2023 동계 세미나

Blind Super Resolution

Sogang University

Vision & Display Systems Lab, Dept. of Electronic Engineering

Presented by

조영수

Outline

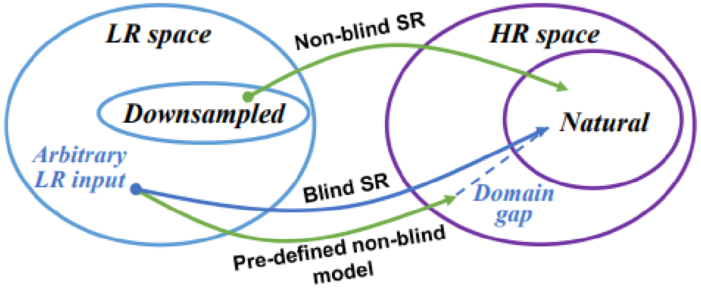
- Introduction
- Background
  - Blind Super Resolution
  - Face Restoration
- Method
  - PDM-SR
    - Learning the Degradation Distribution for Blind Image Super-Resolution (CVPR 2022)
  - ReDegNet
    - From Face to Natural Image: Learning Real Degradation for Blind Image Super-Resolution (ECCV 2022)
- Results
- Conclusion

Introduction

- Super Resolution
  - Restore High-Resolution (HR) image from Low-Resolution (LR) image
  - Ill-posed problem
    - Multiple solution could be obtained from a pixel of low-resolution image
  - According to the number of LR image
    - SISR (Single Image Super Resolution) / MISR (Multi Image Super Resolution)

Figure 1: Example of Single Image Super Resolution

Background



- Blind Super Resolution
  - Recover HR image from the LR image degraded with unknown kernel
  - Degradation Estimation method
    - Real-world LQ images corrupted with unknown degradation parameters
      - Estimating unknown degradation parameter and apply non-blind restoration method
    - Fail due to LQ images with high frequency noises or compression artifacts
  - Data-driven method
    - Main challenge related to the lack of suitable training pairs
      - Capture the paired LQ/HQ images on the same scene by digital camera
      - Learn restoration mapping with unpaired LQ and HQ
        - ✓Learn from only single and suffer from complex degradation
      - Randomly shuffle orders of practical degradation

Background

- Degradation-Learning-Based method
  - Deterministic degradation model
    - Learn the degradation adaptively
      - Reduce the domain gap between real test images and LR images
    - Fail to model the random factor in degradations
  - Probabilistic degradation model
    - Consider the random factors
      - HR images concatenated with random vectors before degraded
    - Unclear relationship between random vectors and degradations

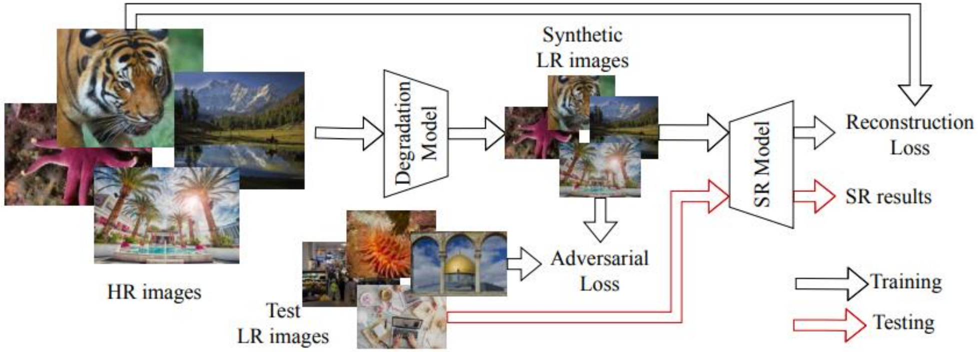
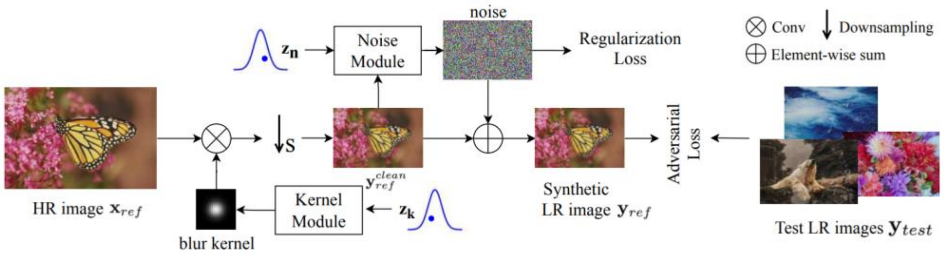


Figure 2: Degradation-learning-based methods

Method



- Problem formulation
  - Probabilistic degradation model (PDM)
    - Parameterize the degradation with two random variables

$$\mathbf{D}(\mathbf{x}) = (\mathbf{x} \otimes \mathbf{k}) \downarrow_s + \mathbf{n},$$

  - Two linear steps
    - Mutually independent
      - Blur kernels with properties of camera lens/noises with properties of sensors

$$\begin{cases} \mathbf{y}^{clean} = (\mathbf{x} \otimes \mathbf{k}) \downarrow_s \\ \mathbf{y} = \mathbf{y}^{clean} + \mathbf{n} \end{cases},$$

