Super Resolution using GAN

Project Members:
Rachit Shah (rshah25)
Sourabh Sandanshi (ssandan)
Swaraj Meshram (smeshram)

Introduction

Image Super-Resolution is the process of denoising, deblurring and upscaling lower resolution images to generate finer details and make higher resolution clean images.

Low Resolution

High Resolution

Dataset

- DIV2K - used for NTIRE 2017 and PIRM 2018 Competitions
- DIV2K Dataset consists of high resolution of RGB images with a large diversity of contents.
- For Low Res data, an unknown image degradation like blurring, image interpolation is applied and images are scaled down by a factor of 4
- Training Set consists of 800 images
- Validation Set consists of 100 images

Preprocessing

- Images are cached and stored in a tensorflow Dataset.
- Random sections on image of size 96x96 for low res and corresponding for high res are cropped.
- Cropped images are randomly flipped horizontally with a 50% probability
- Flipped images are further rotated by 0, 90, 180, 270 degrees depending on a random number.
- This helps reduce the input data size and model training time. It also helps create a more generalized model

Baseline CNN

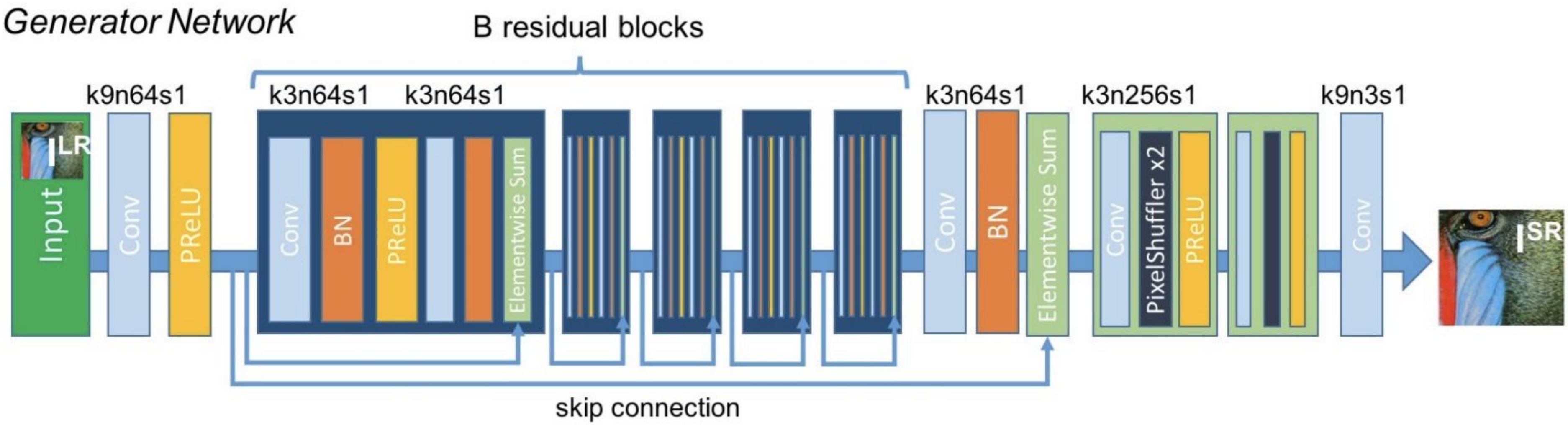
- We normalize the input and denormalize the output using the RGB Mean of the dataset
- There are 2 Convolution Layers with 64 filters of size 3x3 in the middle
- We then have 2 upsampling Convolution to make image 4x (super resolve)
- Final Convolution Layer with 3 filters of size 3x3.

