ComputerVisionLab
Seoul National University
https://cv.snu.ac.kr

Structure-Resonant Discriminator for Image Super-Resolution

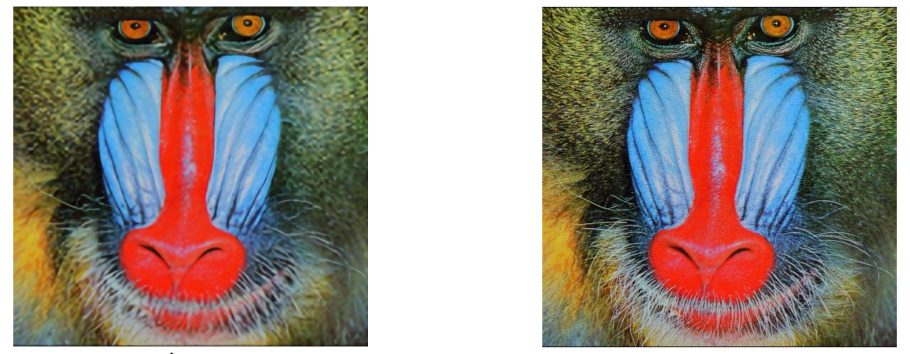
Jaerin Lee and Kyoung Mu Lee
Department of ECE, ASRI, Seoul National University, Seoul, Korea

ICME 2021

INTRODUCTION

Image Super-Resolution

- Conventionally, per-pixel ***Lp*** distances between HR images and SR images are used to train the deep image super-resolution network.
- Applying only the ***Lp*** distance leads to blurry results.

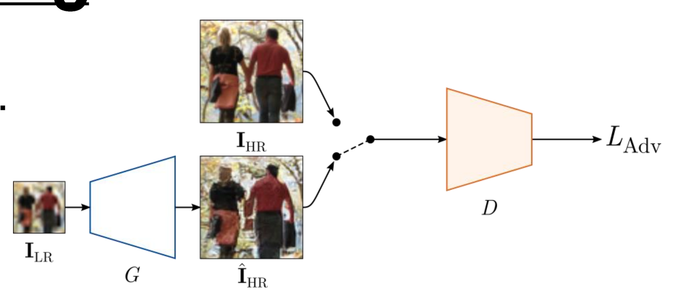


$\hat{\mathbf{I}}_{\mathrm{HR}}$ ×4 (EDSR B. Lee et al., In CVPRW 2017) $\mathbf{I}_{\mathrm{HR}}$

- More loss functions are designed by the computer vision community to alleviate the problem.
  1. **Perceptual loss.**
  2. **Adversarial loss.**

Rethinking Adversarial Training

- In this scheme, the **generator** and the **discriminator** is alternatively trained.
- The structure of the **discriminator** is seldom discussed compared to that of the **generator**.



- **Discriminator is an encoder:** it encodes **ground truth** or **generated** images to extract information of ***realness*** from them.
- **Discriminator is a data model:** it models natural and generated images.
- **Design of a *data model* should be compatible to the structure of the *data*** being modeled in order to achieve effective representation learning.
- **Our goal** is to design a **discriminator** that is **aware of the structure of natural images to provide more appropriate supervision in adversarial training framework – to be used in the image super-resolution problem.**