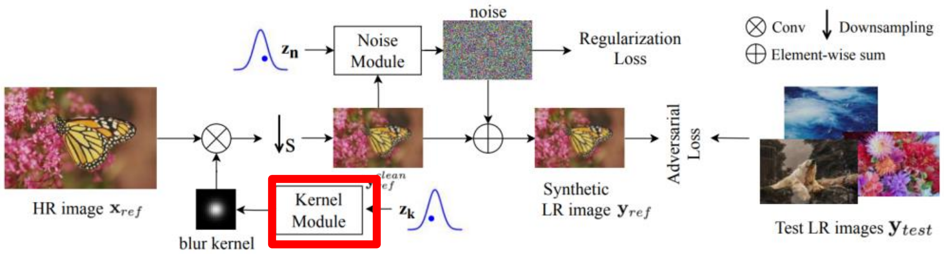
  - Distribution of D
    - Distribution of k and n independently modeled

$$p_D(\mathbf{D}) = p_{k,n}(\mathbf{k}, \mathbf{n}) = p_k(\mathbf{k}) p_n(\mathbf{n}).$$

서강대학교 Sogang University

VDS LAB

Method



- Kernel model
  - Define a priori random variable z_k
    - Generate module to learn the mapping from z_k to k:

$$\mathbf{k} = netK(\mathbf{z}_k) \quad \mathbf{z}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{1})$$

  - Spatially blur kernel
    - Blur kernel for each pixel of x different

$$\mathbf{z}_k \in \mathcal{R}^{f_k \times h \times w} \quad \mathbf{k} \in \mathcal{R}^{(k \times k) \times h \times w}$$

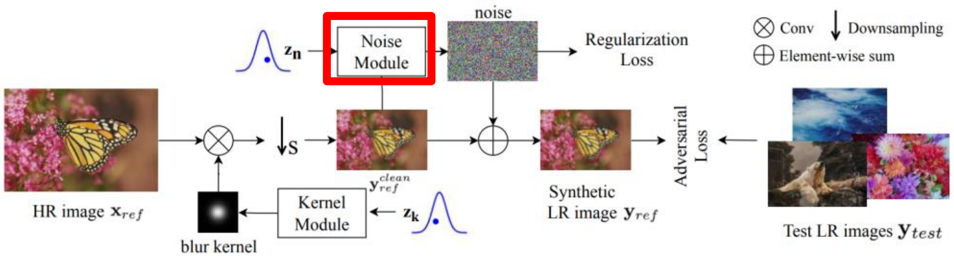
    - General size of convolutional weights 3x3
      - Learned blur kernels spatially correlated
      - Spatial size of convolutional weight set to 1x1
        - ✓Blur kernel independently learned
    - Blur kernel approximated by a spatially invariant one
      - Special case of the spatially variant blur kernel

$$\mathbf{z}_k \in \mathcal{R}^{f_k \times 1 \times 1} \quad \mathbf{k} \in \mathcal{R}^{(k \times k) \times 1 \times 1}$$

서강대학교 SOGANG UNIVERSITY

VDS LAB

Method



- Noise model
  - Adding noise to the blurred and downscaled image y^{clean}
    - Only consider AWGN (additive white gaussian noise)
    - Independent of the content of y^{clean}

$$\mathbf{n} = netN(\mathbf{z}_n) \quad \mathbf{z}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{1})$$

  - Noise in the raw space n_{raw}
    - Combination of shot and read noise
      - Determined by the camera sensors' analog and digital gains

$$\mathbf{n}_{raw} \sim \mathcal{N}(\mathbf{0}, \sigma_{read} + \sigma_{shot}\mathbf{y}^{clean})$$

    - Distribution of n

$$\mathbf{n} = netN(\mathbf{z}_n, \mathbf{y}^{clean}) \quad \mathbf{z}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{1})$$

서강대학교 SOGANG UNIVERSITY

VDS LAB

Method

- Probabilitstic Degradation Model (PDM)
  - Combine the kernel module and the noise module

$$\begin{cases} \mathbf{y}_{ref}^{clean} = (\mathbf{x}_{ref} \otimes netK(\mathbf{z}_k)) \downarrow_s & \mathbf{z}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{1}) \\ \mathbf{y}_{ref} = \mathbf{y}_{ref}^{clean} + netN(\mathbf{z}_n, \mathbf{y}_{ref}^{clean}) & \mathbf{z}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{1}) \end{cases}$$

    - Used to synthesize HR-LR pairs
  - Adversarial training
    - Encourage y_{ref} similar with the test images y_{test}
    - Assumption the noise n with zero mean
      - Added an extra regularizer about noise n
  - Total loss

$$l_{total} = l_{adv} + \lambda l_{reg}$$

Method

- Advantages of using PDM
  1. Able to model more diverse degradations
     - Generate more diverse LR images
       - Better cover the degradations of test images
     - Bridge the gap between training and test datasets
  2. Priori knowledge about the degradations easily incorporate
     - The shape of z_k and k adaptively adjusted based on the observed blur kernel
       - Reduce the learning space of PDM
  3. Formulate as a linear function
     - Better decouple the degradations with image content
       - Focus on learning the degradations
     - Test images heavily blurred
     - Avoid the limitation of extra guidance and focus on learning the degradations

Experiment

- Dataset used

Datasets	Train			Validation	
	LR (used)	HR (used)	paired	LR	HR
2017Track2 [34]	800(401 ∼ 800)	800(1 ∼ 400)	✓	100	100
2018Track2 [35]	800(401 ∼ 800)	800(1 ∼ 400)	✓	100	100
2018Track4 [35]	3200(1600 ∼ 3200)	3200(1 ∼ 1600)	✓	100	100
2020Track1 [26]	2650(1 ∼ 2650)	800(1 ∼ 800)	×	100	100
2020Track2 [26]	2229(1 ∼ 2229)	800(1 ∼ 800)	×	100	0

