Unsupervised Image Super-Resolution using Cycle-in-Cycle Generative Adversarial Networks

Yuan, Yuan and Liu, Siyuan and Zhang, Jiawei and Zhang, Yongbing and Dong, Chao and Lin, Liang

CVPR2018

Motivation

1) **Why applying unsupervised training?**

   unavailability of the down-sampling process

   unavailability of the low-/high-resolution pairs

2) **What is the difference between SR and image-to-image translation?**

   SR accepts an LR image and outputs a HR image with much larger resolution and higher quality and is not just a different style.

cycleGAN model

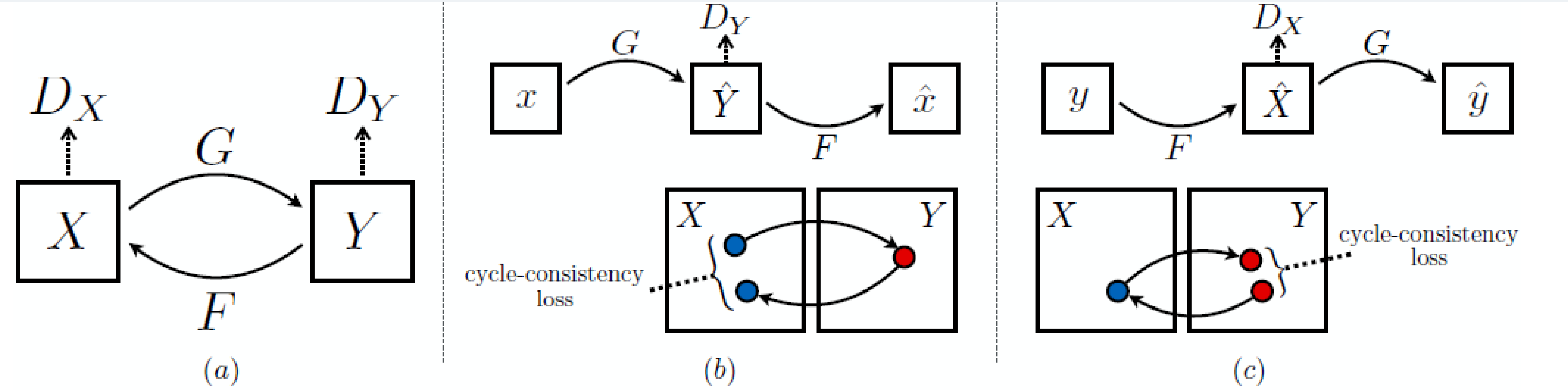
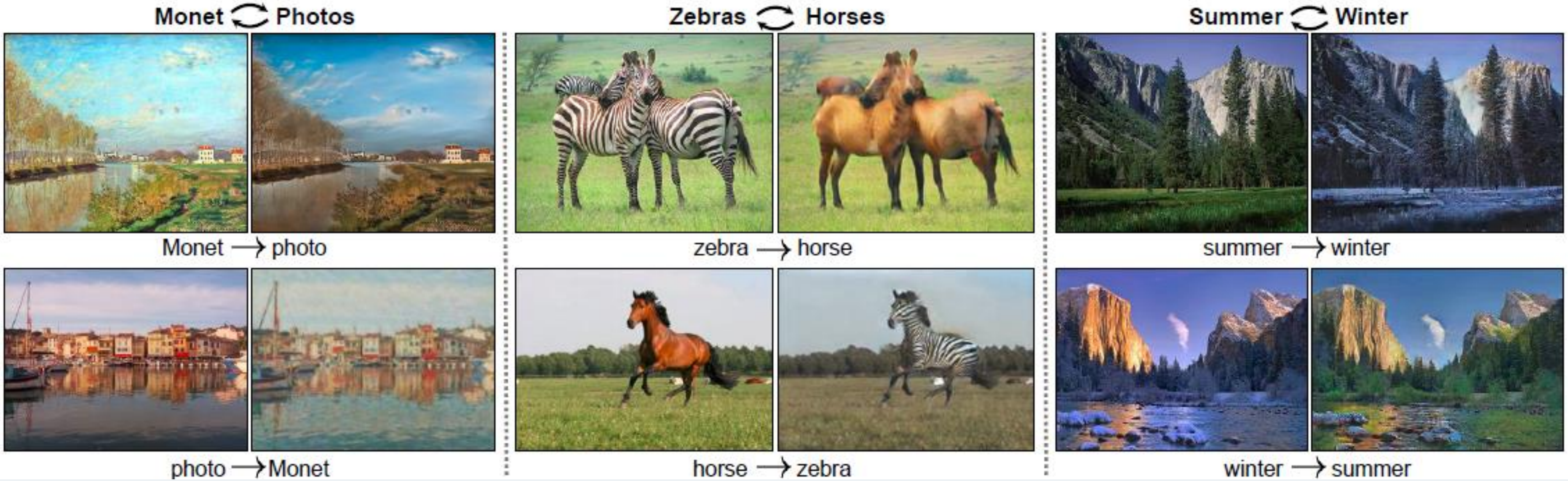


