Hierarchically Aggregated Residual Transformation for Single Image Super Resolution

Zejiang Hou and Sun-Yuan Kung

Department of electrical engineering, Princeton University

{zejiangh, kung}@princeton.edu

Single Image Super Resolution (SISR)

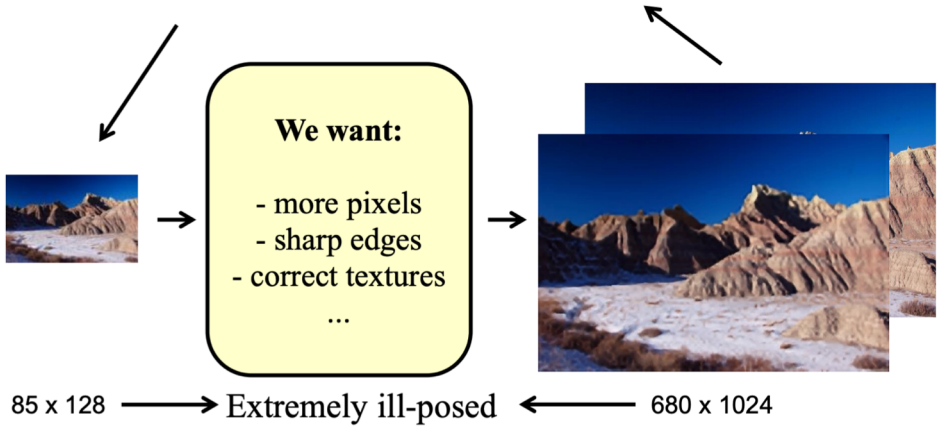
- Problem definition: reconstructing a high-resolution (HR) image from a degraded low-resolution (LR) input.

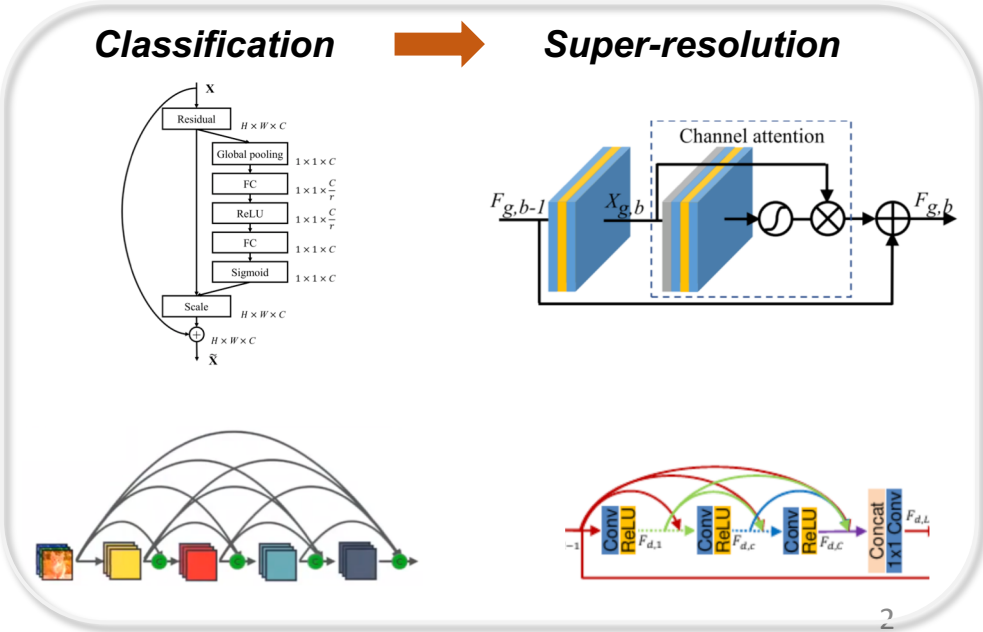
- Lots of recent search on deep learning SISR
  - Deeper and more complex architectures
  - Various attention mechanisms
  - Feedback mechanisms: error back-projection, high-level feature feedback…