Paired

Unpaired

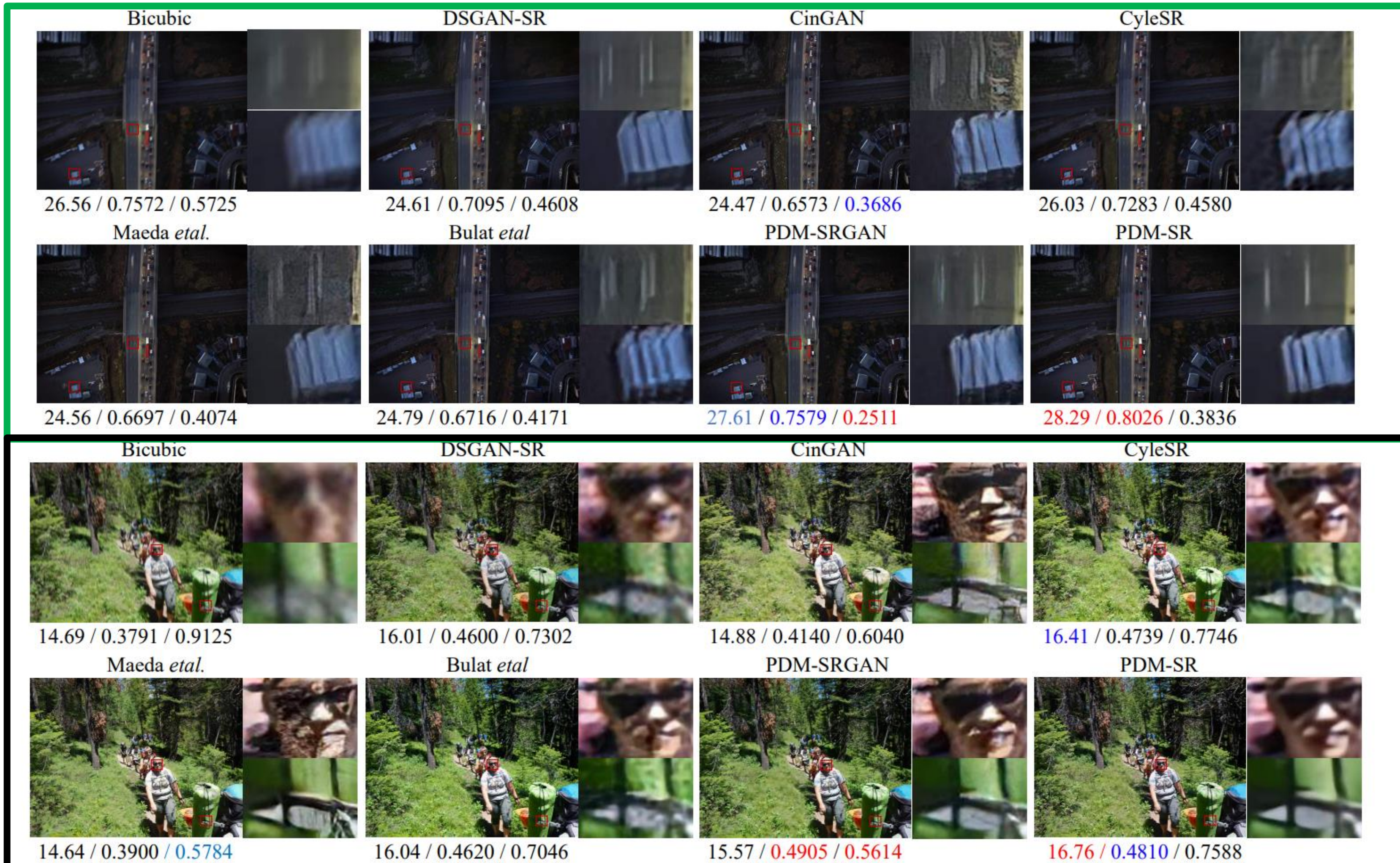
- Comparison between other datasets

Methods	2017Track2			2018Track2			2018Track4			2020Track1		
	PSNR↑	SSIM↑	LPIPS↓	*PSNR↑	*SSIM↑	LPIPS↓	*PSNR↑	*SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Bicubic	21.73	0.5731	0.5430	20.58	0.5304	0.7929	20.26	0.5101	0.7970	25.51	0.6731	0.6414
EDSR [23]	21.58	0.5646	0.4697	20.43	0.5045	0.7778	20.07	0.4771	0.7831	25.34	0.6391	0.6074
DSGAN-SR [12]	20.18	0.5165	0.4314	20.67	0.5829	0.5293	20.26	0.4402	0.5381	23.29	0.6631	0.3295
CinGAN [7]	19.04	0.4451	0.3847	20.10	0.4631	0.4748	20.09	0.4680	0.4903	21.70	0.5814	0.3386
CycleSR [8]	20.70	0.5242	0.4798	21.36	0.5291	0.6390	20.65	0.4980	0.6574	25.48	0.7259	0.3641
Maeda *et al.* [30]	19.23	0.4754	0.3667	19.90	0.4728	0.4897	18.57	0.4085	0.5322	20.06	0.5368	0.4074
Bulat *et al.* [4]	19.84	0.5020	0.4115	20.27	0.4488	0.6668	20.91	0.5408	0.5918	21.49	0.5553	0.4935
PDM-SRGAN	23.43	0.6412	0.2475	20.32	0.5257	0.4074	20.25	0.5307	0.4415	24.56	0.6630	0.2716
PDM-SR	23.69	0.6725	0.3427	20.85	0.5870	0.5240	20.32	0.5611	0.5282	26.80	0.7470	0.3601