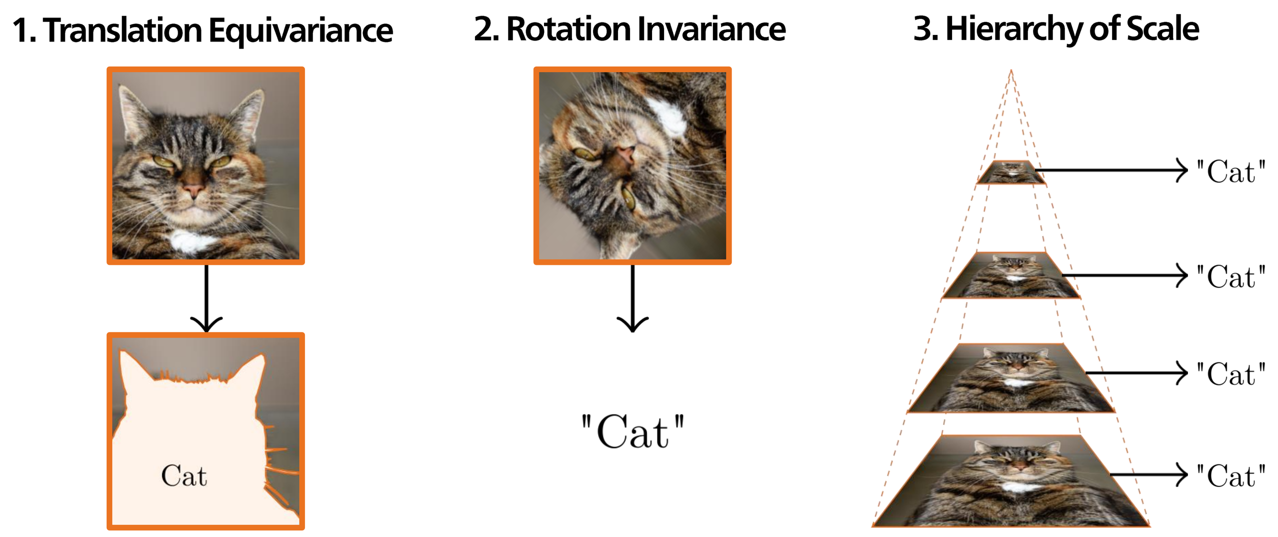
Structual Properties of Natural Images

- ***Three*** the most salient structural features presented in natural images.
- These will guide the design process of our new **discriminator** model.

1. **Translation Equivariance:** Information in a natural image is concentrated (in the form of objects, textures, etc.) in a local region; translation only affects the positions and not the contents.
2. **Rotation Invariance:** The same object can appear in arbitrary orientations.
3. **Hierarchy of Scale:** The same object can appear in various scales.

- These properties discusses about the **symmetries** occur in natural images.
- **Symmetries** are represented in the language of ***group theory*** in mathematics; We will use the mathematical notion of ***invariance*** and ***equivariance***.

DESIGN



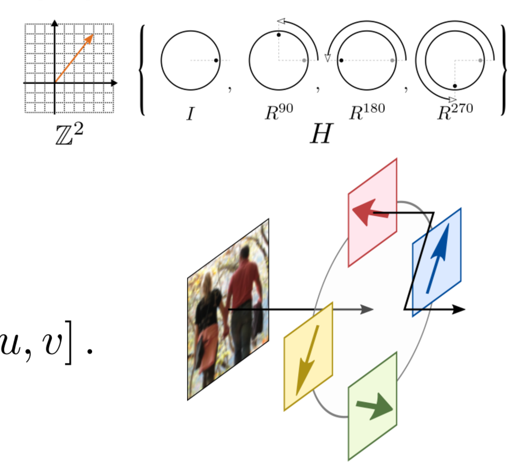
1. Translation Equivariance

- This property can be captured with a translation equivariant map:
  $$\mathbf{F}' := f(\mathbf{F})\,, \quad f(\mathtt{translate}(\mathbf{F}, t)) = \mathtt{translate}(\mathbf{F}', t/s)\,.$$
  $$\mathbf{F} \in \mathbb{R}^{b\times c\times h\times w}\,, \quad \mathbf{F}' \in \mathbb{R}^{b\times c\times h/s\times w/s}\,.$$
- A translation equivariant map
  - Ensures consistent ***local*** feature extraction.
  - Maintains ***positional*** information.

$$p4 := \mathbb{Z}^2 \rtimes H$$
Roto-translation group ***p4*** is a semi-direct product of
1. Discrete translation group $\mathbb{Z}^2$
2. Discrete 4-rotation group H

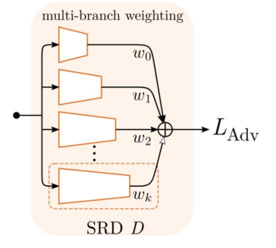
2. Rotation Invariance

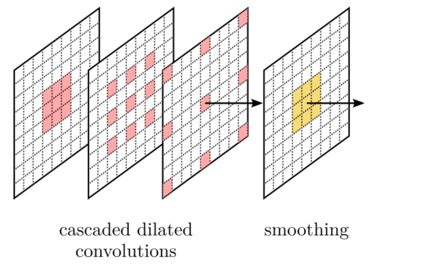
- Rotation invariance is embedded in the lowest layer of the **discriminator** where the gradients for the **generator** is generated from.
- It is built w.r.t. specific discrete roto-translation group, ***p4***.
- Rotation invariant map is a composition of:
  1. **A discrete rotation *equivariant* convolution**
     $$[\mathbf{F}']_\theta = [\mathbf{F} *_{\mathbb{Z}^2\to p4} \mathbf{k}]_\theta := \mathbf{F} * R^\theta \mathbf{k}\,.$$
     $$\mathbf{F} \in \mathbb{R}^{b\times c\times h\times w}\,, \quad \mathbf{F}' \in \mathbb{R}^{b\times c\times |H|\times h\times w}\,.$$
  2. **A rotation *group-wise* pooling layer**
     $$\mathtt{gmaxpool}(\mathbf{F}')[b,c,u,v] := \max_{g\in\mathcal{H}} \mathbf{F}'[b,c,g,u,v]\,.$$
     $$\mathtt{gmaxpool}(\mathbf{F}') \in \mathbb{R}^{b\times c\times h\times w}\,.$$