- Challenges for current DL approaches
  - Naïve employment of classification network
  - Incapability to reconstruct multi-scale objects and leverage multi-scale features within each layer

- Existing multi-scale network
  - Down-sampling/up-sampling operations to resize feature-maps: information loss leads to inferior performance
  - Inception-like multi-scale module: larger kernel size renders the network inefficient

$$y = D(x \otimes k) + n$$

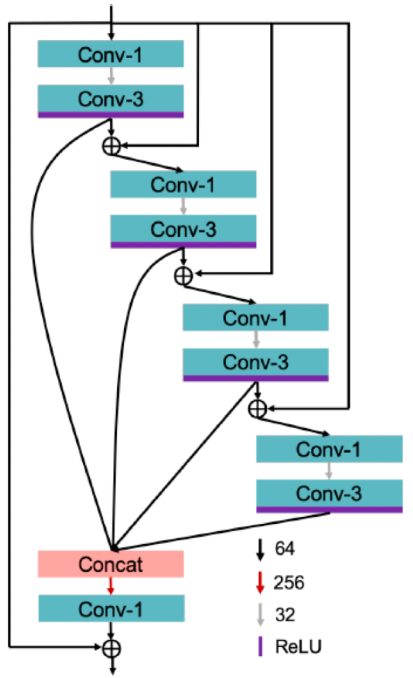




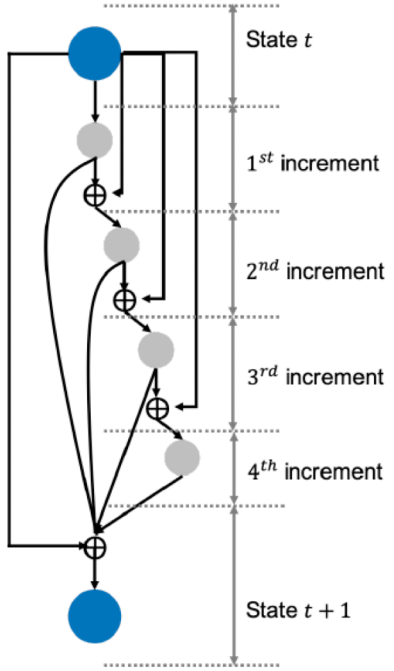
PRINCETON UNIVERSITY

Contribution: HARTnet

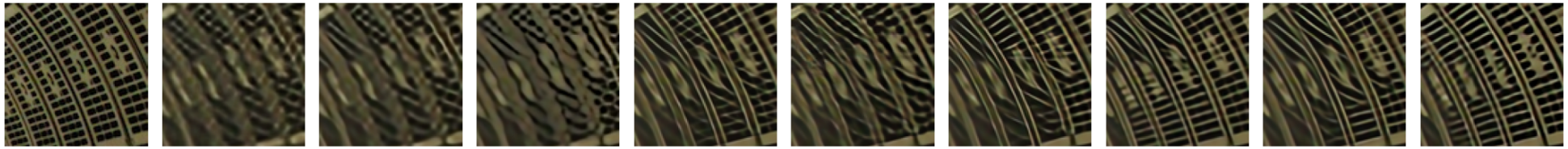
- Hierarchically aggregated residual transformations (HART) building block for multi-scale feature representation.
- Model interpretation from perspective of numerical ordinary differential equation
- Generalizable architecture to handle other image restoration tasks: image denoising, low-light image enhancement.
- State-of-the-art SISR performance