Experiment

2017Track2

2018Track2



서강대학교 SOGANG UNIVERSITY

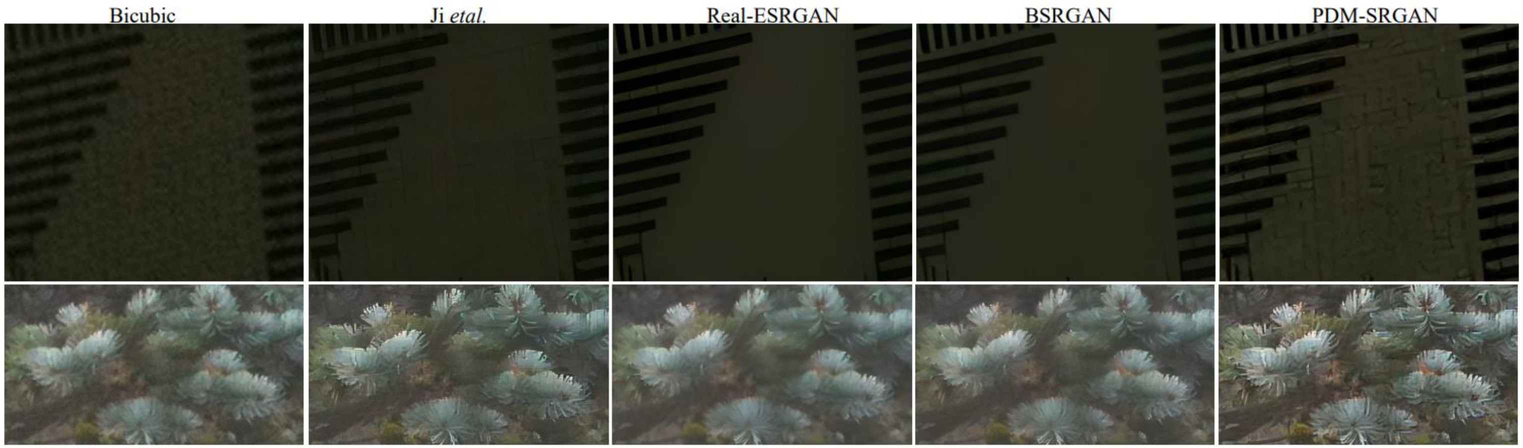
VDS LAB

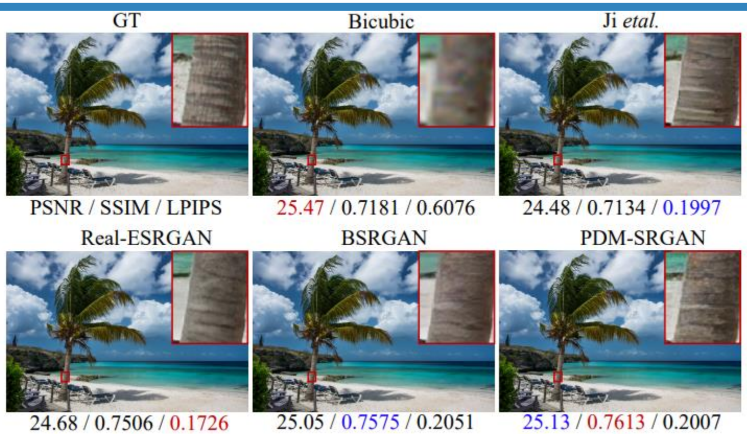
Experiment

2020Track2

2020Track1

- Visual comparison





서강대학교 SOGANG UNIVERSITY

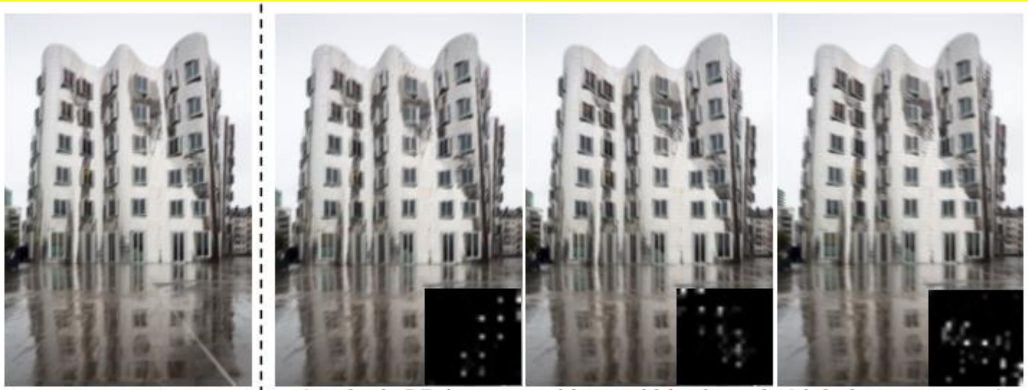
VDS LAB

Experiment

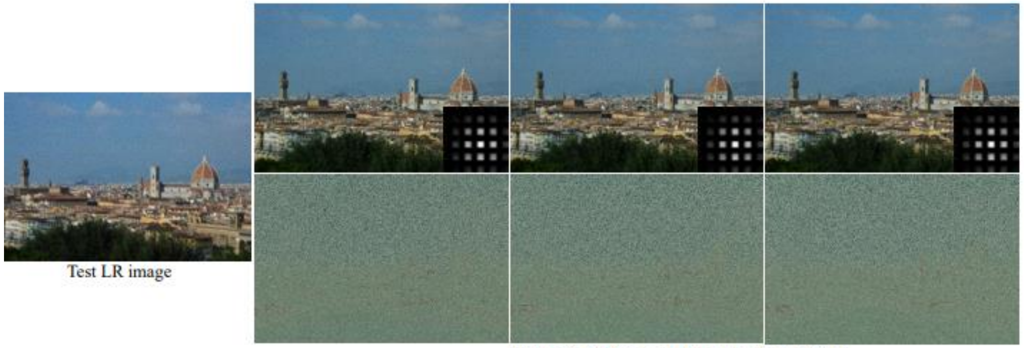
2018Track4

2017Track2

- Learned Degradations
  - Different shapes of blur
    - 2017Track2
      - Asymmetric
    - 2018Track4
      - Symmetric
  - Noise
    - Learned colorful noise
      - Content of the image



