Human Portrait Super Resolution using GANs

Yujie Shu
yujieshu@stanford.edu

1 Introduction

Super resolution can be useful in
- surveillance application
- medical imaging
- satellite image analysis
- photo recovery

Since human precision has a higher standard for faces and can really tell the nuances, we explore different neural network models and train a neural network to do super resolution to match this level of expectation.
- SRGAN is the state-of-the-art in SR
- PGGAN is the state-of-the-art in generating HD faces
- Transfer learning
- Mix loss of perceptual VGG loss and MSE loss
- Our models of SRWGAN-GP and SRPGGAN generates photo-realistic results

We build 4 models from scratch:
- SRResnet as baseline model
- SRGAN
- SRWGAN-GP
- SRPGGAN

2 Data

Dataset:
- CelebA-HQ dataset
- 30k 1024x1024 celebrity images
- Most are cropped front face portrait

Data Preprocession:
- Preprocessed to different resolution: 512x512, 256x256, 128x128, 64x64, 32x32, 16x16
- Split into 28000/1000/1000 as train/dev/test sets
- Choose input LR size of 32x32

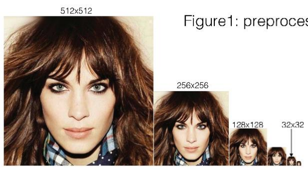

Figure1: preprocessing

3 Methods

SRResnet	SRGAN	SRWGAN-GP	SRPGGAN
Resnet	NSGAN	WGAN-GP	PGGAN
Skip connection	SRResnet as G	transfer learning	WGAN-GP
MSE loss	Mix loss	Mix loss	Mix loss

MSE loss and perceptual VGG loss

$$l^{SR}_{MSE} = \frac{1}{r^2WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}(I^{HR}_{x,y} - G_{\theta_G}(I^{LR})_{x,y})^2 \quad (1)$$

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

SRGAN

$$\min_{\theta_G}\max_{\theta_D} \mathbb{E}_{I^{HR}\sim p_{train}(I^{HR})}[\log D_{\theta_D}(I^{HR})] + \mathbb{E}_{I^{LR}\sim p_G(I^{LR})}[\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR})))] \quad (3)$$

SRWGAN

$$\min_{\theta_G}\max_{\theta_D} \mathbb{E}_{I^{HR}\sim p_{train}(I^{HR})}[D_{\theta_D}(I^{HR})] - \mathbb{E}_{I^{LR}\sim p_G(I^{LR})}[(D(G_{\theta_G}(I^{LR})))] \quad (4)$$

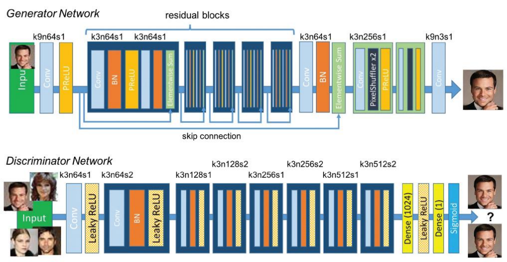

Figure 2: SRGAN model and SRResnet using Generator Architecture

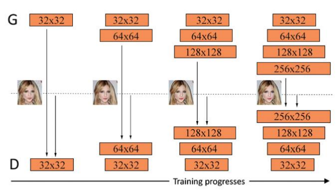

Figure 3: SRPGGAN model

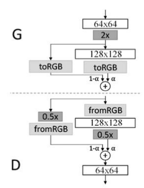

Figure 4: transition

4 Results

- SRResnet has highest PSNR but overly smooth texture
- SRGAN has better result than SRResnet
- SRWGAN-GP and SRPGGAN has better result than SRGAN

Model	PNSR
SRResnet	25.10
SRGAN	24.57
SRWGAN-GP	23.67
SRPGGAN	23.29

Table 1: PSNR Comparison

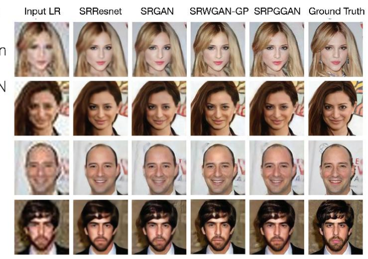

Figure 5: SR with 4x scaling Output Comparison

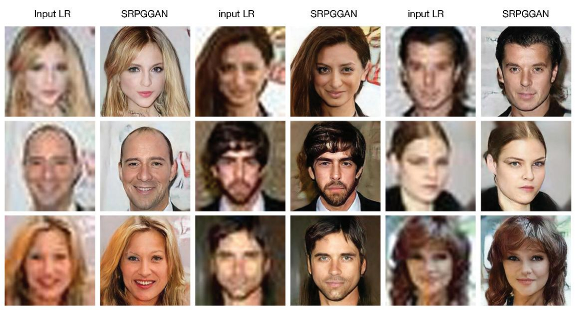

Figure 6: SRPGGAN 8x scaling factor generates photo-realistic results

5 Conclusion

- Successfully implemented SRGAN and SRResnet models with 4x scaling
- Extended SRGAN to SRWGAN-GP with transfer learning to get improved results
- Integrated techniques of PGGAN to grow the generator and discriminator progressively
- Built a SRPGGAN model to do 8x scaling super resolution
- Mixloss of MSE and perceptual VGG loss removed checkerboard artifacts, and
- generated better results than MSE smoothed out images
- Try adding more side view images to the data set for furture work
- More hyperparameter tuning (e.g. VGG scale, adversarial scale, etc)