(a) HART block

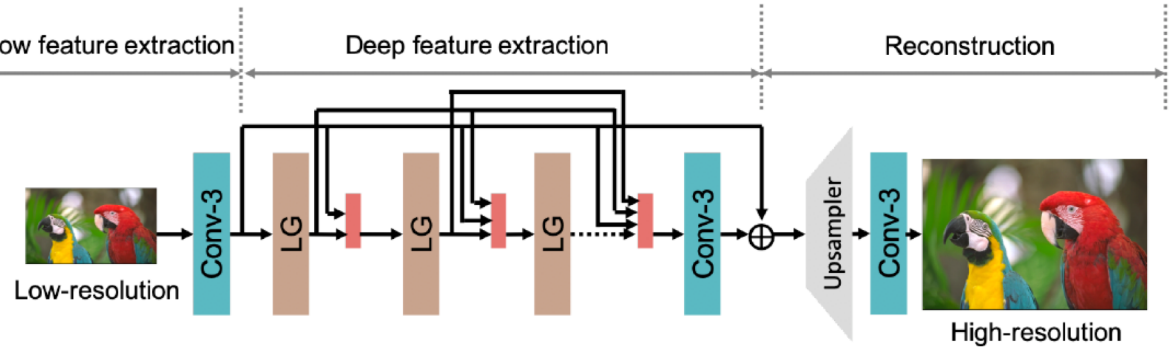


(b) RK4 method



HR | SRCNN[1] | VDSR[6] | LapSRN[7] | EDSR[2] | MSRN[8] | RDN[9] | RCAN[3] | OISR[10] | **HARTnet**

Model Overview



(a) Overall HARTnet architecture

Relating CNN-based SISR as an optimal control problem

Pixel-wise loss function

$$\min_{\{\boldsymbol{\theta}(t)\}_{t=0}^{T}} \mathcal{L}(\mathcal{F}_{REC}(\mathbf{x}_{\uparrow}), \mathbf{I}_{HR}) + \int_0^T \mathcal{R}(\boldsymbol{\theta}(t), t)dt$$

$$s.t. \quad \mathbf{x}_{\uparrow} = \mathcal{F}_{\uparrow}(\mathbf{x}(T) + \mathbf{x}(0))$$ → Upsampler conv.

$$\dot{\mathbf{x}} = f(\mathbf{x}(t), \boldsymbol{\theta}(t)), \quad t \in [0, T]$$ → Discretizing the continuous dynamics as multiple feature extraction modules

$$\mathbf{x}(0) = \mathcal{F}_{LFE}(\mathbf{I}_{LR})$$ → Shallow feature extraction conv.

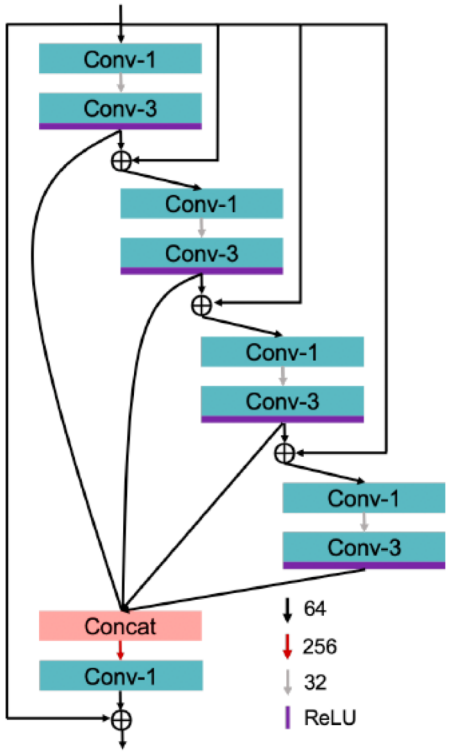
Hierarchically Aggregated Residual Transformations

- The key to multi-scale feature representation: increase the range of receptive fields in each layer
- HART block: replace single 3x3 conv by multiple bottleneck convs connected in hierarchical residual-like fashion
- Split-transform-concatenate strategy
  - Achieve multiple equivalent receptive fields at a granular level
  - Enrich the feature scales in the output of each block

$$\mathbf{O}_1 = \sigma(\mathbf{W}^1_{3\times3} * \mathbf{W}^1_{1\times1} * (\mathbf{X}_{in})) \quad (10)$$