3. Hierarchy of Scale

- The **scale** of an object $\Longleftrightarrow$ The **receptive field size** of a filter
- ***Two*** design features are introduced to the **discriminator**:
  1. **Multi-branch architecture**
     - Multiple **receptive field sizes** for different **scales**.
     - Branch outputs are summarized by a **weighted sum**.
  2. **Cascaded dilated convolutions**
     - Efficient large receptive field by **exponentially increasing dilation**.
     - The **smoothing convolution** mitigates potential gridding artifact.





IMPLEMENTATION

Structure Resonant Discriminator (SRD)

- **Multi-branch structure**, each branch is a composition of:
  1. **Rotation invariant layer**: ***p4***-equivariant conv + group-wise maxpool
  2. **Cascaded dilated convolutions** with exponentially increasing dilation
  3. A single 3X3 **smoothing convolution**
  4. A 1X1 convolution **classifier**

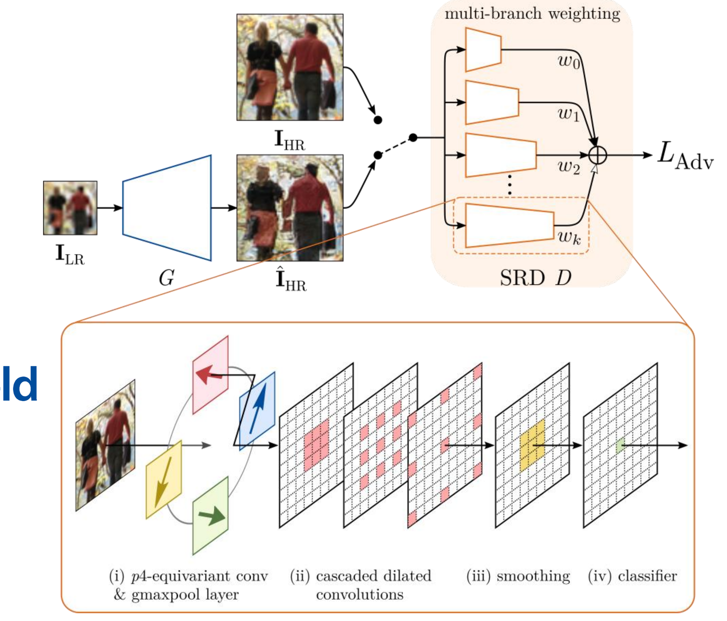
Implementation Details

1. **Multi-branch architecture**

branch ID i	depth l	receptive field size	branch weight w_i (rel.)
1	1*	25×25	1
2	2	65×65	1.25^2
3	3	129×129	1.5^2
4	4	257×257	1.75^2
5	5	513×513	2^2

2. **Cascaded dilated convolutions**
   - Enables exponential **receptive field size** w.r.t. the number of layers.

layer	gconv	conv_1	conv_2	dconv_1	⋯	dconv_l	conv_3	cls
kernel size	3	3	3	3	⋯	3	3	1
stride	1	2	2	1	⋯	1	1	1
dilation	1	1	1	2	⋯	2^l	1	1
receptive field size	3	+2	+4	+16	⋯	$+2^{l+3}$	+8	+0



Training Details

- **Loss function: LSGAN** loss.
  $$L^D_{\mathrm{adv}} = 1/2\ \mathbb{E}\|D(\mathbf{I}_{\mathrm{HR}}) - \mathbf{1}\|_2^2 + 1/2\ \mathbb{E}\|D(\mathbf{I}_{\mathrm{SR}}) - \mathbf{0}\|_2^2\,,$$
  $$L^G_{\mathrm{adv}} = 1/2\ \mathbb{E}\|D(\mathbf{I}_{\mathrm{SR}}) - \mathbf{1}\|_2^2\,,$$
- **SR Network:** **RRDB (ESRGAN** baseline).
- **Dataset:** Mixture of **DIV2K** & **Flickr2K**.
- **Optimizer:** **Adam** with default hyperparameter.