Test LR image | Synthetic LR images and learned blur kernels (right bottom corner)



Test LR image

Synthetic LR images and learned degradations

Method

- ReDegNet
  - Learn the real degradation and transfer to natural ones
    - Degradations from the pairs of real-world LQ and pseudo HQ face images
  - Two sub-networks
    - DegNet
      - Learning the degradation representation Ω
    - SynNet
      - Synthesizing the LQ image with the given HQ input and Ω

$$\Omega_f^{Rea} = \mathcal{F}_{Deg}\left(I_f^{RealL}, I_f^{PseH}; \Theta_{Deg}\right)$$

$$\hat{I}_f^L = \mathcal{F}_{Syn}\left(I_f^{PseH}, \Omega_f^{Rea}; \Theta_{Syn}\right) ,$$

Method

- ReDegNet
  - After jointly learn through degradation disentanglement
    - The synthetic realistic LQ natural images obtained

$$\hat{I}_n^L = \mathcal{F}_{Syn}\left(I_n^H, \Omega_f^{Rea}; \Theta_{Syn}\right)$$

Train

Inference

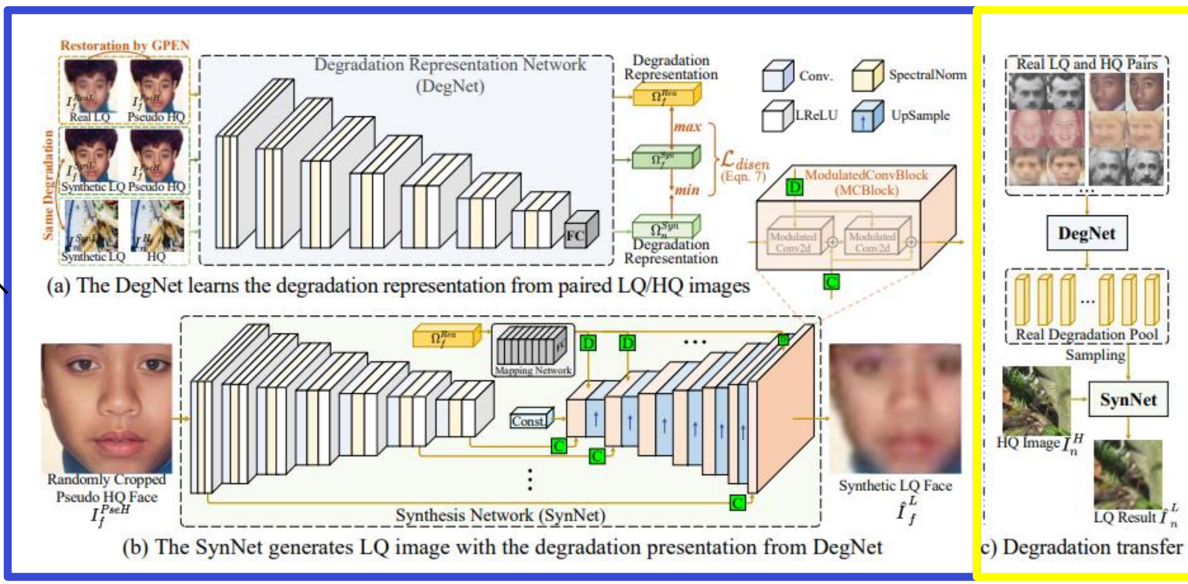
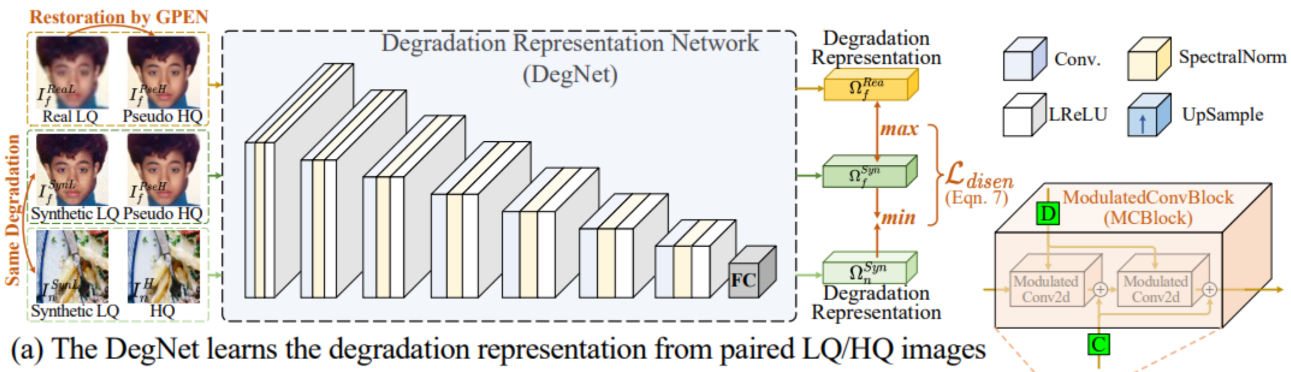


Figure 3: Overview of ReDegNet

Method

- Learning Real Degradation from Face Image
  - Utilized the LQ and HQ pairs
    - Explore the degradation process how the HQ image degraded
  - DegNet
    - Stacked with several convolutional layers
      - Followed by spectral normalization and LeakyReLU
    - Fully convolutional layer at the last
      - Predict the degradation representation vector Ω

