PROJECT PRESENTATION

GAN-BASED MODEL FOR SUPER-RESOLUTION PROBLEM

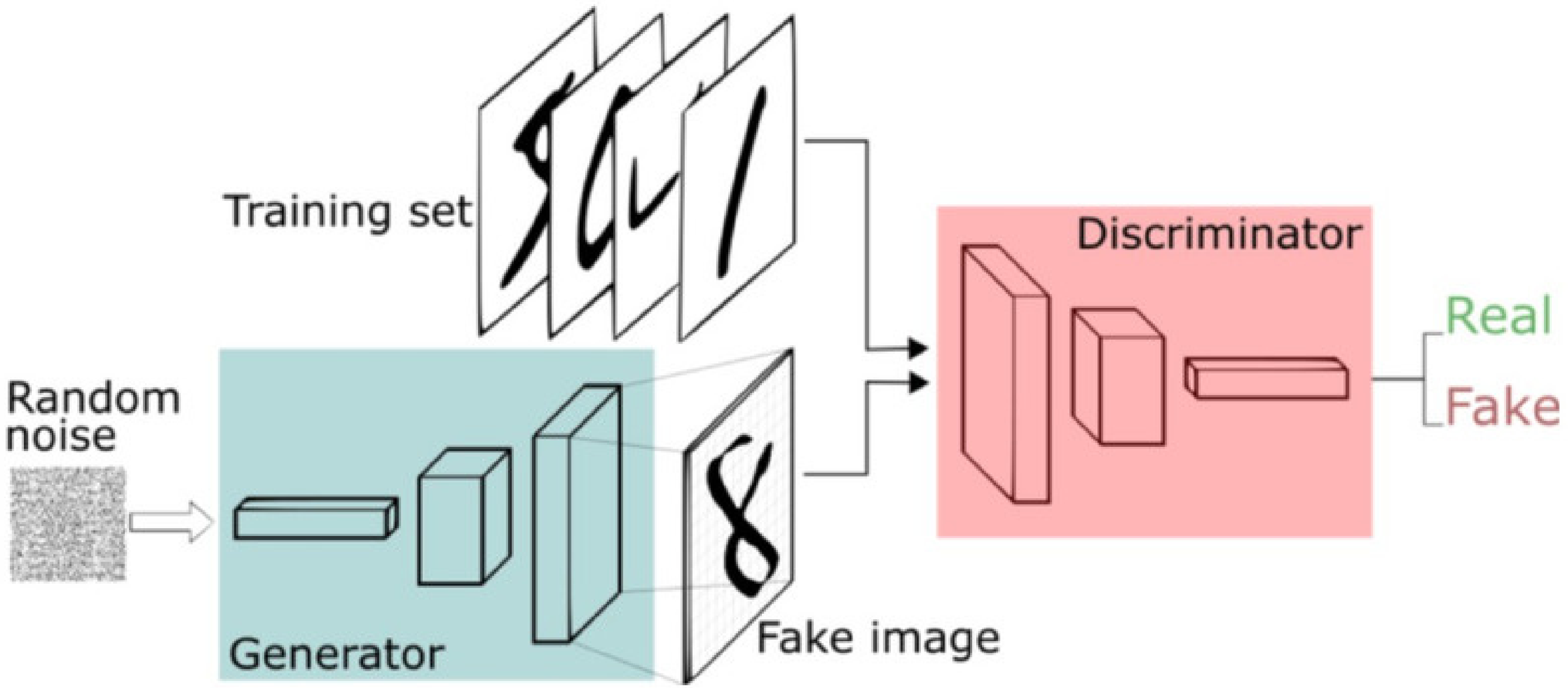
Presenter: Truong Minh Duy

Instructor: Dr. Nguyen Duc Dung

Reviewer: Dr. Tran Tuan Anh

Introduction

Generative Adversarial Networks (GANs) [1]



Generative Adversarial Network (GAN) architecture [2]

Generative Adversarial Networks (GANs) [1]

A powerful generative model:

- Can deal with complex and unstructured data
- Potential to generate a large number of high quality synthetic data

Well applicable in many vision tasks

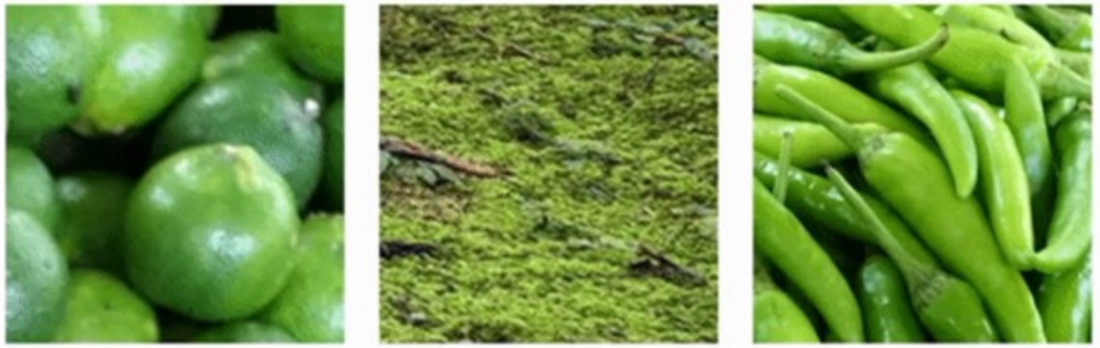
Still have some problems: unstable training, output diversity, etc

Real or Fake?

The photo-realistic image generated by GAN [3]

Single-image super-resolution (SISR)

Reconstruct high-resolution (HR) image from its low-resolution (LR) counterpart



Ill- posed problem

High practical value

Many HR for one LR [4]

Scope

- 4x scale single image super-resolution

- Consider only natural scene pictures

Terminology

- HR/SR/LR image: High-resolution/super-resolution/low-resolution image

- Upsampling/downsampling: The operation to convert image from LR-to-HR/HR-to-LR

2

Related Works

Accuracy and perceptual SR [5]

