ORTHOGONALLY REGULARIZED DEEP NETWORKS FOR IMAGE SUPER-RESOLUTION

Information Processing & Algorithms Laboratory
Everything can be optimized

TIANTONG GUO, HOJJAT SEYED MOUSAVI, VISHAL MONGA | THE PENNSYLVANIA STATE UNIVERSITY

INTRODUCTION & MOTIVATION

Recent advances have seen a surge of deep learning approaches for image super-resolution:

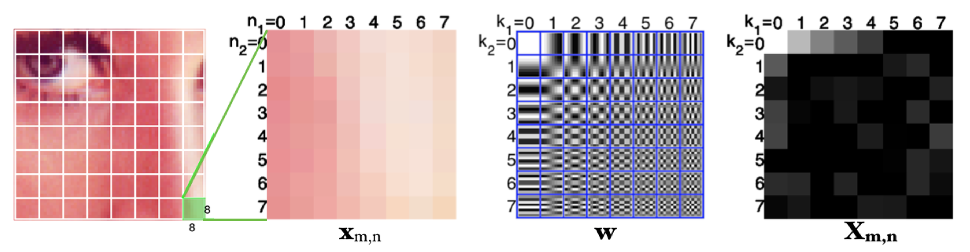
- CNN[1] with residual net structure[2] is trained to learn the relationship between low & high-resolution images.
- Deep learning SR methods[3] work on spatial domain data and aim to reconstruct pixel values.
- ORDSR explores the advantages of transform domain and learning suitable transformation basis functions.

2D DCT & SR

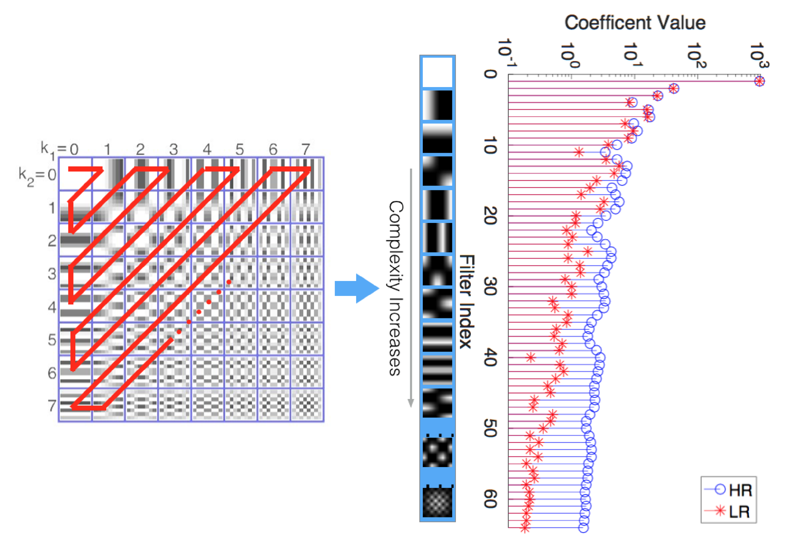
An image $\mathbf{x}(n_1, n_2)$ of size $H \times W$ can be decomposed into blocks of size $N \times N$.

$$\mathbf{X}_{m,n}(k_1, k_2) = \sum_{n_2=0}^{N-1} \sum_{n_1=0}^{N-1} \mathbf{x}_{m,n}(n_1, n_2) \times \mathbf{w}^{\text{dct}}_{k_1,k_2}(n_1, n_2)$$

$$\mathbf{w}^{\text{dct}}_{k_1,k_2}(n_1, n_2) = C_{k_1,k_2} cos\left[\tfrac{\pi}{N}\left(n_1 + \tfrac{1}{2}\right)k_1\right] cos\left[\tfrac{\pi}{N}\left(n_2 + \tfrac{1}{2}\right)k_2\right]$$