Other loss functions
1. L1 loss.
   $$L_1(\mathbf{I}_{\mathrm{HR}}, \hat{\mathbf{I}}_{\mathrm{HR}}) = \mathbb{E}[\|\mathbf{I}_{\mathrm{HR}} - \hat{\mathbf{I}}_{\mathrm{HR}}\|_1]\,,$$
2. L1 perceptual loss.
   $$L_{\mathrm{per}}(\mathbf{I}_{\mathrm{HR}}, \hat{\mathbf{I}}_{\mathrm{HR}}) = \mathbb{E}[\|\mathrm{VGG}(\mathbf{I}_{\mathrm{HR}}) - \mathrm{VGG}(\hat{\mathbf{I}}_{\mathrm{HR}})\|_1]\,,$$

Overall loss function:
$$L_{\mathrm{tot}} = \lambda_1 L_1 + \lambda_{\mathrm{per}} L_{\mathrm{per}} + \lambda_{\mathrm{adv}} L_{\mathrm{adv}}\,.$$

Quantitative Results

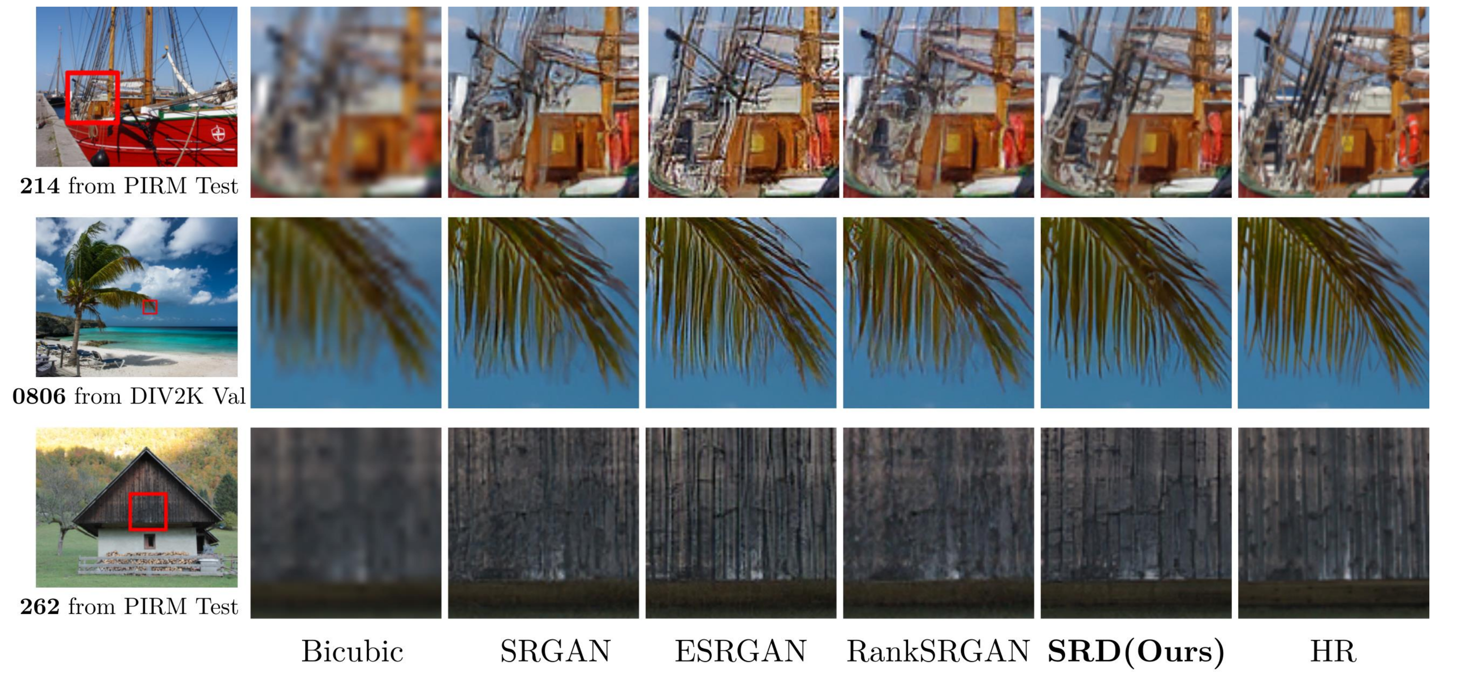
Method	Set5			Set14			BSD100			Urban100		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Bicubic	28.42	0.8104	0.340	26.00	0.7027	0.438	25.96	0.6675	0.524	23.14	0.6577	0.473
EDSR	32.46	0.8968	0.174	28.80	0.7876	0.284	27.71	0.7420	0.372	26.64	0.8033	0.232
RRDB	32.73	0.9011	0.173	28.99	0.7917	0.277	27.85	0.7455	0.366	27.03	0.8153	0.200
EDSR + GAN	29.39	0.8420	0.086	26.47	0.7183	0.146	25.73	0.6662	0.191	23.88	0.7169	0.155
SRGAN (Reproduced)	29.90	0.8486	0.081	26.56	0.7089	0.145	25.49	0.6524	0.177	24.38	0.7305	0.143
ESRGAN	30.44	0.8498	0.075	26.28	0.6981	0.134	25.28	0.6495	0.162	24.34	0.7327	0.123
SROBB	28.93	0.817	0.087	25.43	0.678	0.162	-	-	-	-	-	-
RankSRGAN-NIQE	29.77	0.8363	0.073	26.48	0.7023	0.138	25.49	0.6484	0.177	24.53	0.7279	0.142
RankSRGAN-Ma	28.85	0.8204	0.078	25.79	0.6852	0.145	25.03	0.6390	0.183	24.12	0.7182	0.143
RankSRGAN-PI	29.65	0.8342	0.073	26.46	0.7021	0.137	25.44	0.6484	0.175	24.47	0.7289	0.138
EDSR + RaGAN + **SRD (Ours)**	29.91	0.8473	0.080	26.80	0.7187	0.134	25.78	0.6626	0.184	24.38	0.7326	0.149
RRDB + RaGAN + **SRD (Ours)**	30.46	**0.8523**	0.074	26.73	0.7129	**0.126**	25.80	0.6663	0.164	24.72	0.7441	0.124
RRDB + LSGAN + **SRD (Ours)**	**30.55**	0.8506	**0.063**	**26.93**	**0.7227**	**0.126**	**25.87**	**0.6693**	**0.157**	**25.07**	**0.7542**	**0.117**