PSNR/SSIM: 25.85/0.82

PSNR/SSIM: 22.71/0.70

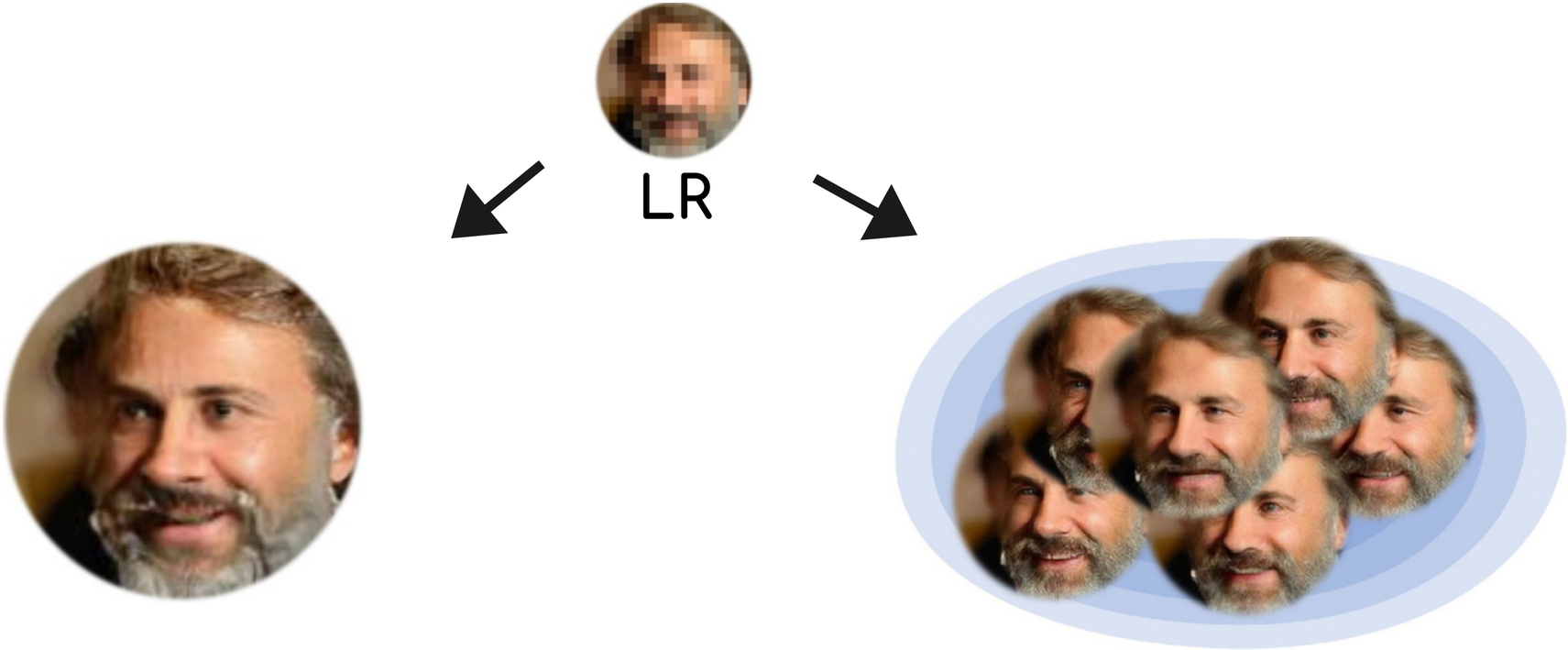
Accuracy

Maximize reconstruction accuracy calculated by mathematical formulation

Perceptual 🎯

Target improve perceptual quality estimated by well-correlated human opinion metrics

Single image and stochastic SR [4]



Single image

Generate exactly one picture

Stochastic

Produce many outputs

GAN-based method for SISR

Most GAN-based model is single image perceptual SR method

They normally improve the visual quality, but suffer from unnatural noise [5]

SRGAN [6]

First framework utilize adversarial loss in SR field

SRFeat [7]

Modify the generator and propose a feature discriminator

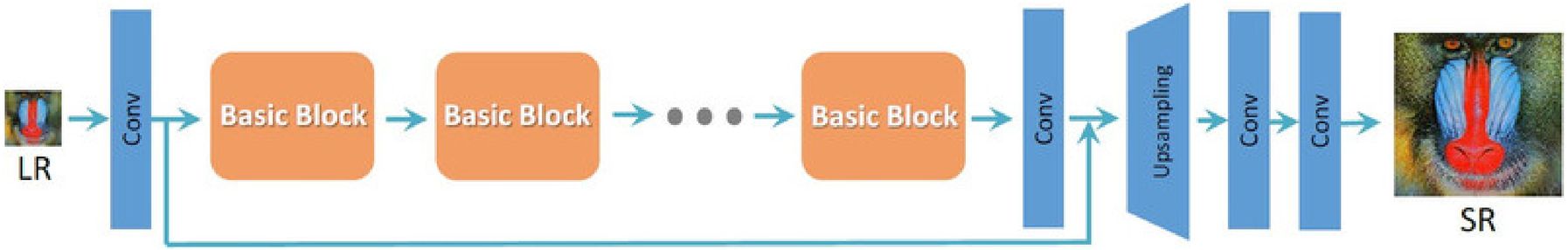
ESRGAN [8]

Our baseline model

ESRGAN: baseline model [8]

Based on its impressive results and our limited resources, we choose ESRGAN as our baseline model

Generator architecture:



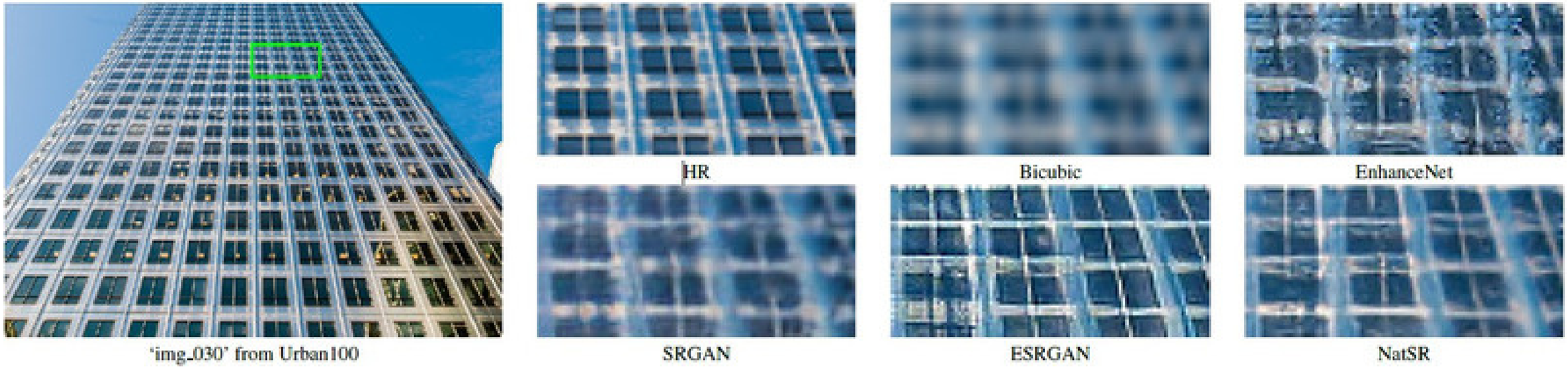
Loss function:

$$L_G^{original} = L_{percep}^{VGG} + \lambda_G \times L_G^{RaGAN} + \eta \times L_1$$

Method

ESRGAN: image quality

Publishing in 2018, ESRGAN is still a **state-of-the-art (SOTA)** model in the perceptual SR field [9].



ESRGAN with other method [9]

Evaluation metrics [5]

Highest priority: NIQE, FID

High priority: LPIPS

Medium priority: PSNR, SSIM

Low priority: BRISQUE

	Full-reference	No-reference
Accuracy-driven	PSNR, SSIM	
Perceptual-driven	LPIPS, FID	BRISQUE, NIQE

Datasets and source code

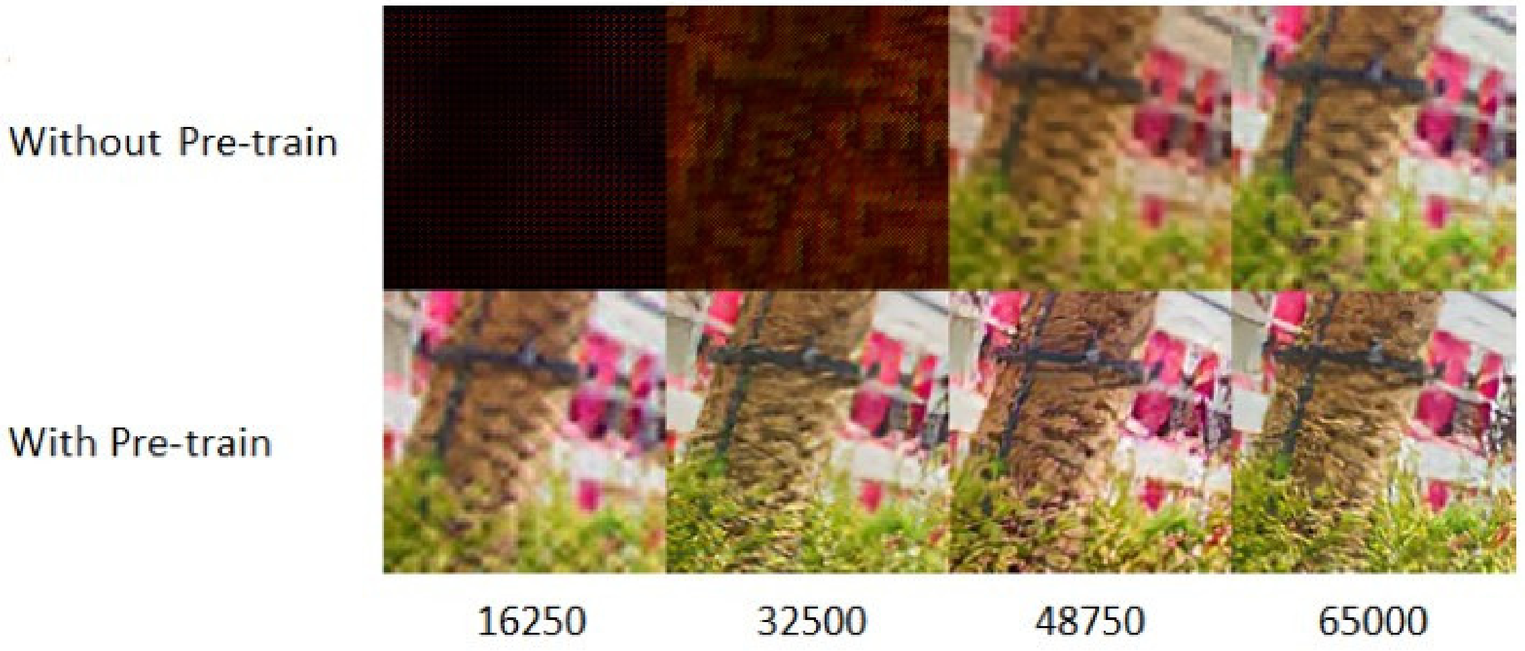
	Quantity	Source	Cropped HR size
Training	3	DIV2K, Flickr2K, OST Training	128 x 128
Validation	1	DIV2K	1024 x 1024
Test	4	Set5, Set14, BSD100, Urban100	From 128 x 128 to 512 x 512