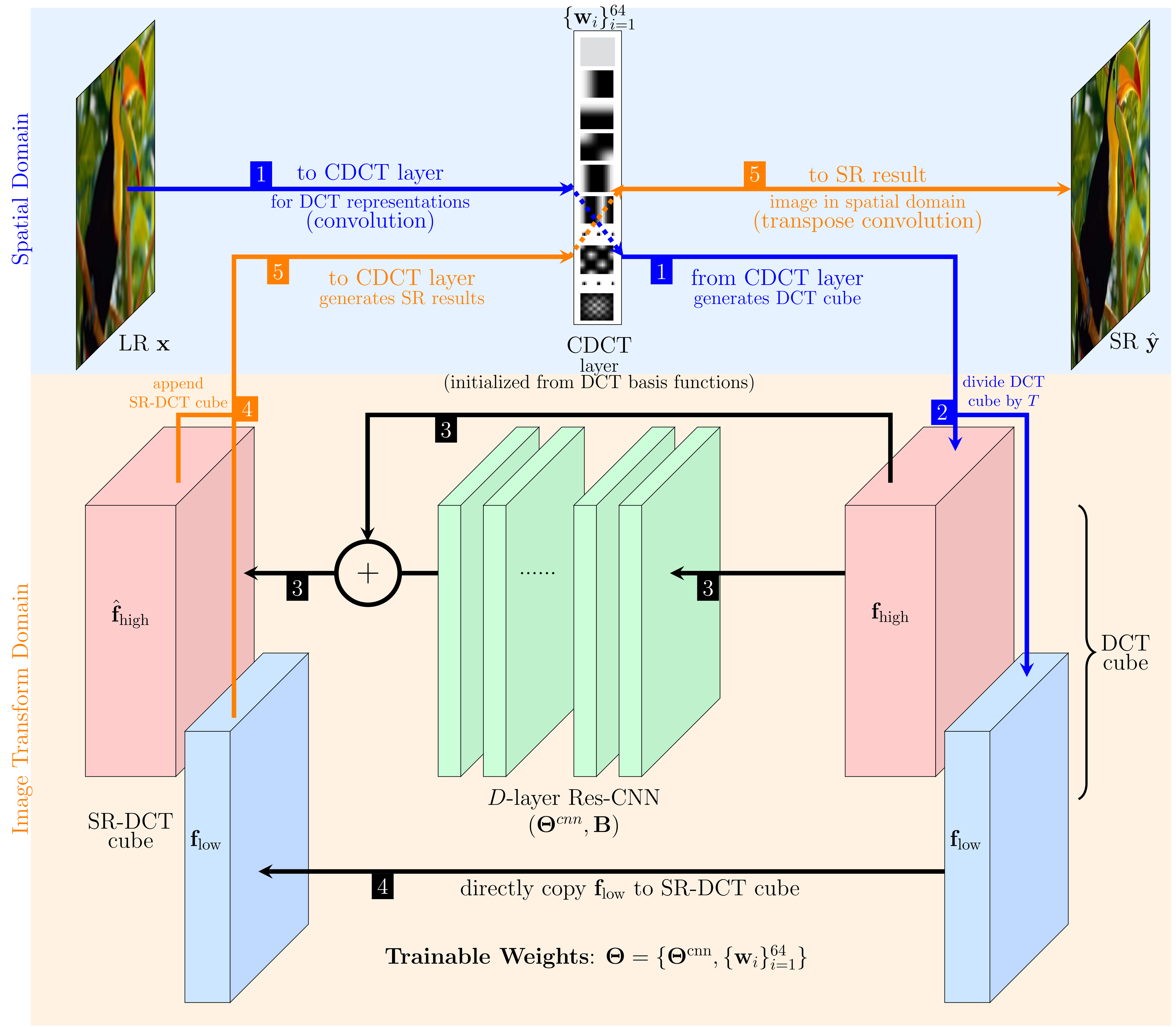


Zig-zag reordering the DCT basis functions as filters with increasing complexity order:



LR and HR image share low-frequency DCT coefficient. SR becomes the problem of **recovering high-frequency DCT coefficients** of the HR image from the corresponding LR ones.