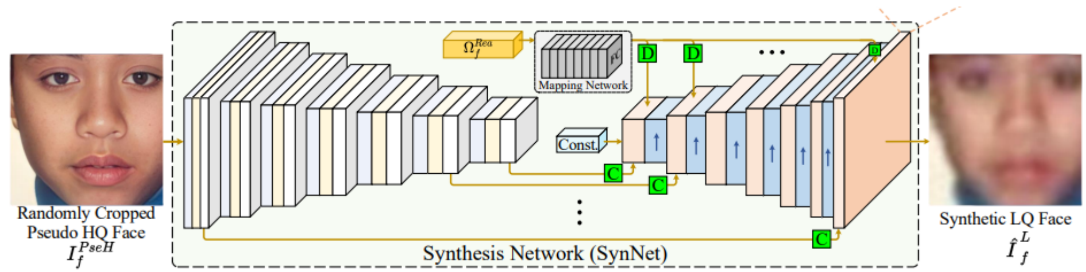


(a) The DegNet learns the degradation representation from paired LQ/HQ images

Figure 4: Overview of DegNet

[1] Karras, Tero, et al. "Analyzing and improving the image quality of stylegan." *CVPR*. 2020.

Method

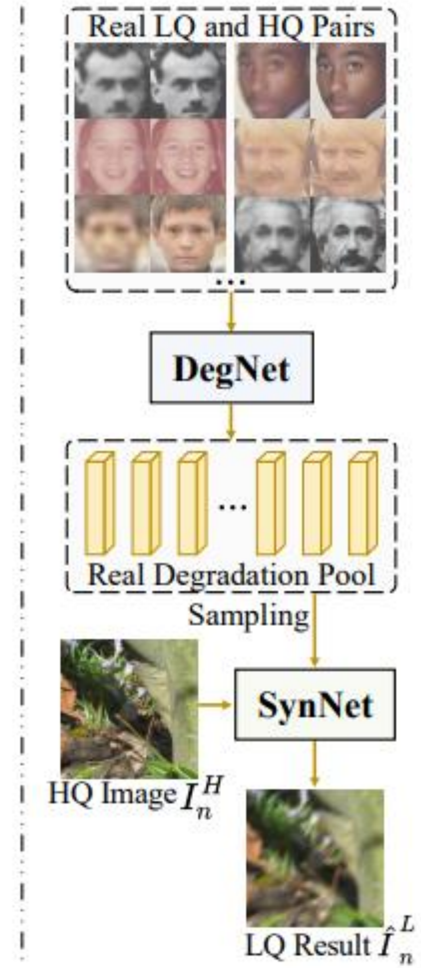


(b) The SynNet generates LQ image with the degradation presentation from DegNet

- Synthesizing the LQ Image
  - Remaining problem
    - How to utilize to control the degradation process
  - Process
    - Inspired by StyleGANs
      - Map the degradation representation Ω to W space
      - Image content of SynNet provided by the features of input HQ images
      - Degraded image reconstructed with modulated convolution operation (MCBlock)
    - With several cascaded MCBlocks
      - The final LQ result synthesized
        - ✓Similar degradation types with the given degradation representation

Method

- Transferring Degradation to Natural Image
  - ReDegNet
    - Extract the real degradation from pairs of face images
    - Generate the corresponding LQ image with the expected degradation styles
  - General face restoration
    - Large degradation representations
      - Extracted from real-world LQ and their pseudo HQ face images
    - Sample from real degradation pool
      - Imitate the real degradation process on natural HQ image
  - Specific restoration
    - Used to fine-tune on the whole image



(c) Degradation transfer

Method

- Learning Objectives
  - Disentanglement loss
    - Extract the degradation-related representations
  - Reconstruction loss
    - Constrain the synthetic results close to the ground-truth
- Disentanglement Loss
  - Control the degradation styles
    - Degradation-aware and content-independent
  - Contrastive learning
    - Minimize the distance of Ωs
      - Image from different content but same degradation parameters
    - Maximizing these negative pairs

Method

- Disengtanglement Loss
  - BSRGAN to control the degradation process
    - Synthesize the degraded face and natural image

$$\Omega_f^{Rea} = \mathcal{F}_{Deg}\left(I_f^{ReaL}, I_f^{PseH}; \Theta_{Deg}\right),$$

Real-world LQ and HQ face pair

$$\Omega_f^{Syn} = \mathcal{F}_{Deg}\left(I_f^{SynL}, I_f^{PseH}; \Theta_{Deg}\right),$$

Synthetic LQ and HQ face pair

$$\Omega_n^{Syn} = \mathcal{F}_{Deg}\left(I_n^{SynL}, I_n^{H}; \Theta_{Deg}\right).$$

Synthetic LQ and HQ natural pair

$$\mathcal{L}_{disen} = \left\|\Omega_f^{Syn} - \Omega_n^{Syn}\right\|_2^2 + \frac{\lambda}{\|\Omega_f^{Syn} - \Omega_f^{Rea}\|_2^2 + \epsilon} + \frac{1}{2}\left\|\Theta_{Deg}\right\|_2^2,$$

Method

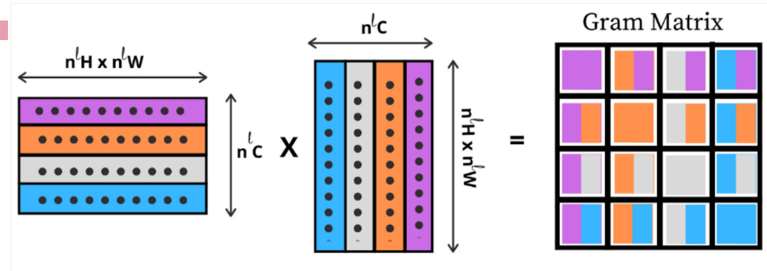


Figure 5: Gram matrix representation

- Reconstruction Loss
  - Mean square error loss $\mathcal{L}_{mse}$+ Realistic loss $\mathcal{L}_{real}$ +Degradation-consistent loss $\mathcal{L}_{cons}$