Source code: use unofficial implementation from [4] with some modification

Pixel wise loss

L1/L2 loss reduce noise but prevent the generation of textures [5]

However, we still use L1 loss in our model because of its effectiveness

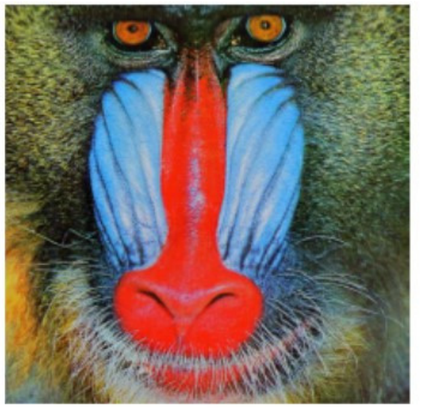


With and without pre-train with L1 loss

Perceptual loss revisited

Bicubic [9]

VGG [9]



HR [9]

ESRGAN uses VGG-based loss, which is trained for image classfication tasks

Thus, it may not the optimal choice for SR problem

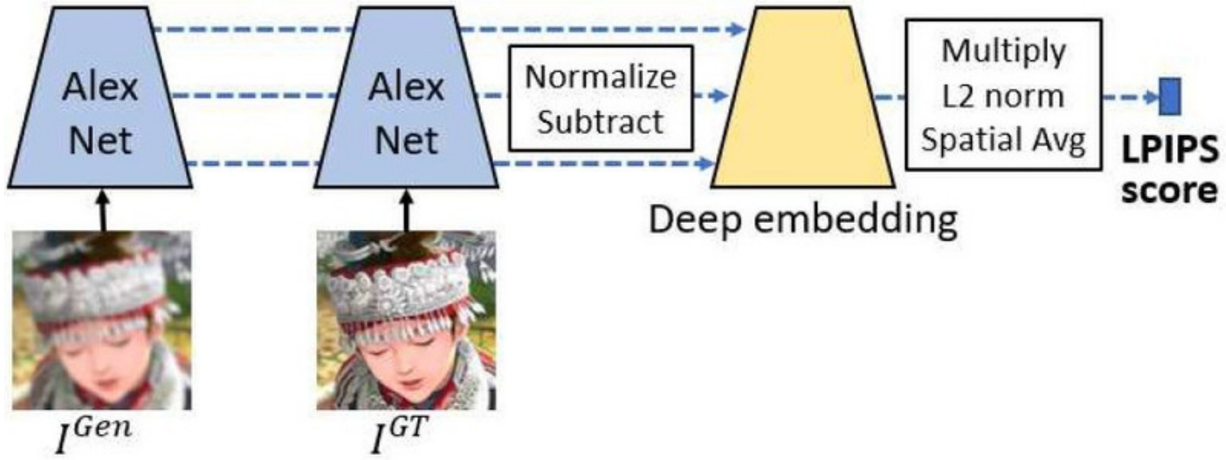
This weakness clearly exposes in a finer scale (16x SR)

Perceptual loss revisited

For human preference , deep features better than the shallow features [10]

A recent studies [10,11] proved that LPIPS [12] and DISTS [13] **rather than VGG** [14] obtain superior results in many vision tasks.

By experiment, we found that LPIPS is a proper choice.



Use LPIPS for SR task [10]

Perceptual loss comparisons

	NIQE ↓	FID ↓	LPIPS ↓	PSNR ↑	SSIM ↑	BRISQUE ↓	Training time ↓
VGG	6.16	28.87	0.2590	24.00	0.63	66.17	**3.5 hours**
LPIPS	4.01	**20.81**	**0.1488**	**24.74**	0.70	**27.58**	**3.5 hours**
DISTS	**2.66**	22.11	0.1576	24.39	**0.71**	33.57	24 hours

Left metric cell means higher priority

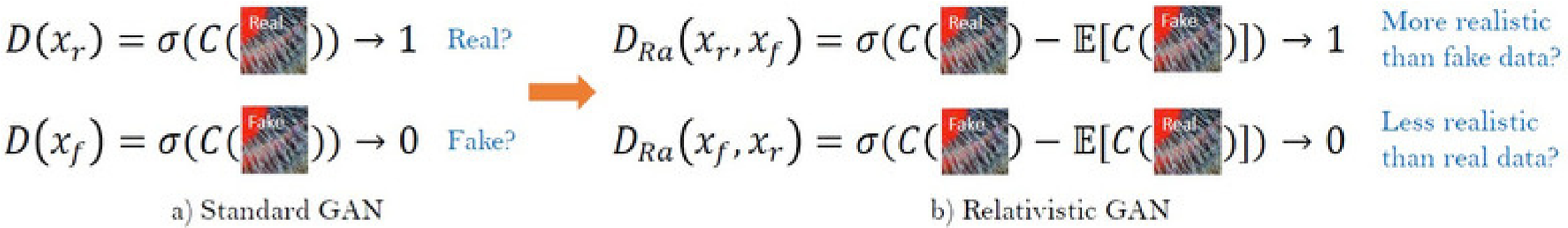
Up arrow: higher is better

Down arrow: lower is better

Bold values highlight the best performance

Relativistic GAN

ESRGAN utilizes relativistic average GAN (RaGAN) [15] to generate more detailed textures compare to standard GAN



Standard and relativistic discriminator [8]

Propose adversarial loss

We propose use the combination of RaGAN [15] and WGANGP [16] called RaGP to obtain more impressive results

$$L_G^{RaGP} = L_G^{RaGAN}$$

$$L_D^{RaGP} = L_D^{RaGAN} + \lambda_g(\left\|\nabla_{\hat{I}} D_{RaGP}(\hat{I})\right\|_2 - 1)^2$$

RaGP formula

Adversarial loss comparison

	NIQE ↓	FID ↓	LPIPS ↓	PSNR ↑	SSIM ↑	BRISQUE ↓
RaGAN	4.01	20.81	0.1488	<u>24.74</u>	**0.70**	<u>27.58</u>
RaHinge	**3.08**	23.37	0.1537	23.40	0.68	36.65
RaLS	3.71	21.92	0.1519	**24.78**	**0.70**	**22.78**
RcGAN	3.59	22.37	0.1565	24.20	<u>0.69</u>	28.69
RcHinge	3.40	<u>20.61</u>	<u>0.1486</u>	24.43	**0.70**	30.96
RcLS	3.49	22.51	0.1524	<u>24.74</u>	**0.70**	29.52
RaGP	<u>3.27</u>	**19.80**	**0.1427**	24.21	0.68	45.21

Bold: the best performance

<u>Underline</u>: the second best

RaGAN and RaGP

Top image: RaGAN - Bottom image: RaGP

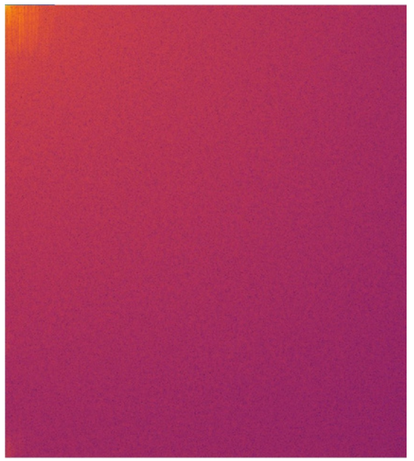
Blue: Generator loss - Orange: Discriminator loss

