

---

# Overview of pix2pix and SRGAN

Arnab Sinha • 03.06.2020

---

---

# SRGAN

---

# Aim

- To generate photo realistic super resolution images from corresponding low resolution images.
- Using appropriate loss functions to obtain perceptual similarity to original HR images.

---

Main loss formula =

$$l^{SR} = \underbrace{l_X^{SR}}_{\text{content loss}} + \underbrace{10^{-3}l_{Gen}^{SR}}_{\text{adversarial loss}}$$

perceptual loss (for VGG based content losses)

Content loss:

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

Here,  $\phi_{i,j}$  is the feature map obtained by the  $j^{\text{th}}$  convolution after the  $i^{\text{th}}$  max pooling layer within the pre-trained VGG network.

So, it is the comparison of the feature map made by fake image and real image. Focuses on perceptual similarity by focusing on generating better high frequency content, which is where MSE loss performs bad.

Adversarial loss:

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

This is the generative component of the adversarial loss. This pushes the generator to create images in the locality where natural images reside.

---

---

# Pix2Pix

---

# AIM

Using conditional GANs to create a model suitable for image to image translation tasks.

A condition is put in the form of an input image and a corresponding output image is formed. This helps us control the output.

---

---

Main loss formula =

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G).$$

Conditional adversarial loss:

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y} [\log D(x, y)] + \mathbb{E}_{x,z} [\log(1 - D(x, G(x, z)))],$$

This is the conditioned loss with the only change that the input image  $x$  is also used to condition the discriminator and generator.

This way, the discriminator compares  $x$  with both the generated output and the real output, providing a much stronger loss function.

L1 loss:

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z} [\|y - G(x, z)\|_1].$$

Calculates the pixel wise loss which makes sure that even though the generator might have reached the locality of the ground truth, it generates an image as close to it as possible.

L2 loss blurs a lot due to the squared error which might be causing the function to minimize faster.

---