Super-Resolution with GANs

Yang Wang(leonwy12@stanford.edu), Yangxin Zhong(yangxin@stanford.edu), Qixiang Zhang(qixiang@stanford.edu)

Introduction

Project Task: Given a low-resolution (LR) image, recover or estimate its high-resolution (HR) version image. This is referred as Super Resolution(SR). SR is an important topic in computer vision as it can be beneficial to many applications such as texture mapping, enlarging consumer photograph, converting NTSC video content to HDTV, and low-resolution face recognition.

Methods: In this project, we use deep neural network to build several models for super-resolution problem, including SRResnet and SRGAN[1] models. We also succeed to improve the performance of SRGAN by replacing its vanilla GAN with a WGAN-GP model. Other attempts to improve the models, such as ratio tuning strategy and combined content loss, are also explored in this project.

Dataset

We used the the RAISE dataset, which consists of 8156 pictures from categories such as "outdoor", "indoor", "landscape", "nature", "people", "objects" and "buildings". The high resolution pictures are paired with their 4x downsampled counterparts to form our training set. We split the data into three parts: 400 for dev set, 400 for test set, and the rest are our training data. During training, we also applied random crop with crop size 24 by 24.

Models

SRGAN[1] model:

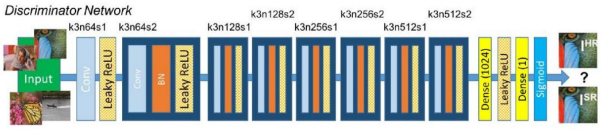

Fig. 1 Discriminator network

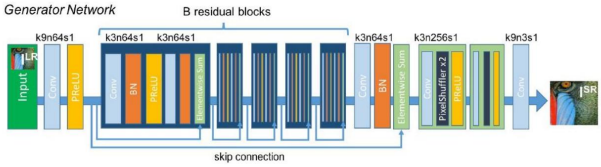

Fig. 2 Generator network

Training GAN is to find the equilibrium of a minimax schema:

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

In our task the generator loss is

$$l_{Gen} = l_{Con} + \gamma \cdot l_{Adv}$$

Discriminator loss is

$$l_{Dis} = \sum_{n=1}^{N} -\log D_{\theta_D}(I^{HR}) + \sum_{n=1}^{N} -\log\left[1 - D_{\theta_D}(G_{\theta_G}(I^{LR}))\right]$$

where γ is a tunable hyperparameter and Content loss may vary for different experiments.

Our contributions and improvements:

- Replace vanilla GAN with WGAN-GP[2] model in SRGAN to stabilize training and improve performance
- Explore different techniques, such as discriminator regularization, ratio tuning, and combined content loss, to improve the perceptual quality of images generated

References

[1] Christian Ledig, Lucas Theis, Ferenc Huszar, 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. arXiv preprint, 2016.

[2] Gulrajani, Ishaan, et al. "Improved training of wasserstein gans." Advances in Neural Information Processing Systems. 2017.

Evaluation

We trained and tested our models on RAISE dataset, evaluating using MSE(Mean Squared Error), PSNR(Peak Signal-to-Noise Ratio), SSIM(Structural Similarity) and MOS(Mean Opinion Score). Among these evaluation metrics, a lower MSE, and higher PSNR, SSIM or MOS values indicate a better performance.

Model	MSE	PSNR	SSIM	MOS
Bilinear	50.18	31.49	0.617	3.33
Bicubic	50.21	31.48	0.621	3.34
SRResnet(MSE)	**40.92**	**32.59**	**0.758**	4.45
SRResnet(Combined)	41.83	32.48	0.746	5.02
SRGAN(MSE)	41.83	32.54	0.754	4.87
SRGAN(MSE+LS)	44.17	32.19	0.730	5.81
SRGAN(MSE+IN)	43.08	32.31	0.736	5.50
SRGAN(VGG54)	57.41	30.77	0.676	**7.05**
SRGAN(VGG54+RT)	57.71	30.74	0.658	6.85
SRGAN(Combined)	53.13	31.15	0.677	6.86
SRWGAN-GP(MSE)	44.70	32.04	0.733	5.59
SRWGAN-GP(VGG54)	60.22	30.51	0.687	6.90
SRWGAN-GP(VGG54+RT)	68.96	29.82	0.665	**7.32**
SRWGAN-GP(Combined)	60.60	30.48	0.685	**7.34**

Table. 1 Experiment Results

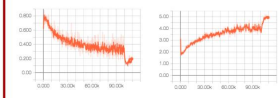

Fig2: Discriminator loss and adversarial loss of SRGAN using VGG54 as content loss

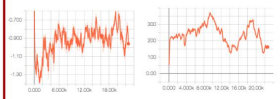

Fig3: Negative critic loss and adversarial loss of SRWGAN which combines VGG54, VGG22 and MSE as content loss

Fig. 1 Result Comparison (LR, SRResnet, SRGAN, SRWGAN, HR)

Discussion:

- SRResnet variants achieve better performance in metrics such as MSE, PSNR, SSIM, however, in human eyes, they generate images that are clearly blurry and obtained much lower scores in MOS.
- Training a SRResnet with deeper layer VGG content loss often results in generating colorful texture which is not realistic in human eyes.
- Using GAN models helps relieve the situation above by "differentiating" it from real high resolution images thus penalizes it generating unrealistic texture.
- Training of GAN is difficult and unstable, often results in very low discriminator loss and high adversarial loss which prevents generator from learning.
- WGAN-GP relieves the problem by stabilizing the training of critics and generator using more continuous metric between distributions.