ORDSR: NETWORK STRUCTURE



REGULARIZATION TERMS & COST FUNCTION

$$\mathbf{L} = \underbrace{\frac{1}{2}\|F(\mathbf{x}) - \mathbf{y}\|_2^2}_{\text{MSE loss}} + \underbrace{\sigma\frac{1}{2}\sum_l \|\mathbf{W}_l\|_2^2}_{\text{weight decay}} + \underbrace{\gamma\frac{1}{2}\sum_{(i,j),i\neq j} \|vec(\mathbf{w}_i)^T vec(\mathbf{w}_j) - \epsilon\|_2^2}_{\text{orthogonality constraint}} + \underbrace{\lambda\frac{1}{2}\sum_t \|var(\mathbf{w}_t) - var(\mathbf{w}_t^{\text{dct}})\|_2^2}_{\text{complexity order constraint}}$$

Pairwise Orthogonality Constraint - keeping the orthogonality properties of transform layer by forcing pairwise orthogonality

$$\forall i \neq j, \|vec(\mathbf{w}_i)^T vec(\mathbf{w}_j) - \epsilon\|_2^2 = 0$$

Complexity Order Constraint - keeping the frequency order of the DCT cube

$$\|var(\mathbf{w}_t) - var(\mathbf{w}_t^{\text{dct}})\|_2^2 = 0$$

$\mathbf{w}_t$ are the CDCT filters and $\mathbf{w}_t^{\text{dct}}$ is the DCT counterparts.
Modified Back-Propagation - training the network with desired gradient terms:

$$\frac{\partial \mathbf{L}}{\partial \mathbf{W}_l^a} = - < (\hat{\mathbf{y}} - \mathbf{y}), \frac{\partial \mathbf{y}}{\partial \mathbf{W}_l^a} >_F + \sigma < \mathbf{W}_l, \frac{\partial \mathbf{W}_l}{\partial \mathbf{W}_l^a} >_F,$$

$$\frac{\partial \mathbf{L}}{\partial \mathbf{w}_i^a} = - < (\hat{\mathbf{y}} - \mathbf{y}), \frac{\partial \mathbf{y}}{\partial \mathbf{w}_i^a} >_F + \gamma \sum_{(j)} \left(vec(\mathbf{w}_i)^T vec(\mathbf{w}_j) - \epsilon\right) \mathbf{w}_j^a + \lambda \frac{\partial var(\mathbf{w}_i)}{\partial \mathbf{w}_i^a}\left(var(\mathbf{w}_i) - var(\mathbf{w}_i^{\text{dct}})\right)$$

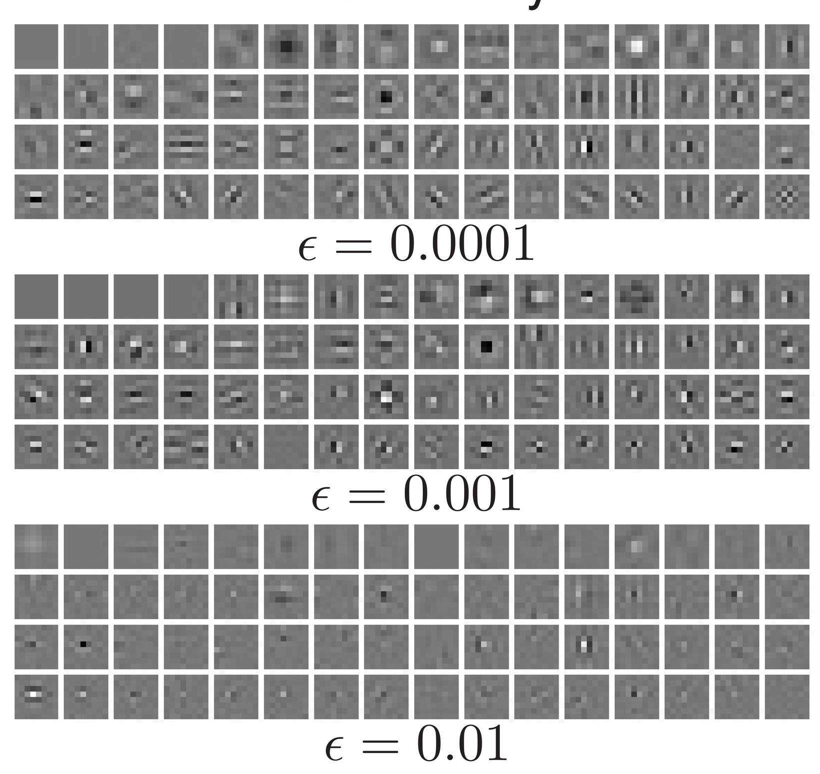
INFERENCE PROCEDURE

For an input LR image $\mathbf{x}$, ORDSR generates its SR version $\hat{\mathbf{y}}$ as follows:

1. The input $\mathbf{x}$ is convolved with CDCT layer producing a DCT cube $\{\mathbf{f}_i\}_{i=1}^{64}$
2. The DCT cube of $\mathbf{x}$ is divided into $\mathbf{f}_{\text{low}}$ and $\mathbf{f}_{\text{high}}$ corresponding to low and high-frequency spectra using a threshold T
3. A D-layer CNN takes $\mathbf{f}_{\text{high}}$ as input and recovers the missing high-frequency information using a residual network structure, generating $\hat{\mathbf{f}}_{\text{high}}$.
4. The $\hat{\mathbf{f}}_{\text{high}}$ is appended to $\mathbf{f}_{\text{low}}$ forming the SR-DCT cube $\{\hat{\mathbf{f}}_i\}_{i=1}^{64}$.
5. The SR-DCT cube $\{\hat{\mathbf{f}}_i\}_{i=1}^{64}$ is transpose convolved with the filters in the CDCT/transform layer (to perform the IDCT/inverse transform) generating $\hat{\mathbf{y}}$.

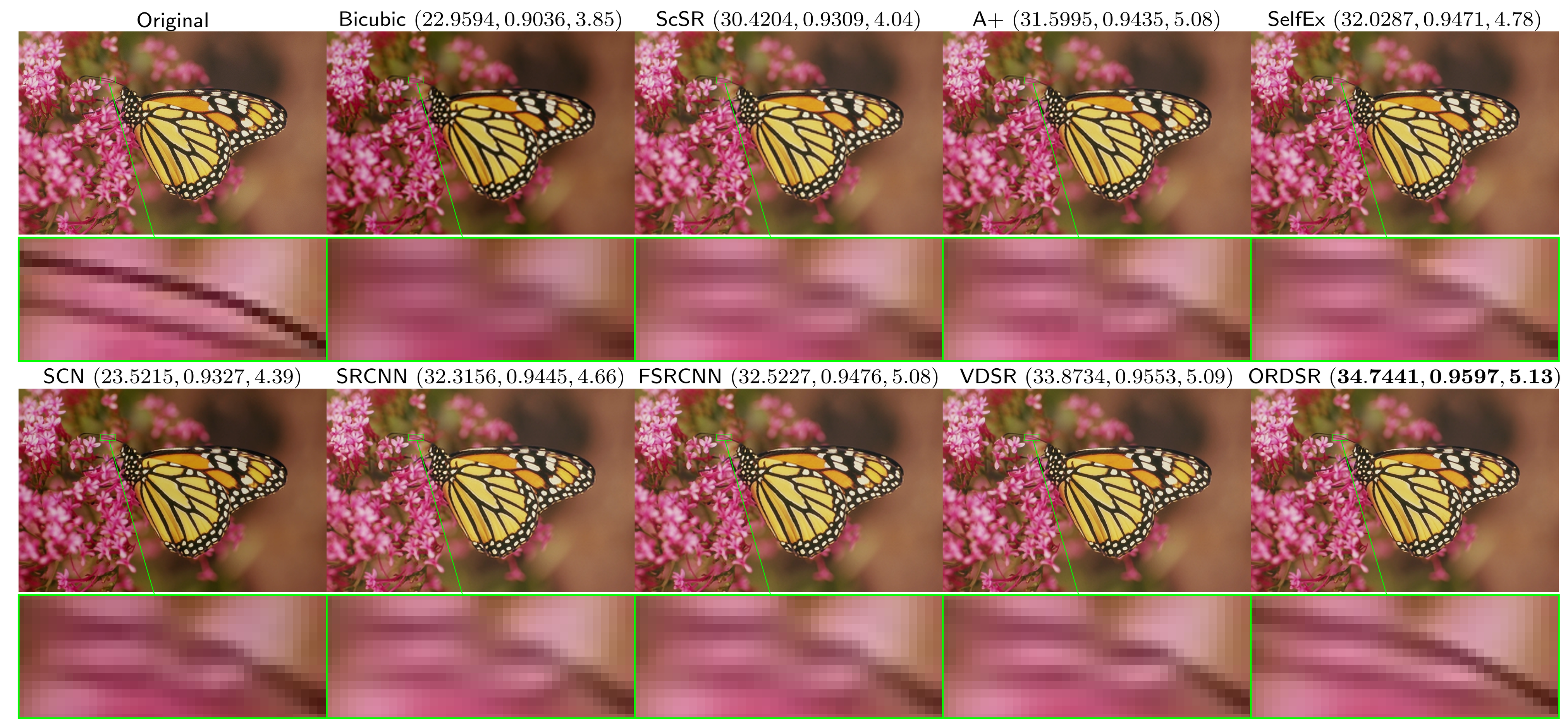
EXPERIMENTAL RESULTS

Filters of CDCT layers:

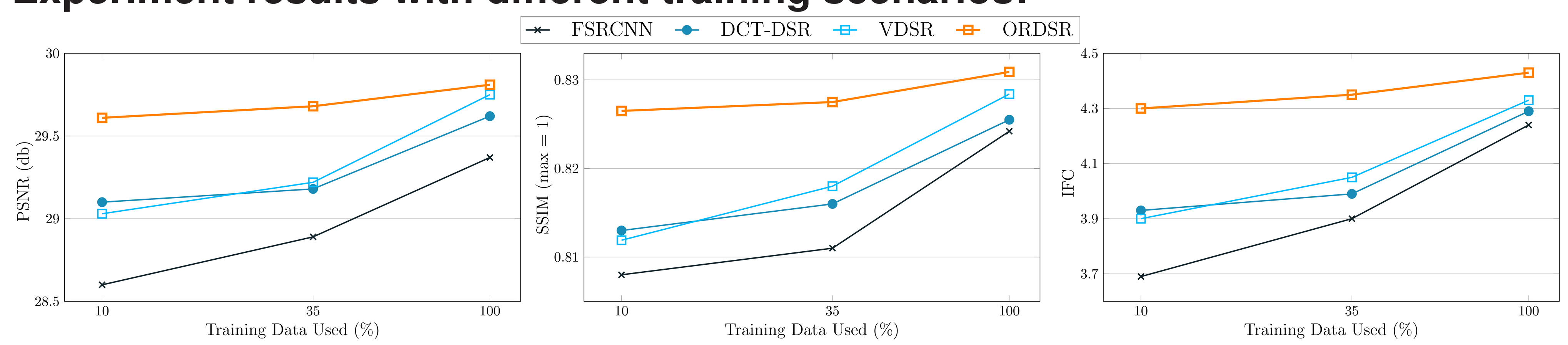


		Bicubic [Baseline]		ScSR [TIP 10]		FSRCNN [ECCV 16]	
Set5	x2	33.64	0.9292	35.78	0.9485	36.94	0.9558
	x3	30.39	0.8678	31.34	0.8869	33.06	0.9140
	x4	28.42	0.8101	29.07	0.8263	30.55	0.8657
Set14	x2	30.22	0.8683	31.64	0.8940	32.54	0.9088
	x3	27.53	0.7737	28.19	0.7977	29.37	0.8242
	x4	25.99	0.7023	26.40	0.7218	27.50	0.7535
		SRCNN [PAMI 16]		VDSR [CVPR 16]		ORDSR [proposed]	
Set5	x2	36.66	0.9542	**37.52**	**0.9586**	37.48	0.9574
	x3	32.75	0.9090	33.66	0.9212	**33.74**	**0.9221**
	x4	30.48	0.8628	31.35	0.8820	**31.38**	**0.8847**
Set14	x2	32.42	0.9063	33.02	0.9102	**33.04**	**0.9109**
	x3	29.28	0.8209	29.77	0.8308	**29.81**	**0.8300**
	x4	27.40	0.7503	28.01	0.7664	**28.06**	**0.7664**

Best results shown in red, second best shown in blue.



Experiment results with different training scenarios:



REFERENCES & CONTACT

[1] Chao Dong et al. Learning a deep convolutional network for image super-resolution. In *ECCV*. Springer, 2014.
[2] Jiwon Kim et al. Accurate image super-resolution using very deep convolutional networks. In *CVPR*, 2016.
[3] T. Guo et al. Deep learning based image sr with coupled backpropagation. In *GlobalSIP*, 2016.
[4] Radu Timofte et al. Ntire 2017 on sisr. In *CVPRW, IEEE*, 2017.

Website: http://signal.ee.psu.edu/ORDSR.html
Contact: tiantong@psu.edu hojjat@psu.edu
Supported by NSF Career Award to V. Monga.