Work in Progress

SIPL Signal and Image Processing Lab

Learning Super-Resolution space

Chen Goldenberg and Uri Savir, Supervised by Idan Kligvasser

Challenge of NTIRE 2021

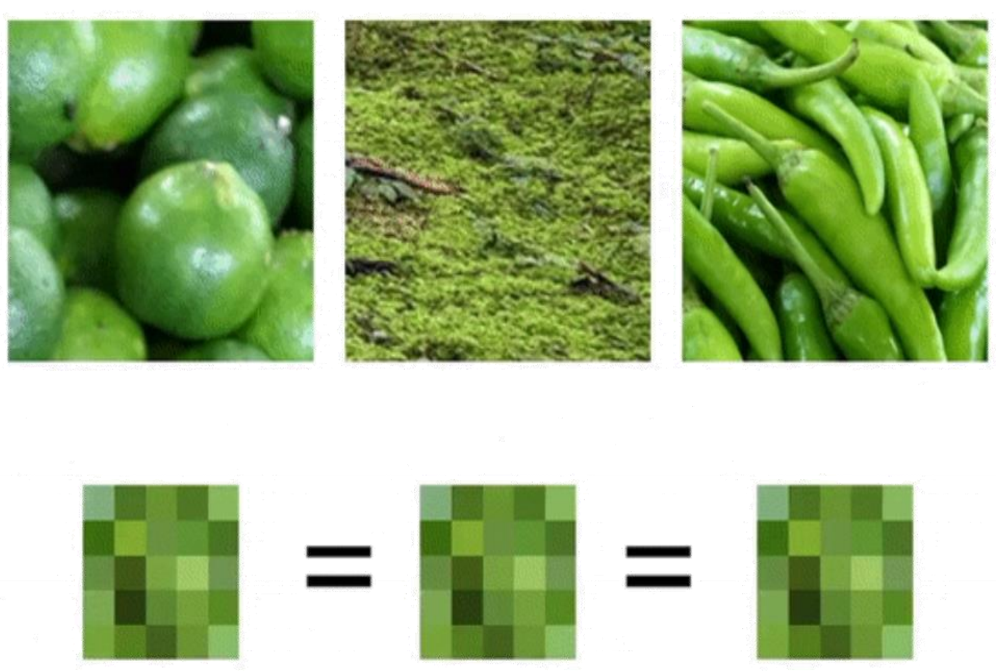
Introduction

- Usually, super-resolution (SR) is trained using pairs of high- and low-resolution images
- Infinitely many high-resolution images can be down-sampled to the same low-resolution image
- That means that the problem is ill-posed and cannot be inverted with a deterministic mapping
- Instead, one can frame the SR problem as learning a stochastic mapping, capable of sampling from the space of plausible high-resolution images given a low-resolution image

space of plausible SR

Goals

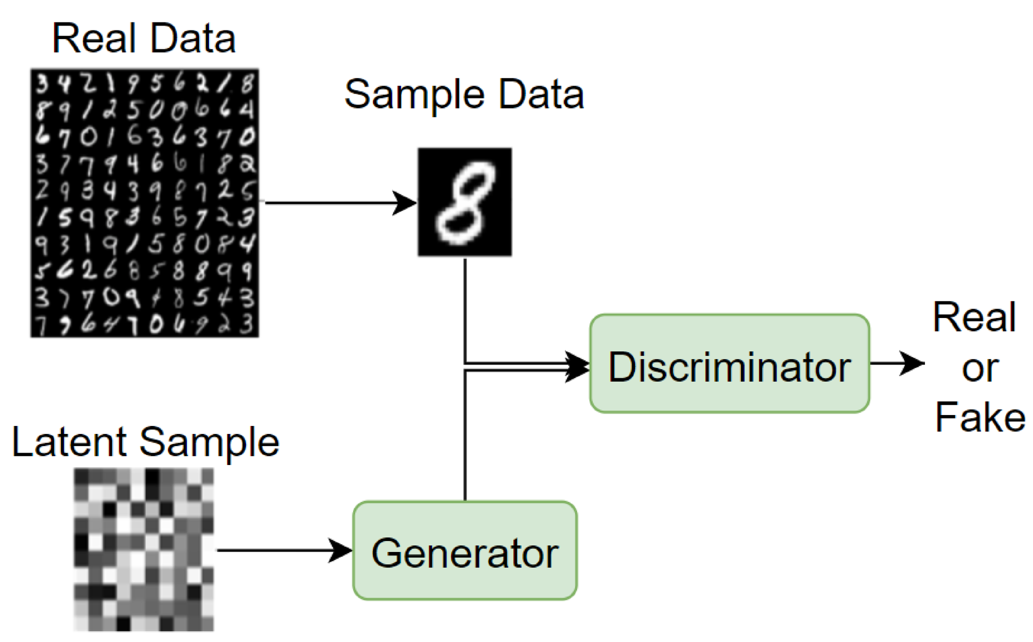
- The goal of this challenge is to develop a super-resolution method that can actively sample from the space of plausible super-resolutions
- Each individual SR prediction should achieve highest possible photo-realism, as perceived by humans
- The solution should be capable of sampling an arbitrary number of SR images capturing meaningful diversity
- Each individual SR prediction should be consistent with the input low-resolution image



Evaluation principles

- Visuality evaluation as perceived by humans
- Spanning the SR Space in order to get meaningful diversity
- Obtain a LR-PSNR of:
  - $\mathrm{PSNR}(\mathrm{SR}\downarrow, \mathrm{LR}) \geq 45\ \mathrm{dB}$