Layer (type)	Output Shape	Param #
input_2 (InputLayer)	[(None, None, None, 3)]	0
normalize (Lambda)	(None, None, None, 3)	0
conv2d_5 (Conv2D)	(None, None, None, 64)	1792
conv2d_6 (Conv2D)	(None, None, None, 64)	36928
conv2d_7 (Conv2D)	(None, None, None, 256)	147712
lambda_2 (Lambda)	(None, None, None, 64)	0
activation_2 (Activation)	(None, None, None, 64)	0
conv2d_8 (Conv2D)	(None, None, None, 256)	147712
lambda_3 (Lambda)	(None, None, None, 64)	0
activation_3 (Activation)	(None, None, None, 64)	0
conv2d_9 (Conv2D)	(None, None, None, 3)	1731
denormalize (Lambda)	(None, None, None, 3)	0

Total params: 335,875
Trainable params: 335,875
Non-trainable params: 0

Residual CNN

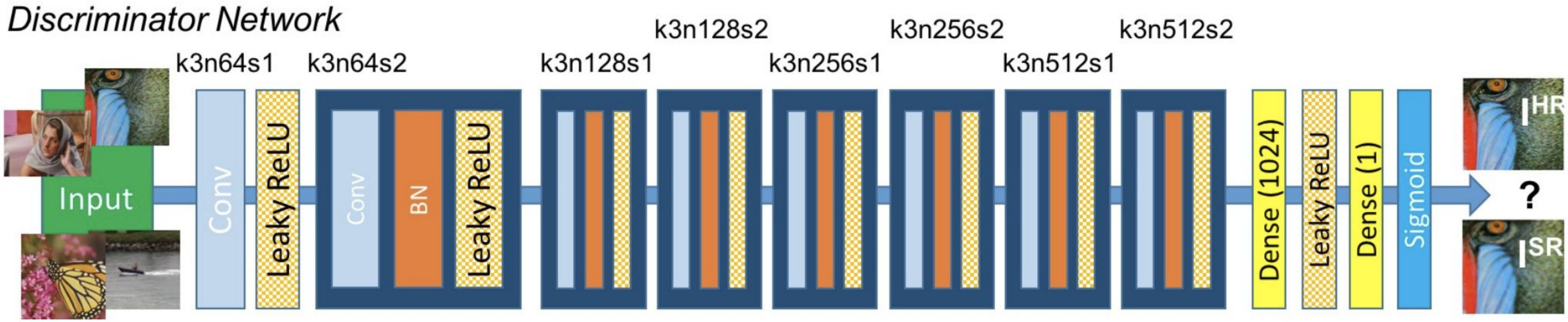
ResCNN with input convolution layer, 16 residual blocks with skip connections, middle convolution, 2x upsampling blocks to super resolve(4x), final convolution layer. We changed some kernel sizes, activation and normalized and denormalized input-output respectively.



Reference: Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

SRGAN

- Use ResCNN as generator & use its weights as pretraining
- We use similar discriminator network as original paper with some changes to the discriminator blocks.
- We added some more blocks with different number of filters and added some skip connections.



Hyperparameters

- We followed multiple papers to decide on Model architecture and hyperparameters
- As GANs take a lot of time to train, We only tuned certain parameters.
- Learning Rate - 0.001 vs 0.0001
- Dense Layer in Discriminator - Present or Absent
- LeakyReLU vs ReLU

