by-side comparison was performed, presenting:
   - The low-resolution input image.
   - The upscaled image using bicubic interpolation.
   - The SRGAN-generated image.
   - The ground truth high-resolution image.

   These comparisons demonstrate the superior visual quality of SRGAN outputs, with more natural textures and reduced artifacts.

4.2 Results

The experimental results demonstrate the effectiveness of the SRGAN model in generating high-quality super-resolved images.

Figures 2 and 3 illustrate the comparison between the ground truth, low-resolution (bicubic), SRGAN output, and bicubic interpolation. The following observations can be made:

- **Bicubic Interpolation:** While bicubic interpolation slightly improves resolution compared to the low-resolution input, it fails to recover fine details and produces overly smooth outputs.
- **SRGAN Output:** The SRGAN-generated images successfully recover high-frequency details, such as edges and textures, resulting in sharper and more visually appealing outputs compared to bicubic interpolation.
- **Ground Truth:** The SRGAN outputs closely match the ground truth, showcasing the model's ability to generate perceptually realistic and high-quality images.

Conclusion

Image super-resolution has wide-ranging applications in fields such as medical imaging, satellite data processing, and video streaming. The paper demonstrates the implementation of the **Super-Resolution Generative Adversarial Network (SRGAN)** framework to enhance low-resolution images into high-resolution counterparts. The SRGAN framework combines the strengths of residual learning and adversarial training, enabling it to generate images that are both structurally accurate and visually realistic. SRGAN model's Generator Network(SRResNet) and Discriminator Network were able to provide high resolution image.

The model was trained using a combination of **content loss** (based on VGG19 features) and **adversarial loss**, balancing pixel accuracy and perceptual realism. In conclusion, the SRGAN framework successfully addresses the challenges of image super-resolution, offering a practical and effective solution for generating high-quality images. Its ability to bridge the gap between low-cost acquisition and high-resolution outputs makes it a valuable tool for a wide range of applications in industry and research.

References

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

Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, pages 184–199. Springer, 2014.

Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.

Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016.

Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6228–6237, 2018.

Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018.