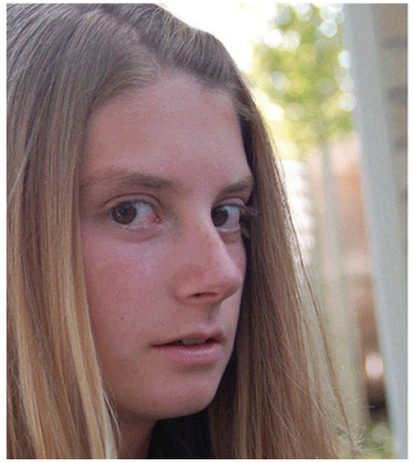
Frequency artifacts

Spatial domain

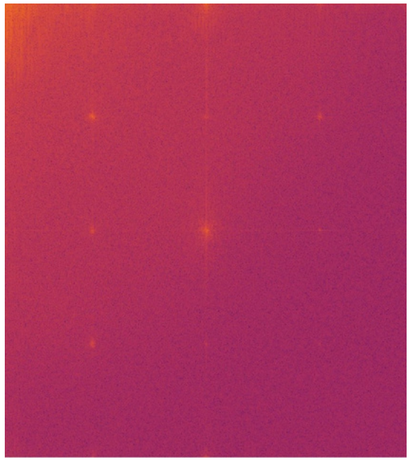
Frequency domain

Real





Fake



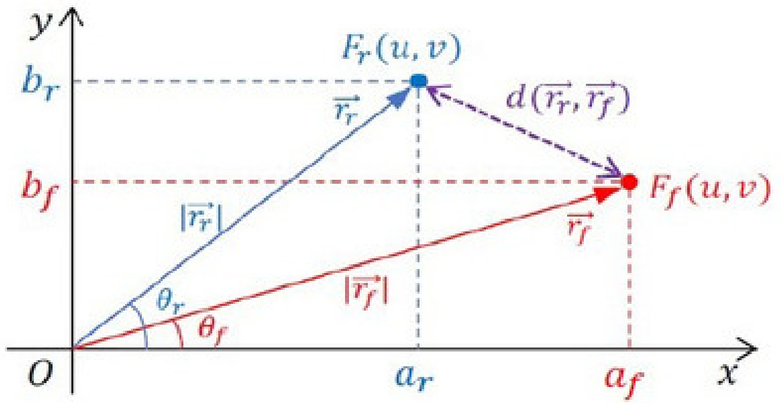
GAN-based models normally work well on spatial domain but reduce their performance on frequency domain [17,18]

Survey

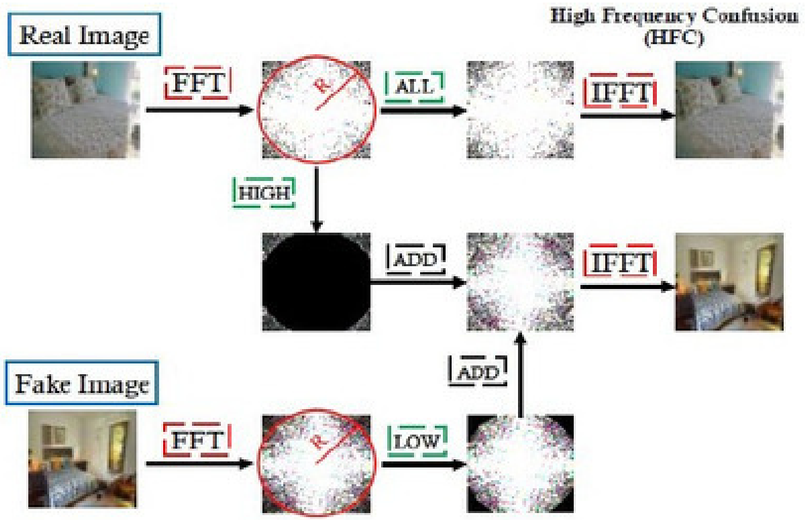
We can group most of current solutions into three main group:

- **Regularize loss function**: spectral loss [19], focal frequency loss [20], etc
- **Change the inputs:** high frequency confusion [21], frequency separation [22], Fourier features mapping [23], etc
- **Use additional module:** SSD-GAN [18], frequency separation network [24], etc

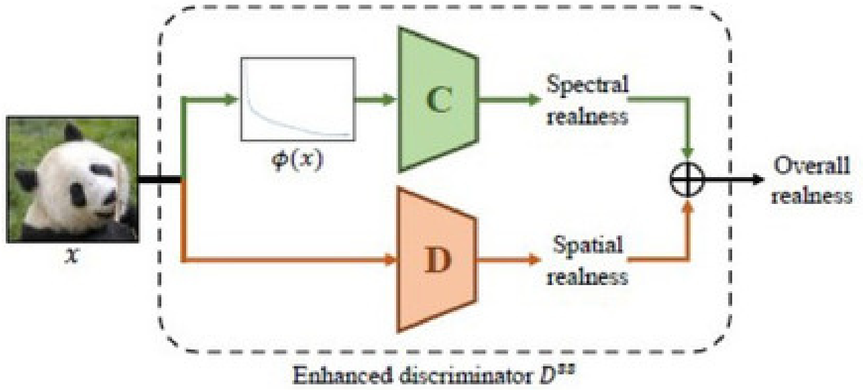
Survey



(a) Regularize loss function



(b) Change the inputs



(c) Use additional module

Three main strategies to alleviate frequency artifacts [18,20,21]

Propose FFT loss

We design a novel FFT loss to alleviate this problem

Comparision with an alternative way called spectrall loss in CVPR 2020 [19], our method outperforms in two aspects:

- We achieve a superior perceptual score
- Our model reconstruct the spectral distribution better

$$L_{FFT} = \frac{\left\| F(G(I^{LR})) - F(I^{HR}) \right\|}{\max(\left\| F(G(I^{LR})) \right\|, \left\| F(I^{HR}) \right\|)}$$