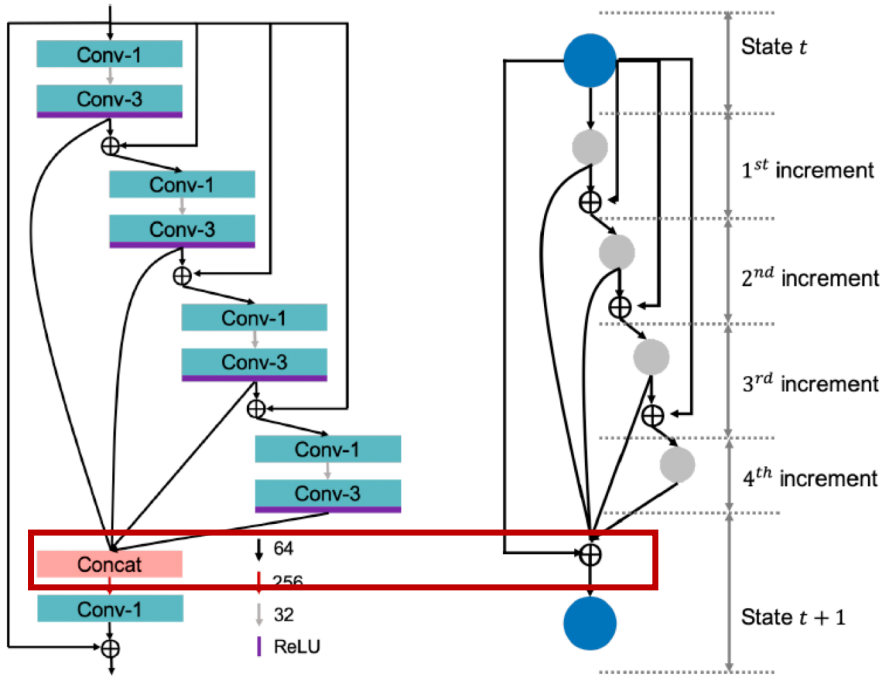
$$\mathbf{O}_i = \sigma(\mathbf{W}^i_{3\times3} * \mathbf{W}^i_{1\times1} * (\mathbf{X}_{in} + \mathbf{O}_{i-1})), \quad 2 \le i \le S \quad (11)$$

$$\mathbf{X}_{ms} = \mathbf{W}_{1\times1} * [\mathbf{O}_1, ..., \mathbf{O}_S] + \mathbf{X}_{in} \quad (12)$$



Model interpretation

- CNN-based SISR can be recast as optimal control
- Deep CNN corresponds to a dynamic system described by an ODE
- Feature propagation can be understood as applying an numerical method to solve the ODE
- Bridging HARTnet with 4th –order Runge-Kutta: smaller local truncation error, more accurate approximation to the dynamic system



(a) HART block (b) RK4 method

Minor difference in the aggregation step

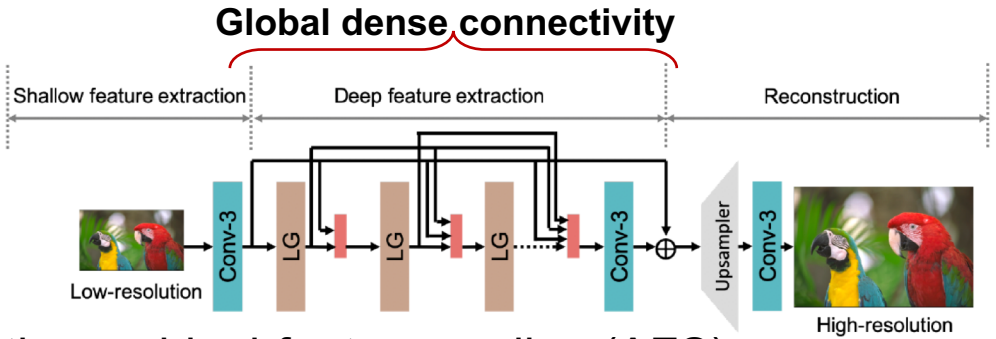
HARTnet: concatenation & 1x1 conv

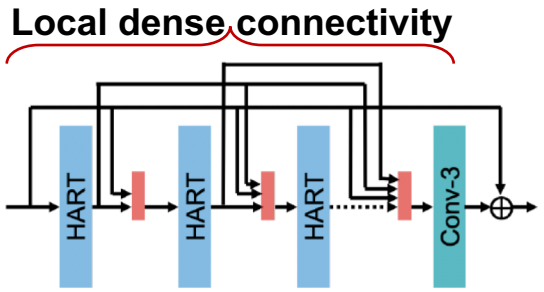
RK4: weighted averaging

Aggregation	Scale	Set5		Set14	
		PSNR	SSIM	PSNR	SSIM
Weighted Avg.	x4	32.39	0.896	28.69	0.784
Concat. & 1x1 Conv.	x4	32.50	0.900	28.80	0.790