Method	DIV2K-Val			PIRM-Val			PIRM-Test			OST300		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Bicubic	26.66	0.8521	0.409	26.50	0.6980	0.465	26.45	0.6892	0.481	25.74	0.6647	0.512
EDSR	29.25	0.9017	0.271	28.29	0.7716	0.309	28.23	0.7632	0.325	27.00	0.7289	0.373
RRDB	29.44	0.9043	0.256	28.71	0.7849	0.292	28.61	0.7756	0.310	27.30	0.7411	0.362
EDSR + GAN	26.62	0.8513	0.133	25.88	0.6848	0.141	25.78	0.6712	0.149	25.25	0.6589	0.190
SRGAN (Reproduced)	26.60	0.8481	0.126	25.61	0.6757	0.144	25.47	0.6599	0.153	24.90	0.6461	0.180
ESRGAN	26.61	0.8479	0.115	25.18	0.6599	0.144	25.04	0.6452	0.152	24.63	0.6422	0.169
RankSRGAN-NIQE	26.53	0.8421	0.128	25.76	0.6739	0.139	25.62	0.6584	0.145	24.97	0.6415	0.184
RankSRGAN-Ma	25.60	0.8261	0.145	25.22	0.6610	0.143	25.11	0.6468	0.150	24.55	0.6261	0.192
RankSRGAN-PI	26.48	0.8431	0.122	25.64	0.6724	0.136	25.48	0.6564	0.143	24.91	0.6410	0.180
EDSR + RaGAN + **SRD (Ours)**	26.80	0.8518	0.127	25.92	0.6830	0.141	25.82	0.6688	0.148	25.19	0.6538	0.185
RRDB + RaGAN + **SRD (Ours)**	26.82	0.8471	0.118	25.88	0.6837	0.130	25.79	0.6710	0.137	25.29	0.6611	0.185
RRDB + LSGAN + **SRD (Ours)**	**27.05**	**0.8529**	**0.107**	**26.09**	**0.6935**	**0.123**	**26.00**	**0.6815**	**0.129**	**25.31**	**0.6624**	**0.166**