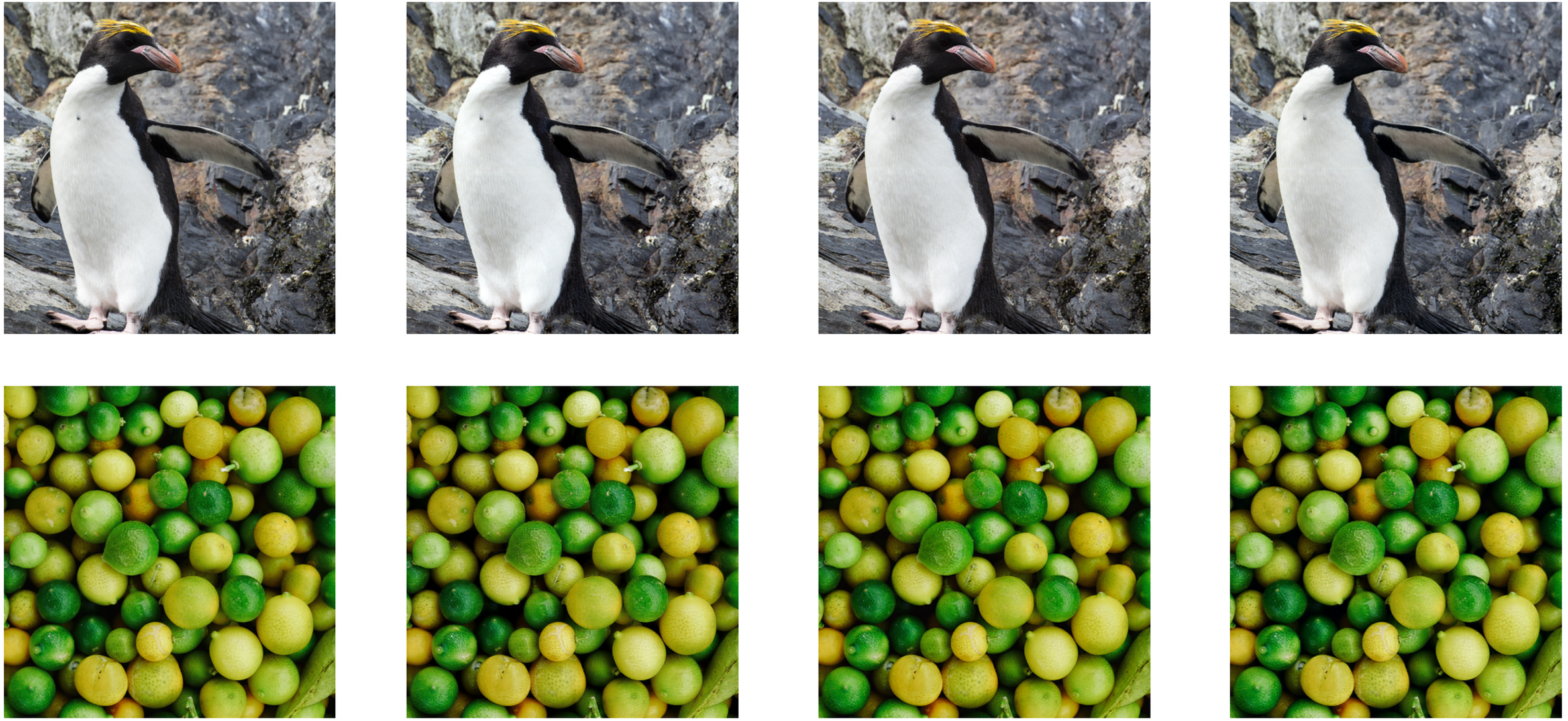
FFT loss formula

FFT vs spectral loss

	NIQE ↓	FID ↓	LPIPS ↓	PSNR ↑	SSIM ↑	BRISQUE ↓
Baseline model						
RaGP	3.27	19.80	0.1427	24.21	0.68	45.21
Hyper = 0.3						
FFT	**3.20**	**17.54**	**0.1410**	**25.67**	**0.74**	**37.99**
spectral	3.44	20.38	0.1457	24.47	0.69	39.32
Hyper = 0.5						
FFT	3.75	**17.73**	**0.1358**	**25.60**	**0.74**	**28.32**
spectral	**3.17**	20.12	0.1491	24.19	0.67	34.10

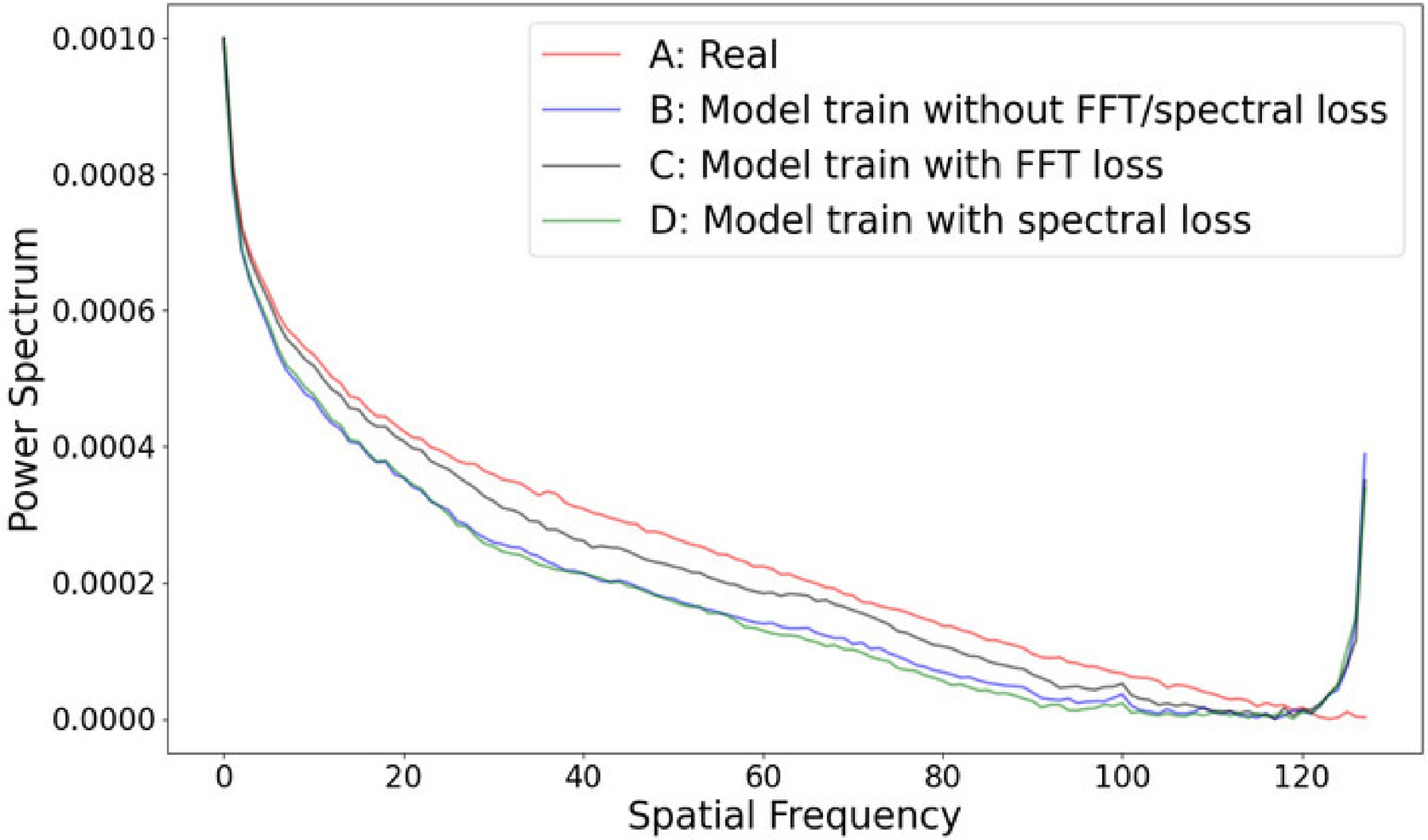
The effect of FFT/spectral loss on the standard setting (RaGP)

Generated images on spatial domain



A B C D

Generated images on frequency domain



The 1D power spectrum of top row images in previous slide

Frequency spectrum discrepancy

	RaGAN	WGANGP	RaGP
original	0.00758	0.00736	0.00801
add FFT loss	0.00535	**0.00291**	**0.00374**
add spectral loss	**0.00521**	0.00840	0.00826