1. Mean square error loss $\mathcal{L}_{mse}$

Feature space

Pixel space

$$\mathcal{L}_{mse} = \ell_{mse}(\hat{I}_f^L, I_f^{RealL}) = \boxed{\frac{1}{\mathcal{CHW}} \left\| \hat{I}_f^L - I_f^{RealL} \right\|^2} + \boxed{\sum_{i=1}^{4} \frac{0.1}{\mathcal{C}_i \mathcal{H}_i \mathcal{W}_i} \left\| \Phi_i(\hat{I}_f^L) - \Phi_i(I_f^{RealL}) \right\|^2}$$

2. Realistic loss $\mathcal{L}_{real}$
   - Style loss + Adversarial loss
     - Style loss
       - ✓ Applied Gram matrix on the feature spaces of VGG-19

$$\mathcal{L}_{style} = \sum_{i=1}^{4} \frac{1}{\mathcal{C}_i \mathcal{H}_i \mathcal{W}_i} \left\| \Phi_i(\hat{I}_f^L)^T \Phi_i(\hat{I}_f^L) - \Phi_i(I_f^{RealL})^T \Phi_i(I_f^{RealL}) \right\|^2$$

Method

$$\Omega_f^{Rea} = \mathcal{F}_{Deg}\left(I_f^{RealL}, I_f^{PseH}; \Theta_{Deg}\right)$$

$$\hat{I}_f^L = \mathcal{F}_{Syn}\left(I_f^{PseH}, \Omega_f^{Rea}; \Theta_{Syn}\right),$$

- Reconstruction Loss

  2. Realistic loss $\mathcal{L}_{real}$

     - Adversarial Loss
       - Adopt the discriminator from SNGAN
       - Difficult to distinguish with only synthetic LQ image
         - ✓ Additional condition
           - The HQ image and their degradation representation

$$\mathcal{L}_D = -\mathbb{E}[\min(0, -1 + D(I_f^{RealL}, I_f^{PseH}, \Omega_f^{Rea}))] - \mathbb{E}[\min(0, -1 - D(\hat{I}_f^L, I_f^{PseH}, \Omega_f^{Rea}))]$$

$$\mathcal{L}_G = -\mathbb{E}[D(\mathcal{F}_{Syn}(I_f^{PseH}, \mathcal{F}_{Deg}(I_f^{RealL}, I_f^{PseH}; \Theta_{Deg}); \Theta_{Syn}), I_f^{PseH}, \Omega_f^{Rea})]$$

     - Total Loss of Realistic Loss

$$\mathcal{L}_{real} = 0.1 \cdot \mathcal{L}_{style} + \mathcal{L}_G$$

Method

- Reconstruction Loss

3. Degradation-consistent loss $\mathcal{L}_{cons}$

  - Degradation representation Ω_n^{Syn}, Ω_f^{Syn}

    - Reduce the difference between degradation representation

$$\mathcal{L}_{cons} = \ell_{mse}(\underbrace{\mathcal{F}_{Syn}(I_n^H, \Omega_f^{Syn}; \Theta_{Syn}), I_n^{SynL}}_{\text{Face degradation with natural image}}) + \ell_{mse}(\underbrace{\mathcal{F}_{Syn}(I_f^{PseH}, \Omega_n^{Syn}; \Theta_{Syn}), I_f^{SynL}}_{\text{Natural degradation with face image}})$$

- Total Loss

$$\mathcal{L} = \lambda_{disen}\mathcal{L}_{disen} + \lambda_{mse}\mathcal{L}_{mse} + \lambda_{real}\mathcal{L}_{real} + \lambda_{cons}\mathcal{L}_{cons},$$

Result

- Quantitative comparison
  - Real-world Pairs
    - LPIPS performance improved
      - More consistent with human perception
  - Real-world
    - RealSRSet: BSRGAN
    - RealLQSet: Internet and LQ frames extracted from 480p video

Methods	Real-world Pairs									Real-world LQ	
	RealSR-Canon			RealSR-Nikon			DRealSR			RealSRSet	RealLQSet
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	NIQE↓	NIQE↓
RealSR	25.58	.723	.458	**25.49**	.693	.459	**27.69**	**.759**	.438	**4.82**	5.62
BSRGAN	**25.61**	**.768**	.363	24.51	.711	.391	26.64	.744	.380	5.60	5.36
Real-ESRGAN	24.95	**.768**	.366	24.50	**.716**	.388	26.57	.753	.374	5.75	5.24
Ours	25.57	.765	**.362**	25.43	**.716**	**.385**	26.91	.758	**.373**	4.85	**4.93**
Ours (-D)	24.63	.749	.463	24.35	.684	.460	26.32	.740	.425	6.45	6.27
Ours (U)	25.05	.752	.428	24.72	.708	.421	26.35	.741	.404	5.81	5.93