Figure 3: (a) Our model contains two mapping functions $G : X \rightarrow Y$ and $F : Y \rightarrow X$, and associated adversarial discriminators D_Y and D_X. D_Y encourages G to translate X into outputs indistinguishable from domain Y, and vice versa for D_X and F. To further regularize the mappings, we introduce two *cycle consistency losses* that capture the intuition that if we translate from one domain to the other and back again we should arrive at where we started: (b) forward cycle-consistency loss: $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$, and (c) backward cycle-consistency loss: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$

J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *2017 IEEE International Conference on Computer Vision (ICCV)*,2017.

cycleGAN


Monet ⇄ Photos
Zebras ⇄ Horses
Summer ⇄ Winter
Monet → photo
zebra → horse
summer → winter
photo → Monet
horse → zebra
winter → summer

CinCGAN model

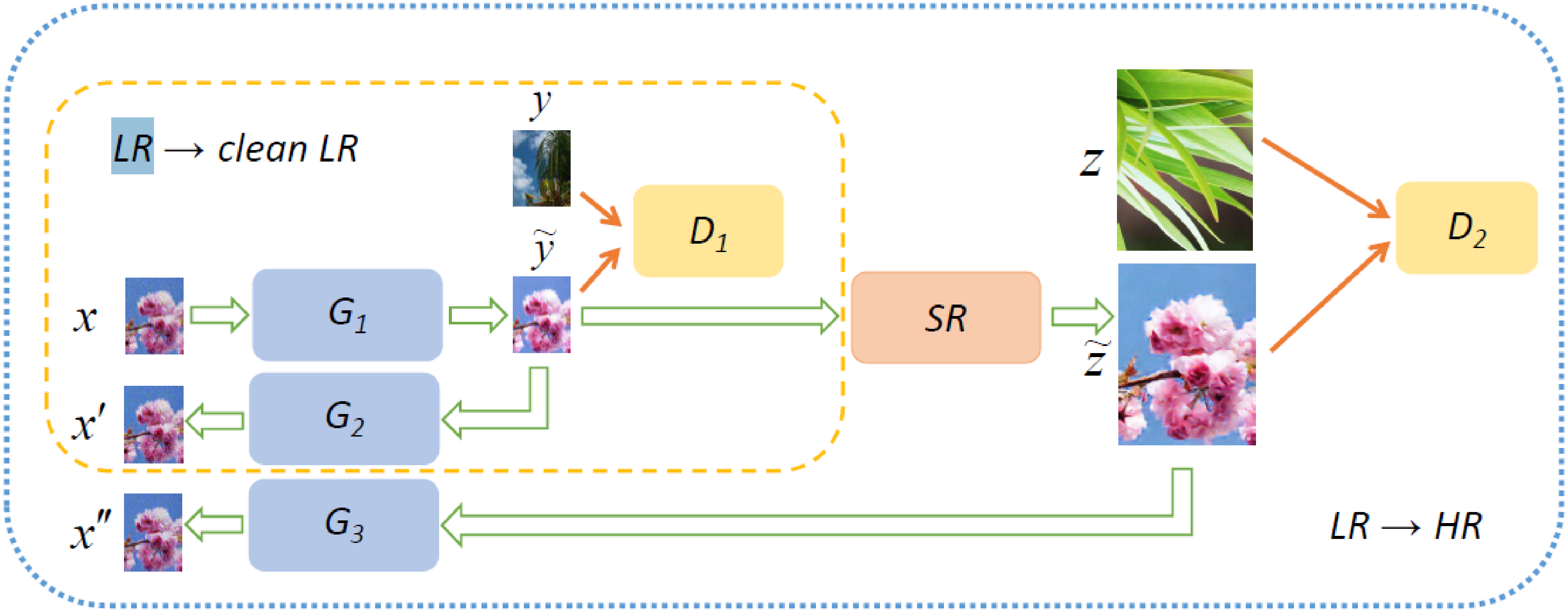


Figure 2. The framework of the proposed CinCGAN, where G_1, G_2 and G_3 are generators and *SR* is a super-resolution network. D_1 and D_2 are discriminators. The G_1, G_2 and D_1 compose the first *LR→clean LR* CycleGAN model, mapping the degrade LR images to clean LR images. The G_1, *SR*, G_3 and D_2 compose the second *LR→HR* CycleGAN model, mapping the LR images to HR images.