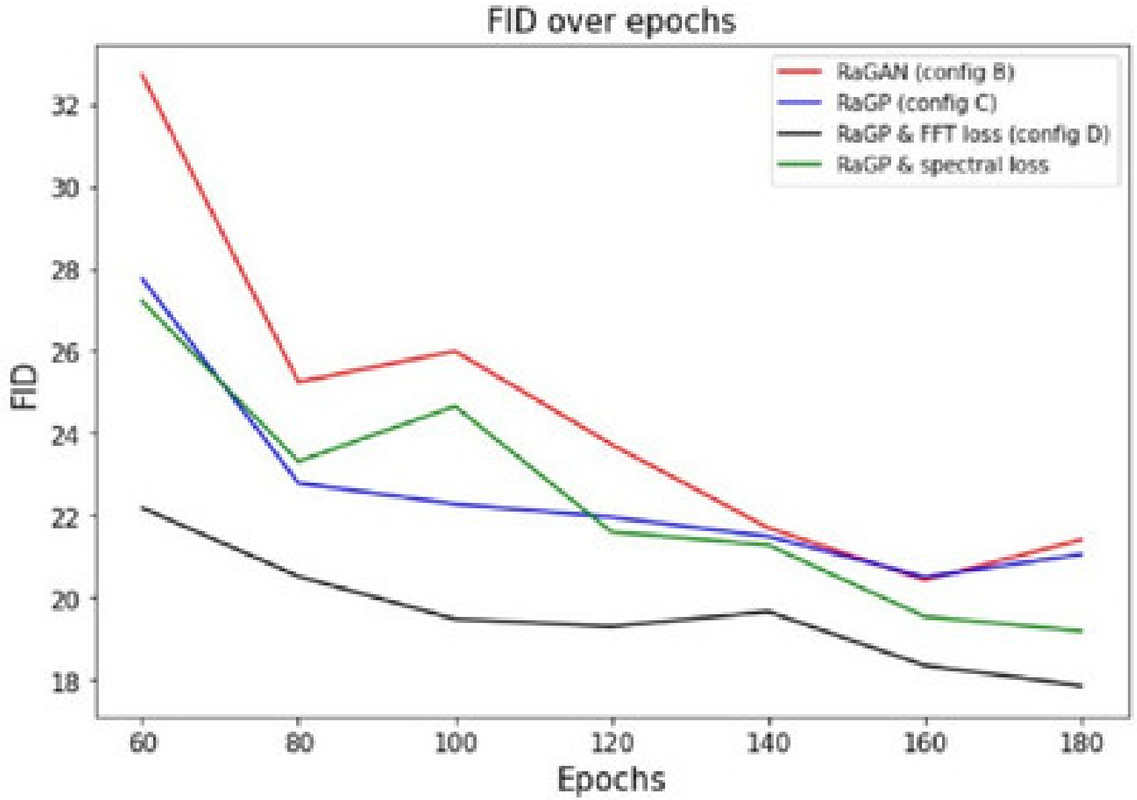
Test on several settings

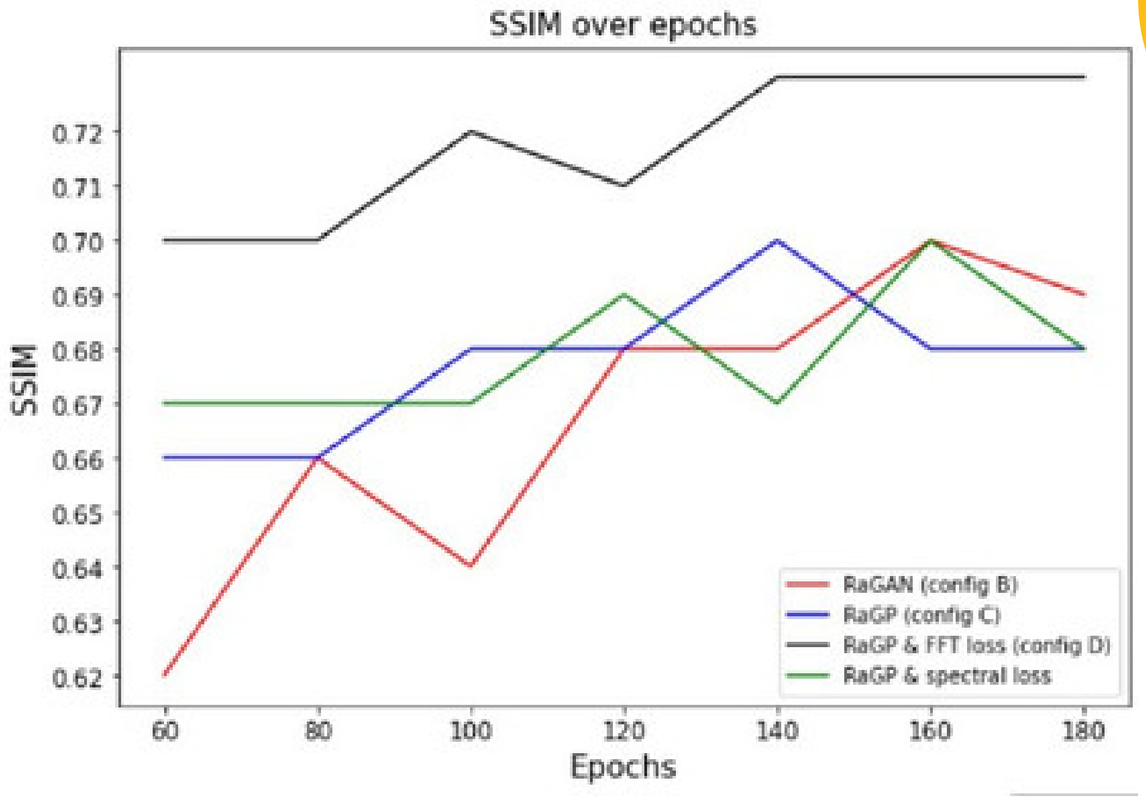
	Set5	Set14	BSD100	Urban100
RaGP	0.00034	0.00066	0.00581	0.00457
add FFT loss	**0.00017**	**0.00031**	**0.00253**	**0.00327**
add spectral loss	0.00033	0.00070	0.00590	0.00465

Test on several datasets

The score (lower is better) reports the difference between 1D power spectrum generating from real and fake images

Metric scores during training procedure





FID↓: lower is better

SSIM↑: higher is better

More experiments

Not only spectral loss [18], our model can achieve better results than

- Focal frequency loss [20]
- High frequency confusion [21], frequency separation with Sinc filter [22]
- SSD-GAN [18]
- Another Fourier transform: DCT

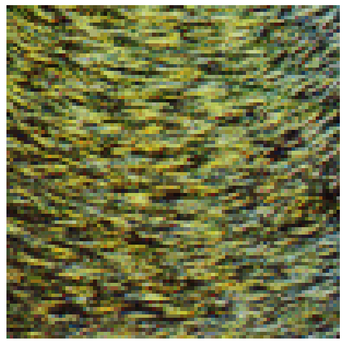
Final results

	NIQE↓	FID ↓	LPIPS ↓	PSNR↑	SSIM ↑	BRISQUE ↓
Set14						
Baseline	5.30	153.30	0.1541	23.86	0.64	52.13
Our	**4.62**	**132.27**	**0.1149**	**25.33**	**0.69**	**48.08**
BSD100						
Baseline	4.79	88.31	0.1945	22.65	0.59	55.29
Our	**4.50**	**78.16**	**0.1514**	**23.66**	**0.64**	**48.38**
Urban100						
Baseline	5.23	45.71	0.1746	20.72	0.62	**59.51**
Our	**4.80**	**30.94**	**0.1294**	**22.10**	**0.68**	59.90

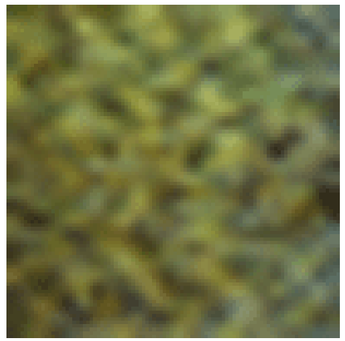
Baseline model and our model on various test sets

