Anime Super Resolution Using GANs

Jin Huang (huang86)

CS230 - Deep Learning, Stanford University

Introduction

- Anime images and videos are at most 1080p. The old classic ones are even at 480p.
- Anime resides in a lower dimension in terms of color and features, usually with low frequency content. It’s a different and unique field to focus on. Most importantly, I’m an anime fan.
- The best modern approach for super resolution problem is SRGAN. New solutions show up to improve the training of GANs, such as WGAN with gradient penalty.

Data

- Only images larger than 512 x 512.
- 11228 random images from kaggle dataset [https://www.kaggle.com/ myanimelist/dragon-ball-illustrations/butaru](https://www.kaggle.com/myanimelist/dragon-ball-illustrations/butaru)
- Train/Dev/Test set sizes are 108K/2K/2K.

Preprocessing

- CenterCrop to 2x4 x 2x4
- RandomCrop to 128 x 128 (SRGAN) or 96 x 96 (SRWGAN-GP) - Original HR
- Bicubic downsample to 32 x 32 (SRGAN) or 24 x 24 (SRWGAN-GP) - LR Input

Method

Generator

- Input & output pixel values [0, 1].
- Same as the last layer.
- Using R-residual blocks, ref. [8].

Generator Adversarial Loss

\[ \frac{1}{m} \sum_{i=1}^{m} \text{L}_{\text{Gan}} \]

SRGAN-GP

\[ \frac{1}{m} \sum_{i=1}^{m} \text{L}_{\text{Gan}}^{\text{GP}} \]

Fig. 1: Generator Network Details

SRGAN Discriminator

- Discriminator network.
- Input & output pixel values [0, 1].
- SRGAN-GP has no BatchNorm layers.
- SRGAN-GP has no sigmoid.
- Batch normalization creates correlation between samples in the same batch. It impacts the effectiveness of the gradient penalty.

SRGAN Discriminator Loss

\[ \frac{1}{m} \sum_{i=1}^{m} \text{L}_{\text{GAN}} + \frac{1}{m} \sum_{i=1}^{m} \text{L}_{\text{Grad}} \]

SRWGAN-GP Discriminator Loss

\[ \frac{1}{m} \sum_{i=1}^{m} \text{L}_{\text{GAN}}^{\text{GP}} + \frac{1}{m} \sum_{i=1}^{m} \text{L}_{\text{Grad}}^{\text{GP}} \]

Fig. 4: SRGAN Discriminator Network Details

Fig. 5: SRWGAN-GP Discriminator Network Details

Training

Fig. 6: SRGAN Training

Fig. 9: SRWGAN-GP Training

HyperParameters

- 500 epochs.
- Default initializations from PyTorch.
- ADAM: betas = (0.9, 0.999), epsilon = 1e-8
- SRGAN uses learning rate 1e-4.
- SRWGAN-GP uses 1e-4.
- SRGAN has label smoothing.
- SRWGAN-GP uses 1e-3.

Fig. 7: Losses

Fig. 8: Generator Loss / Adversarial Loss

Findings

- SRGAN-GP’s MSE loss decreases smoothly.
- SRGAN-GP has stable gradients.
- SRGAN-GP converges faster.
- SRGAN gets higher PSNR and SSIM at peak.

Fig. 11: Original HR / SRWGAN-GP / SRGAN / Deep RockNet

- Deep RockNet gets very blurry results with high PSNR and SSIM.
- SRGAN results are more blurry than SRWGAN-GP ones by average.
- SRGAN fails in more cases. SRWGAN-GP produces good results even when SRGAN fails. See the last row in Fig. 10.

Conclusions

- Anime and reality photos are on different distributions. Anime Super Resolution is a different problem from realistic photo Super Resolution.
- SRGAN is working on Anime Super Resolution without features from pre-trained models.
- SRWGAN-GP overall outperforms SRGAN on Anime Super Resolution.

Future Work

- Explore more GAN training tricks.
- Experiment with other GANs.
- Work on Anime video Super Resolution. Use RNN to get consistent video frames.