Solution pipeline : three steps

- First, deblur and denoise the input images at low resolution by learning a mapping from an LR image set X to a "clean" LR image set Y.
- Second, up-sampling images generated by step one with a pre-trained deep model.
- In the end, combining and fine-tune these two models simultaneously to get the final HR images.

LR Image Restoration

The purpose of the first CycleGAN framework is to map an LR image x to a clean LR image y.

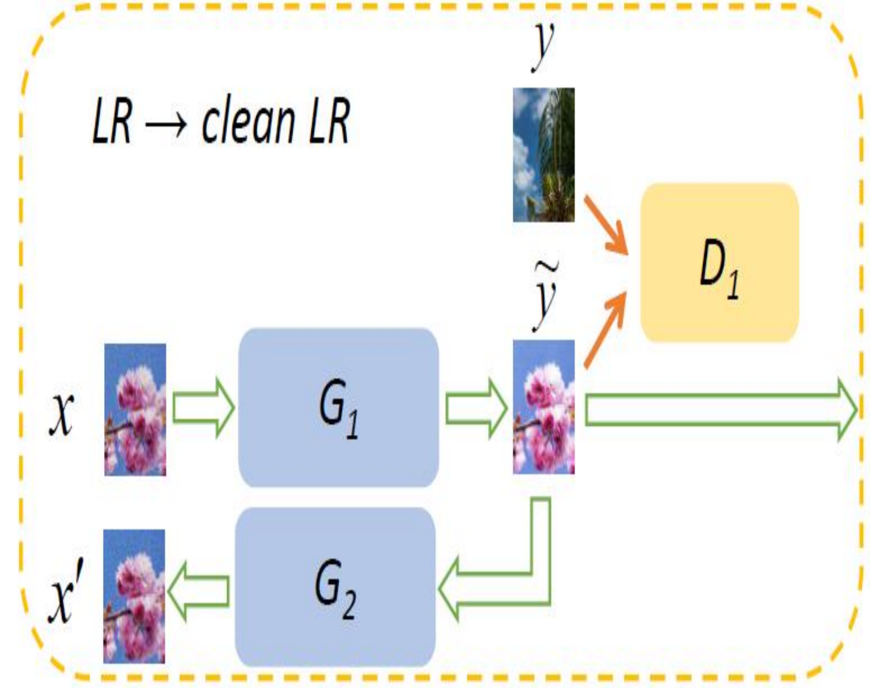
$$\mathcal{L}_{GAN}^{LR} = \frac{1}{N}\sum_{i}^{N} ||D_1(G_1(x_i)) - 1||_2, \quad (1)$$

$$\mathcal{L}_{cyc}^{LR} = \frac{1}{N}\sum_{i}^{N} ||G_2(G_1(x_i)) - x_i||_2. \quad (2)$$

$$\mathcal{L}_{idt}^{LR} = \frac{1}{N}\sum_{i}^{N} ||G_1(y_i) - y_i||_1. \quad (3)$$

$$\mathcal{L}_{TV}^{LR} = \frac{1}{N}\sum_{i}^{N} (||\nabla_h G_1(x_i)||_2 + ||\nabla_w G_1(x_i)||_2), \quad (4)$$

$$\mathcal{L}_{total}^{LR} = \mathcal{L}_{GAN}^{LR} + w_1\mathcal{L}_{cyc}^{LR} + w_2\mathcal{L}_{idt}^{LR} + w_3\mathcal{L}_{TV}^{LR} \quad (5)$$