Visualization

baboon from Set14



HR



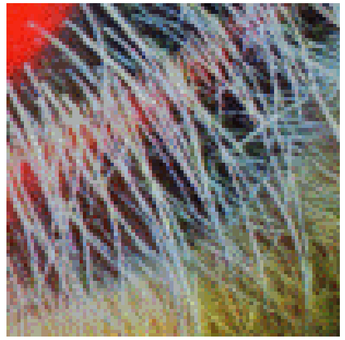
Bicubic

Baseline

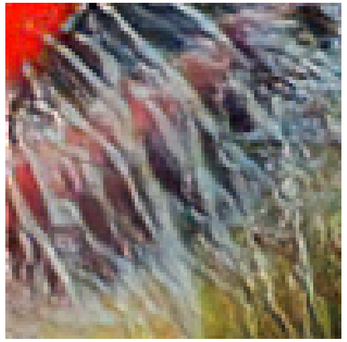
Our

Visualization

baboon from Set14

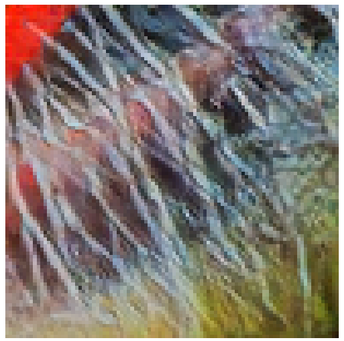


HR



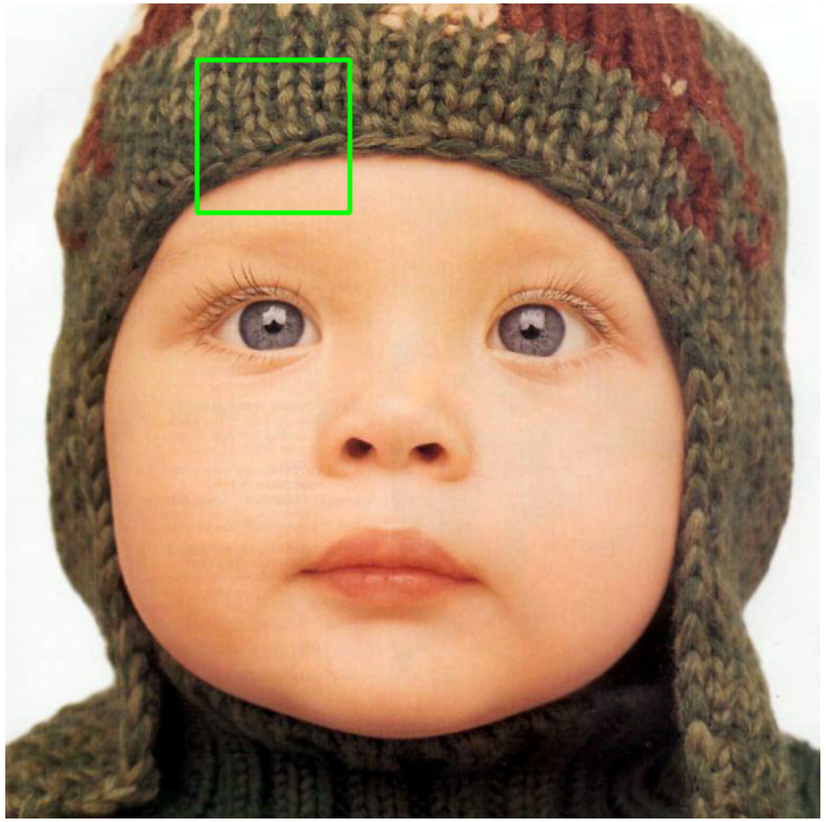
Baseline

Bicubic



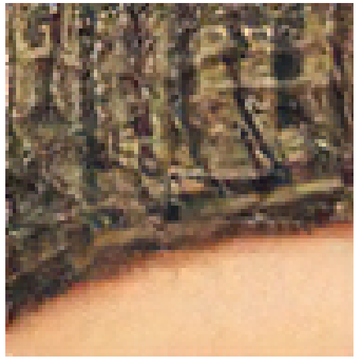
Our

Visualization

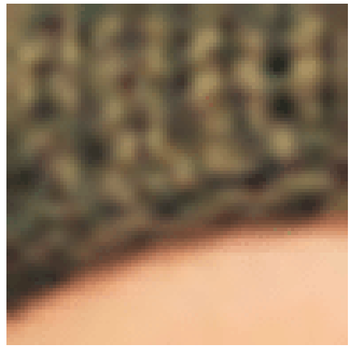


baby from Set5

HR



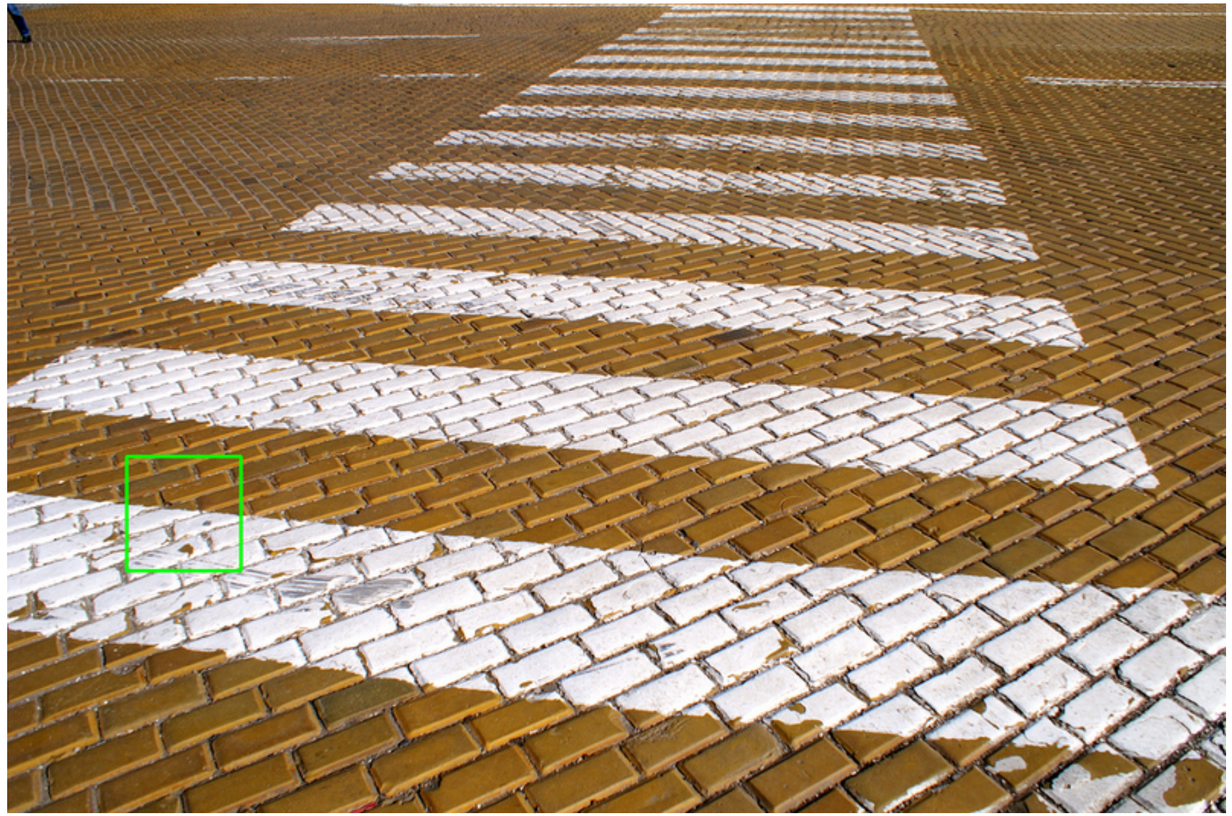
Baseline



Bicubic

Our

Visualization



img_091 from Urban100

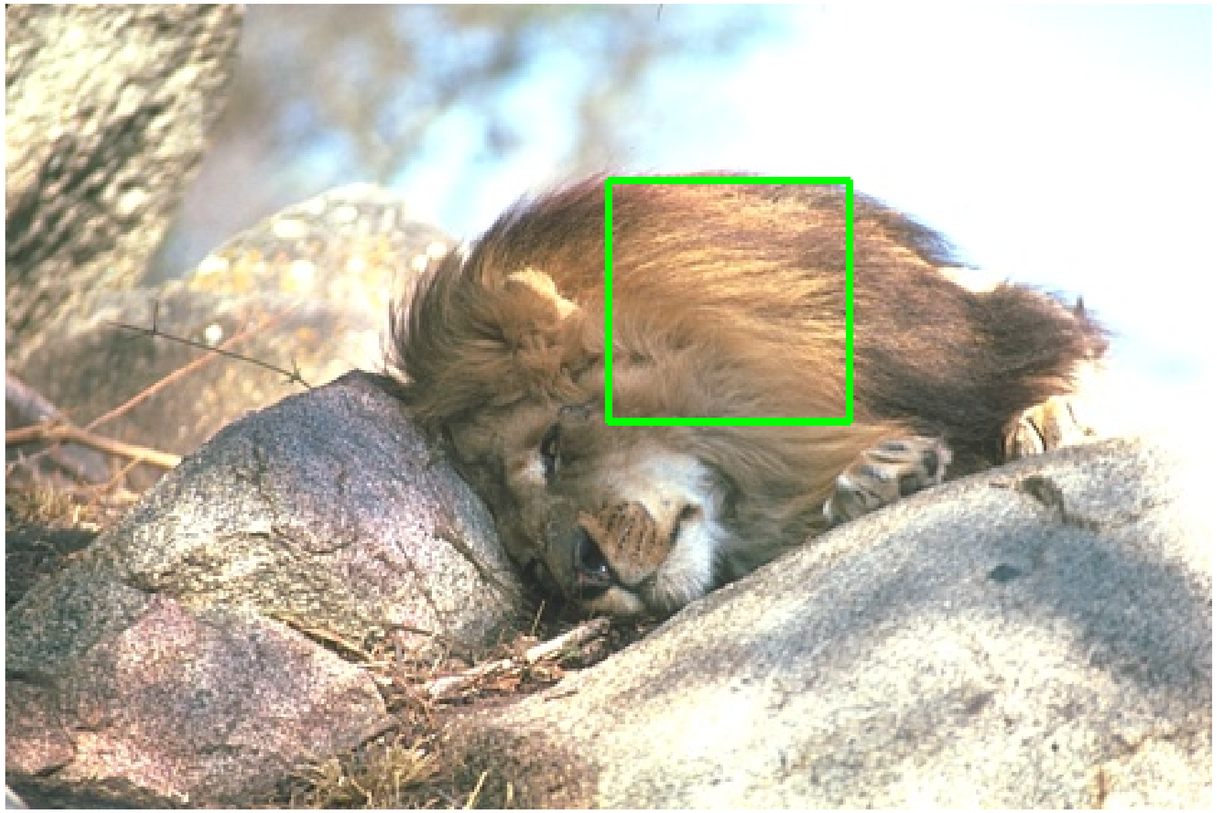
HR

Baseline

Bicubic

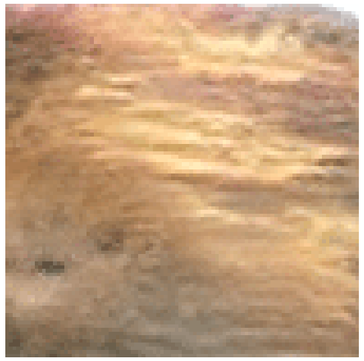
Our

Visualization

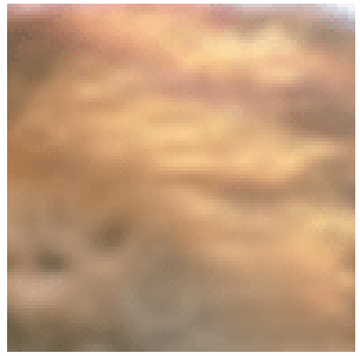


105025 from BSD100

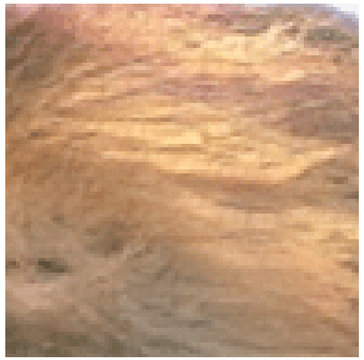
HR



Baseline



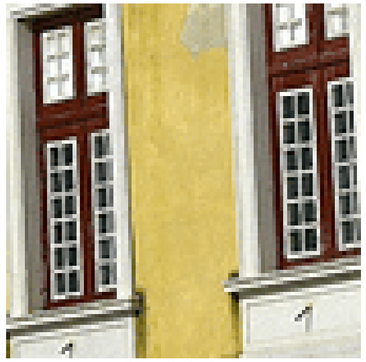
Bicubic



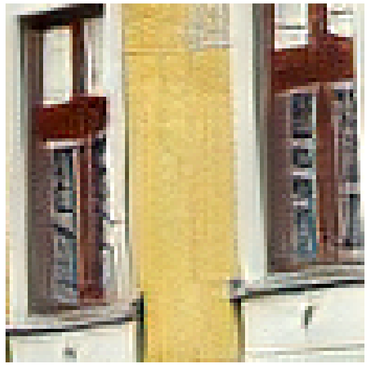
Our

Visualization

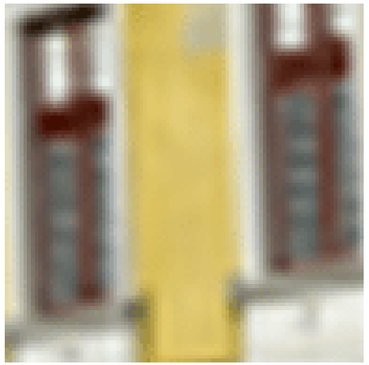
img_054 from Urban100



HR



Baseline



Bicubic

Our

Compare to pretrained model

	NIQE ↓	LPIPS ↓	PSNR ↑	SSIM ↑	Training images	Training iterations
Bicubic	5.20	0.409	26.70	0.77	None	None
EDSR	4.46	0.270	<u>28.98</u>	<u>0.83</u>	800	300K
RRDB	5.08	0.253	**29.44**	**0.84**	14K	1000K
RankSRGAN	**2.45**	0.128	26.55	0.75	800	900K
ESRGAN-pretrain	2.61	0.124	26.22	0.75	14K	1000K
SRFlow	3.57	<u>0.120</u>	27.09	0.76	3450	400K
Our	<u>2.55</u>	**0.109**	26.57	0.76	6K	74K

Note: All batch sizes are 16 and 1K equals 1000

Discussion

Limitation

We cannot find a proper architecture for ESRGAN

Failed experiments: attention mechanism [25,26,27,28], receptive field [29], Realness GAN [30], task-image downscaling [31], capsule network [32], U-Net discriminator [33], etc

Problem

- The model tends to diverge
- The model can only enhance low-priority metric
- The model cannot preserve the performance when scaling

My lessons

- Careful literature review
- Do not too bias in one direction
- Prepare sufficient knowledge and experience
- Catch up with latest trend

References

[1] Ian Goodfellow et al. “Generative Adversarial Nets”. In: *Advances in Neural Information Processing Systems.* Vol. 27. 2014, pp. 2672–2680.

[2] Thalles Silva. 2019. An intuitive introduction to Generative Adversarial Networks (GANs)" url: "https://bit.ly/2YrPvh5" (visited on 10/5/2020).

[3] James Vincent. 2019. url: "https://bit.ly/3iBTMEU" (visited on 12/26/2020).

[4] Andreas Lugmayr et al. “SRFlow: Learning the Super-Resolution Space with Normalizing Flow”. In: *ECCV*. 2020.