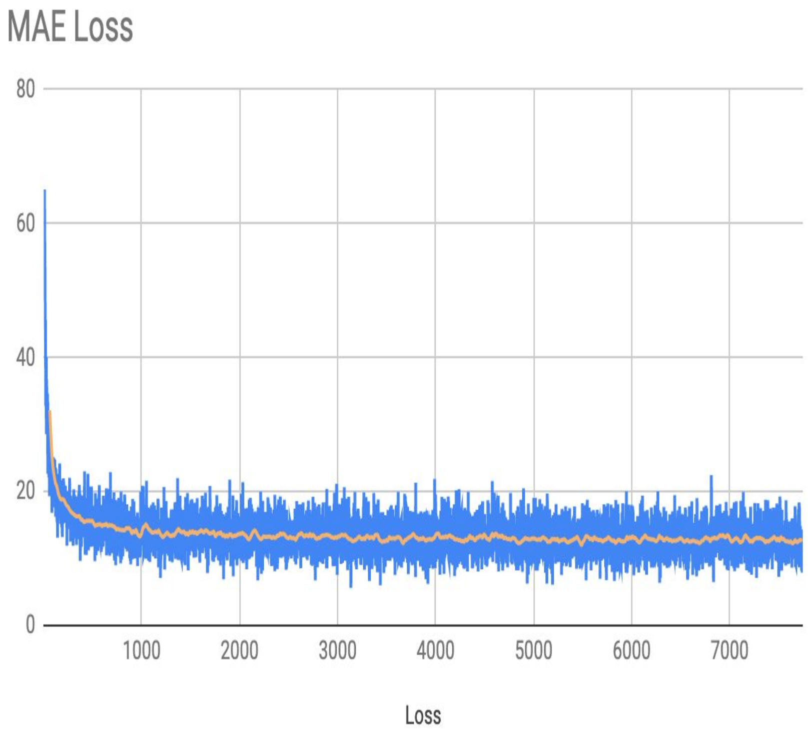
Evaluation Criteria

PSNR - ratio between the maximum possible value (power) of a signal and the power of distorting noise that affects the quality of an image's representation.

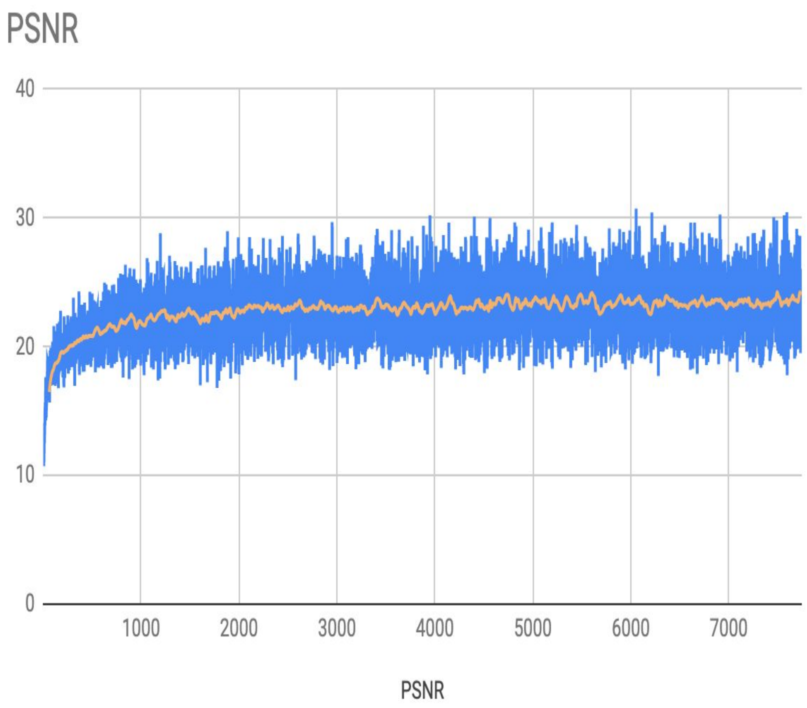
SSIM - reference metric that evaluates noise from a reference image (our original high resolution image) and a processed image (the output of GAN model).

Loss - MAE for CNN, MSE for ResCNN, VGG MSE Loss + 0.001 * Binary Cross Entropy Loss for Generator and Binary Cross Entropy Loss for Discriminator in GAN