CinCGAN model

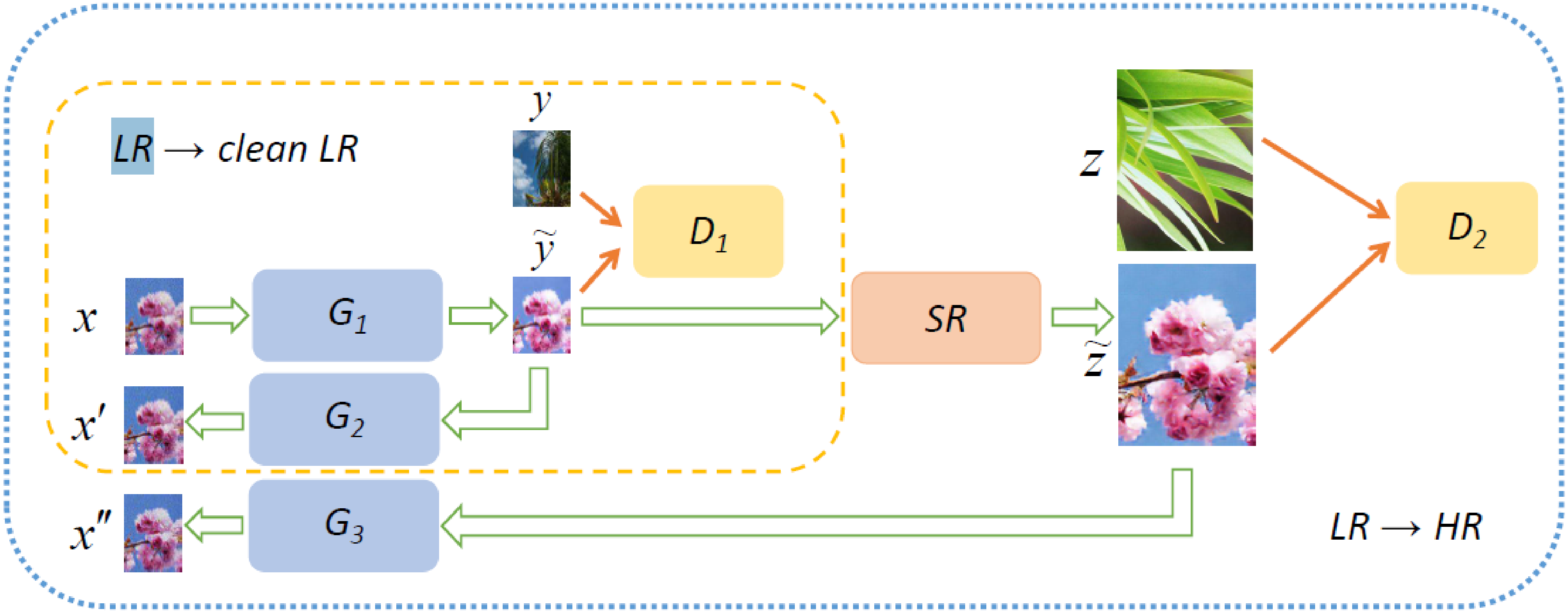


Figure 2. The framework of the proposed CinCGAN, where G_1, G_2 and G_3 are generators and *SR* is a super-resolution network. D_1 and D_2 are discriminators. The G_1, G_2 and D_1 compose the first *LR→clean LR* CycleGAN model, mapping the degrade LR images to clean LR images. The G_1, *SR*, G_3 and D_2 compose the second *LR→HR* CycleGAN model, mapping the LR images to HR images.