[5] Yochai Blau et al. “The 2018 PIRM Challenge on Perceptual Image Super-Resolution”. In: *Proceedings of the European Conference on Computer Vision(ECCV) Workshops*. Sept. 2018

[6] C. Ledig et al. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network". 2017.

[7] Seong-Jin Park et al. “SRFeat: Single Image Super-Resolution with Feature Discrimination”. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. Sept. 2018.

[8] Xintao Wang et al. “ESRGAN: Enhanced Super-Resolution Generative Ad-versarial Networks”. In: *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. Sept. 2018.

[9] Ma et al. "Structure-Preserving Super Resolution With Gradient Guidance". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* June. 2020.

[10] Younghyun Jo, Sejong Yang, and Seon Joo Kim. “Investigating Loss Functions for Extreme Super-Resolution”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June. 2020.

[11] Shiqi Wang, Keyan Ding, Kede Ma and Eero P. Simoncelli. "Comparison of Full-Reference Image Quality Models for Optimization of Image Processing Systems". In: *International Journal of Computer Vision 129 (2021)*, pp. 1258 - 1281

References

[12] Richard Zhang et al. “The Unreasonable Effectiveness of Deep Features as a Perceptual Metric”. In: Conference on Computer Vision and Pattern Recognition (CVPR). 2018.

[13] Keyan Ding et al. “Image Quality Assessment: Unifying Structure and Texture Similarity”. In: *Computing Research Repository (CoRR)*. 2020

[14] Karen Simonyan and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *International Conference on Learning Representations*. 2015.

[15] Alexia Jolicoeur-Martineau. “The relativistic discriminator: a key element missing from standard GAN”. In: *ArXiv* abs/1807.00734 (2018).

[16] Ishaan Gulrajani et al. “Improved Training of Wasserstein GANs”. In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017.

[17] Joel Frank et al. “Leveraging Frequency Analysis for Deep Fake Image Recognition”. In: *Proceedings of the 37th International Conference on Machine Learning*. Vol. 119. June 2020, pp. 3247–3258.

[18] Yuanqi Chen et al. “SSD-GAN: Measuring the Realness in the Spatial and Spectral Domains”. In: *Association for the Advancement of Artificial Intelligence (AAAI)*. 2021.

[19] Ricard Durall, Margret Keuper, and Janis Keuper. "Watch Your Up-Convolution: CNN Based Generative Deep Neural Networks Are Failing to Reproduce Spectral Distributions". In *Conference on Computer Vision and Pattern Recognition (CVPR)*. June. 2020.

THANK YOU