Learning Curves CNN

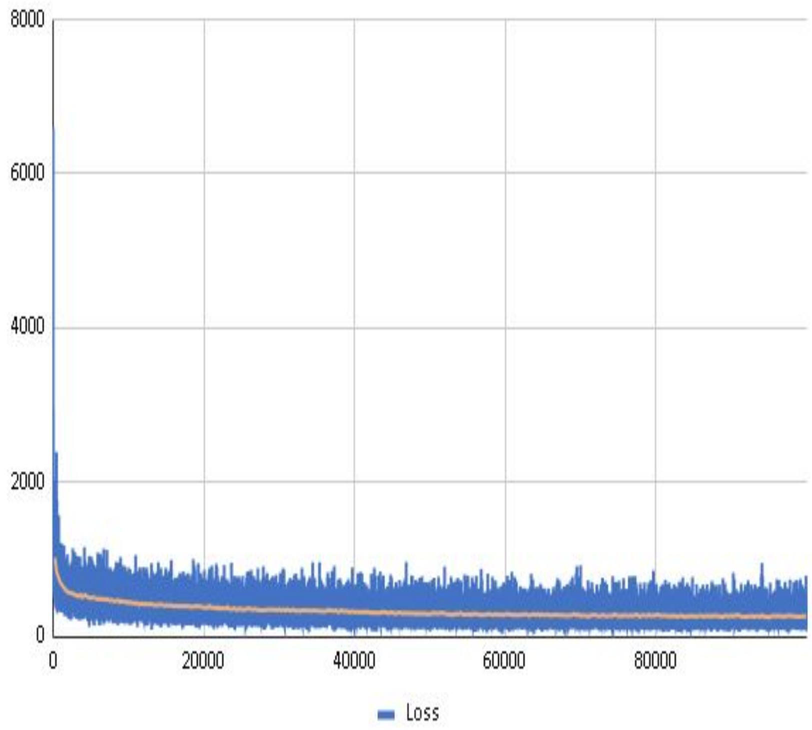


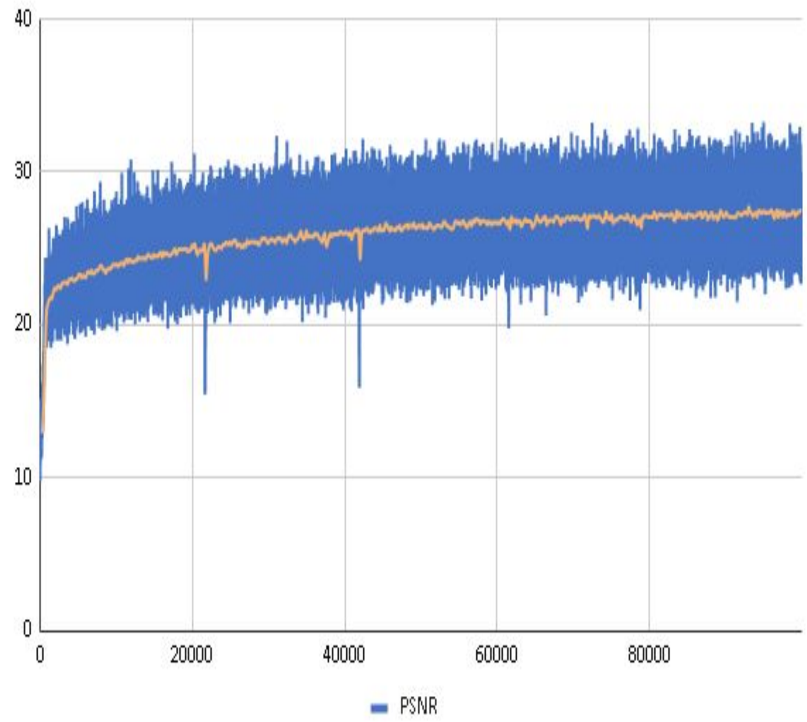
MAE LOSS



PSNR

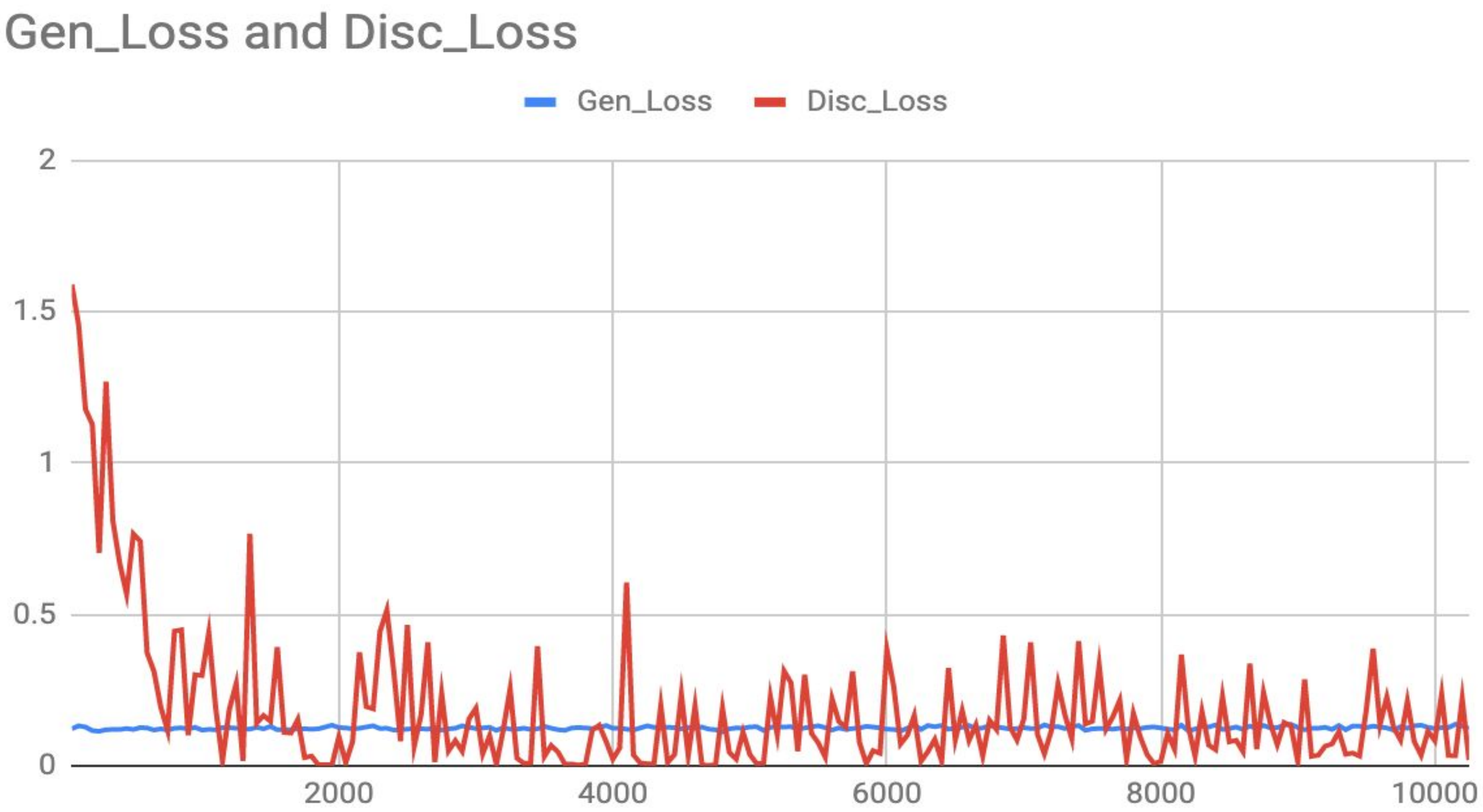
Learning Curves ResCNN





MSE LOSS **PSNR**

Learning Curves GAN



Results

Conclusion

- For most images, we can recreate close to original images
- Baseline CNN marginally better than Low Res
- GAN tends to add more details
- ResCNN is consistent but doesn't have good color reproduction
- PSNR and SSIM are not good indicators when it comes to how humans perceive images
- Results are subjective

Thank You