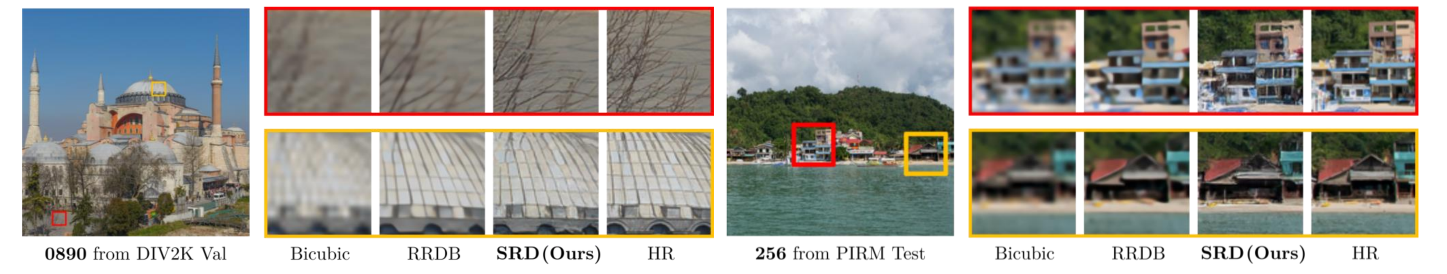
RESULTS

Qualitative Results

- Exhibits the **least distortion** and **visual artifacts** that degrades quality of the SR, compared to the **perceptual quality-oriented methods**.

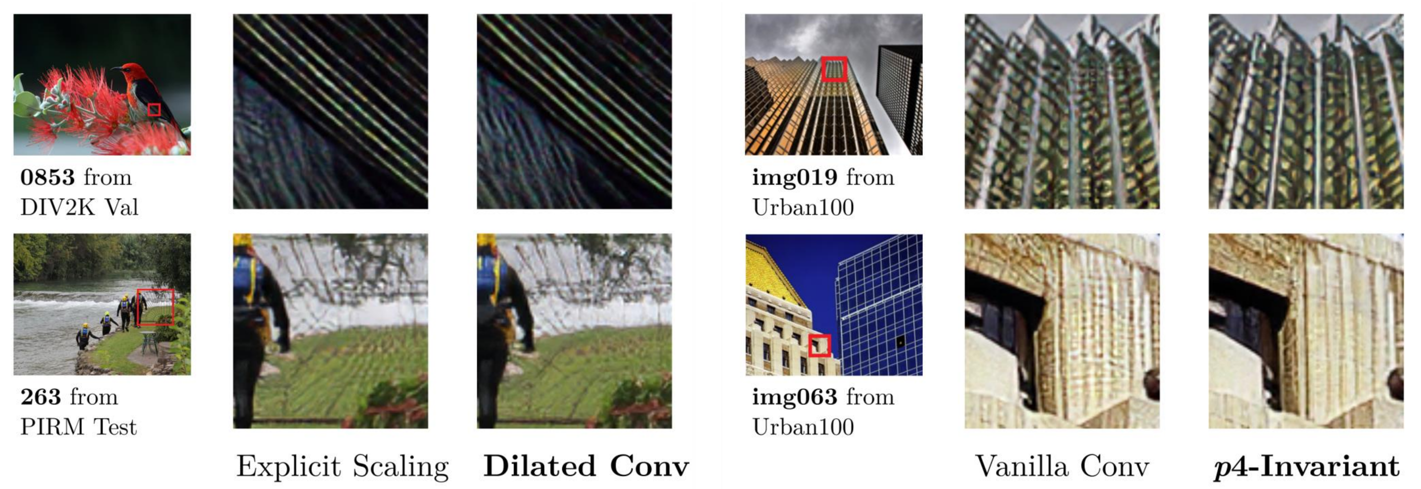


- **Visually pleasing** and **sharp** results compared to **PSNR-oriented methods.**



Ablation Studies

- Demonstrate each additional components in the **SRD** network.
1. **Cascaded dilated convolution** vs. **explicit input scaling** for multi-scale architecture.
2. ***p4*-invariant layer** vs. **standard convolution** for first layer convolution of the discriminator.



Conclusion

- **Discriminator** is also a **data model** and the design of it should be aware of the **structure of natural images**, which constitutes its input space.
- Emphasized ***three*** the most salient structural features of natural images.
- By ***only*** modifying the **discriminator**, significant performance boost in both **accuracy** and **perceptual quality** in image super-resolution is achieved.