Jointly Restoration and Super-Resolution

For the desired size , author directly adopt EDSR as the SR network stacked after G1.

$$\mathcal{L}_{GAN}^{HR} = \frac{1}{N}\sum_{i}^{N} ||D_2(SR(G_1(x_i))) - 1||_2, \quad (6)$$

$$\mathcal{L}_{cyc}^{HR} = \frac{1}{N}\sum_{i}^{N} ||G_3(SR(G_1(x_i))) - x_i||_2, \quad (7)$$

$$\mathcal{L}_{TV}^{HR} = \frac{1}{N}\sum_{i}^{N} (||\nabla_h SR(G_1(x_i))||_2 + ||\nabla_w SR(G_1(x_i))||_2).$$

$$\mathcal{L}_{idt}^{HR} = \sum_{i} ||SR(z') - z||_2.$$

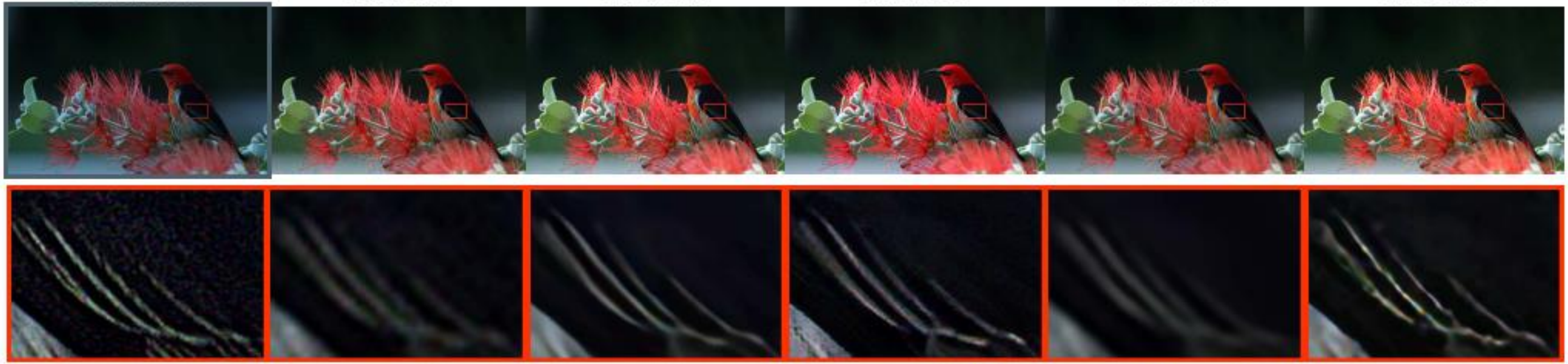
$$\mathcal{L}_{total}^{HR} = \mathcal{L}_{GAN}^{HR} + \lambda_1 \mathcal{L}_{cyc}^{HR} + \lambda_2 \mathcal{L}_{idt}^{HR} + \lambda_3 \mathcal{L}_{TV}^{HR} \quad (10)$$

B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee. Enhanced deep residual networks for single image super-resolution.In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, volume 1, page 3, 2017.

Experimental Results

(a) ground truth PSNR/SSIM | (b) bicubic 23.22/0.64 | (c) EDSR+ [17] 26.23/0.68 | (d) SRGAN+ [16] 24.06/0.58 | (e) BM3D+EDSR 23.06/0.65 | (f) **CinCGAN (ours)** 24.83/0.65

Experimental Results



(a) ground truth	(b) bicubic	(c) EDSR$^+$	(d) SRGAN$^+$ [16]	(e) BM3D+EDSR	(f) **CinCGAN (ours)**
PSNR/SSIM	26.81/0.83	30.28/0.88	29.05/0.85	26.84/0.86	28.26/0.84

Experimental Results

Table 1. Quantitative evaluation on NTIRE 2018 track 2 dataset of the proposed CinCGAN model, in terms of PSNR and SSIM.

method	bicubic	FSRCNN [4]	EDSR [17]	EDSR^{+}	SRGAN^{+} [16]	BM3D+EDSR	**CinCGAN (ours)**
PSNR	22.85	22.79	22.67	25.77	24.33	22.88	24.33
SSIM	0.65	0.61	0.62	0.71	0.67	0.68	0.69

THANKS