Result

- Visual comparison

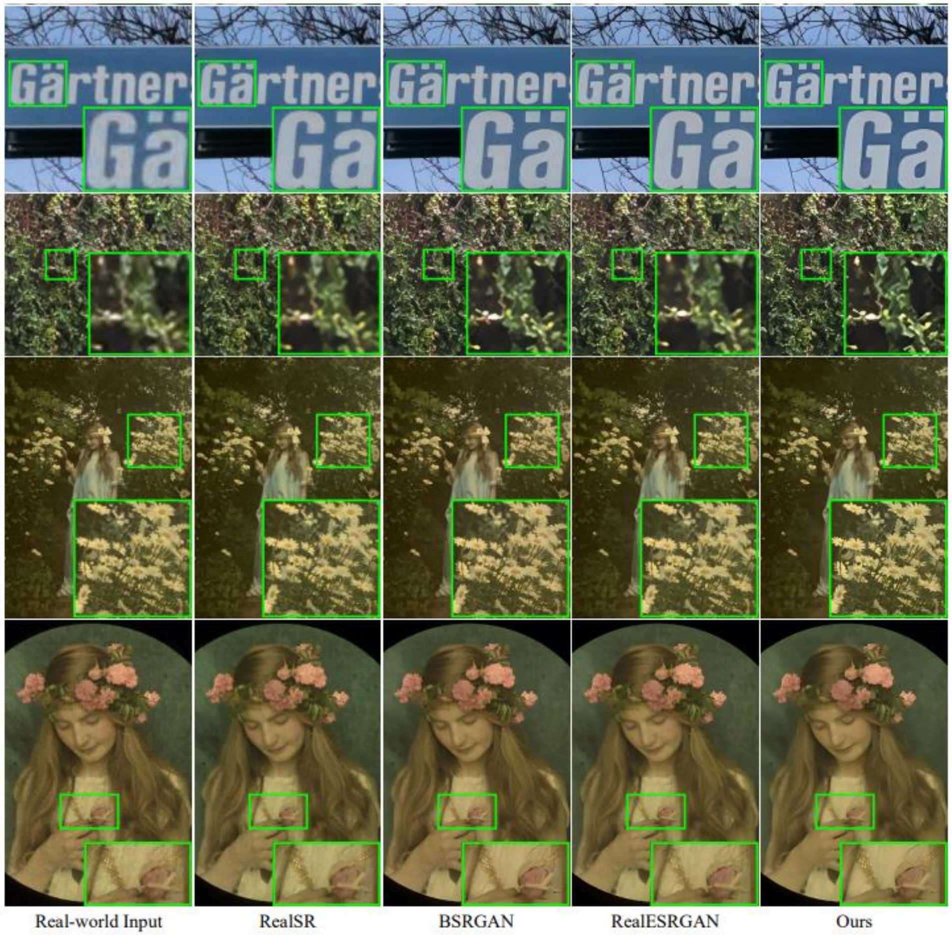


Figure 6: Visual comparison of these competing methods on real-world LQ images

서강대학교 SOGANG UNIVERSITY

VDS LAB

Result

- Visual comparison
  - Combine face image features and natural image features
    - Handle the general restoration with limited real degradation
    - Fine-tune the model for some specific scenarios that have face images

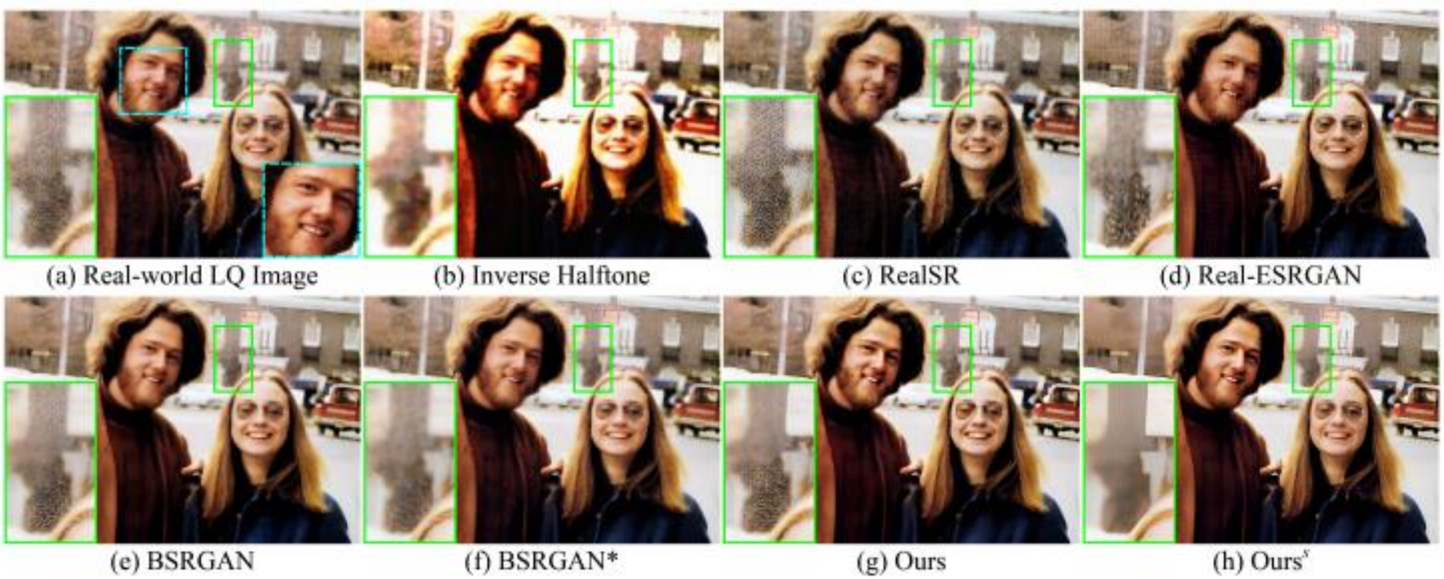


Figure 7: Restoration results on real-world LQ image

Conclusion

- PDM-SR
  - Probabilistic degradation model (PDM)
    - Decouple the degradation with the content of the image
    - Model the random factors in degradations
      - Generate HR-LR training sample with a larger variety of degradations
        - ✓Better cover the degradations of test LR images
      - Flexible formulation of degradations
- ReGeNet
  - Model the real degradations
    - The real-world LQ face images and pseudo HQ counterparts
  - Transfer real degradation process to HQ natural images
  - Learn degradation from non-facial regions and facial regions
    - Applied to general and specific restoration

Thank you