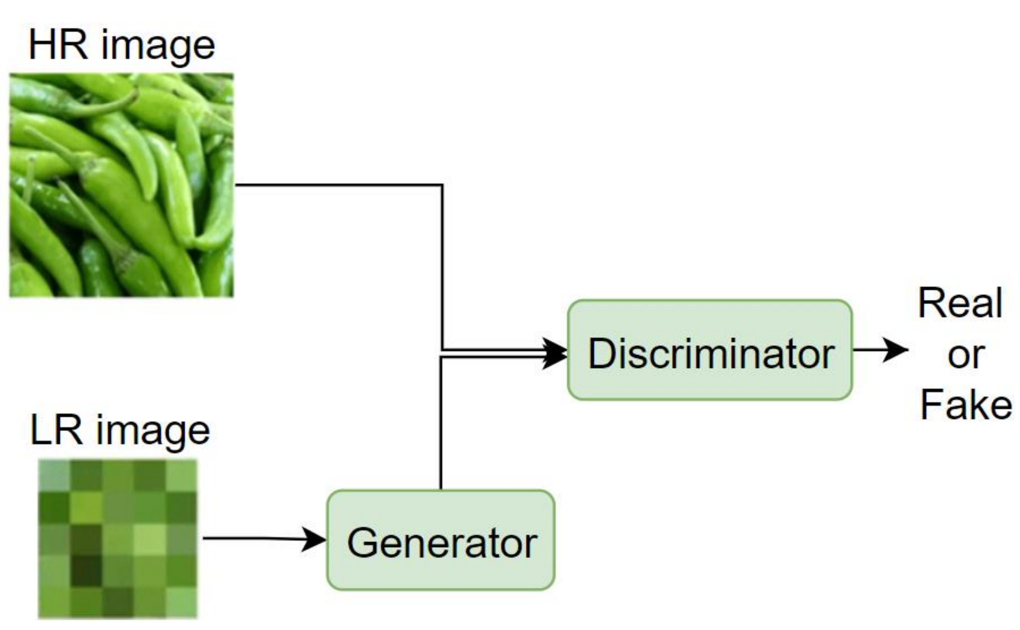
Generative Adversarial Network - GAN



- Adversarial training of Generator & Discriminator
- The Generator generate fake samples, tries to fool the Discriminator
- The Discriminator tries to distinguish between real and fake samples
- The training is adversarial: train them against each other

SR-GAN

- Super Resolution GAN
- The training set is Low-Resolution images, and the ground-truth is their High-Resolutions
- Low-Resolution images are the input for the Generator
- The Generator generates using LR images the SR images to feed the Discriminator and try to fool it
- The Discriminator inputs consists of both HR and SR images, and is trained to distinguish between real HR to fake generated SR



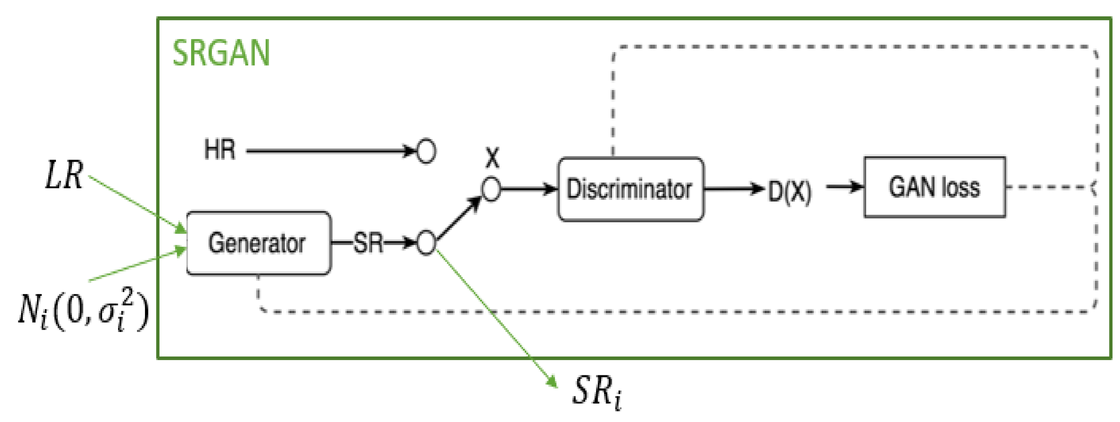
SRGAN scheme

Challenge Rules

- The method cannot be limited to a maximum number of different SR samples
- All SR samples must be generated by a single model
  - No ensembles are allowed
- No self-ensembles during inference
  - e.g. flipping and rotation
- All SR samples must be generated using the same hyper-parameters
- Deterministic methods will naturally score zero in the diversity measure

Solution Method

- Modifying the SRGAN architecture and loss function in order to fulfil the competition challenge
- To get diversity between outputs:
  - Use random noise as input to the generator in addition to the LR image
  - Added “diversity” element to the loss function:
    - Takes two sequential outputs and insert each one to a pretrained classification network, then uses L1 between the outputs
    - Negative coefficient - maximization required



$$LOSS = \alpha \cdot \mathcal{L}_{SRGAN}(SR_i, GT) + \beta \cdot \mathcal{L}_1(SR_i \downarrow, LR) - \gamma \cdot \mathcal{L}_{var}\big(\phi(SR_i), \phi(SR_j)\big)$$

Our solution scheme

- In order to stay consistent to the original LR - added PSNR element to the loss function:
  - Down sampling (bicubic) the SR output to the same dimensions as LR
  - Calculate PSNR between the LR and the SR-Down-Sampled images

June 2021