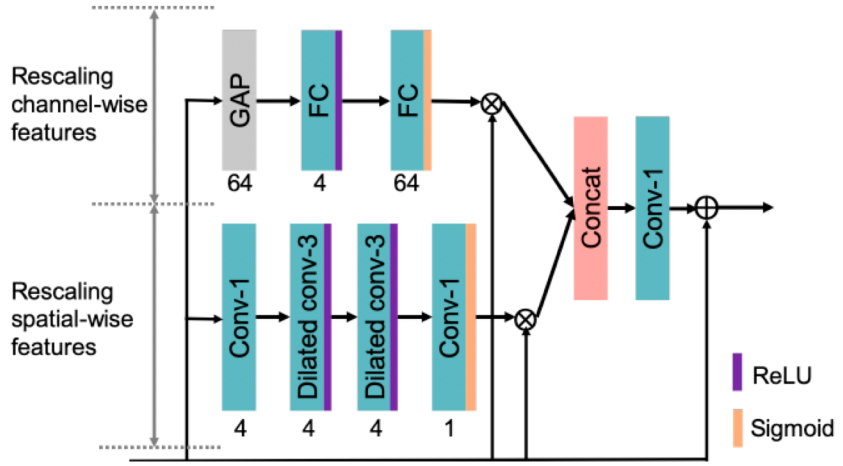
Building HARTnet by cascading HART blocks

- Local and global dense connectivity (DC)
  - Facilitate low-level feature reuse and preservation





- Adaptive residual-feature scaling (AFS)
  - Recalibrate both channel-wise and spatial-wise features to concentrate on the informative textural region

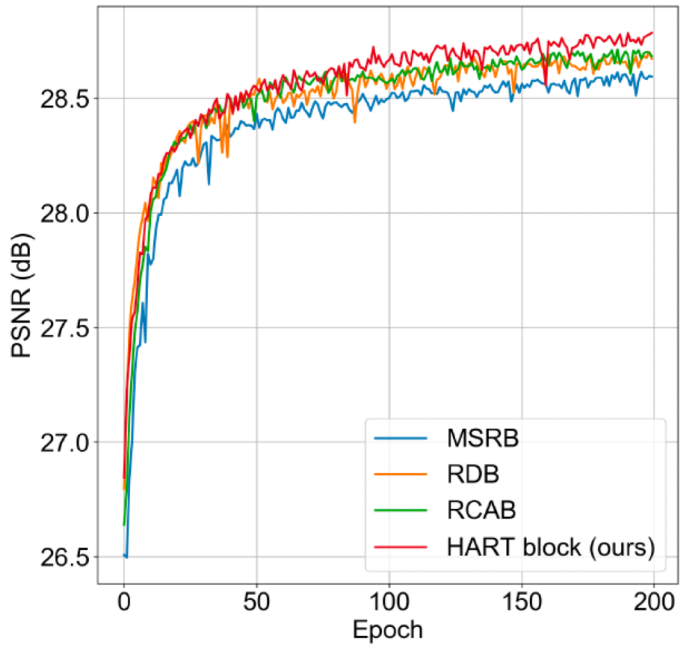


Experimental results

Ablation study

Module	Combination of HART, AFS, DC							
HART	X	✓	X	X	✓	✓	X	✓
AFS	X	X	✓	X	✓	X	✓	✓
DC	X	X	X	✓	X	✓	✓	✓
PSNR	32.22	32.39	32.39	32.41	32.42	32.45	32.43	32.50

Convergence

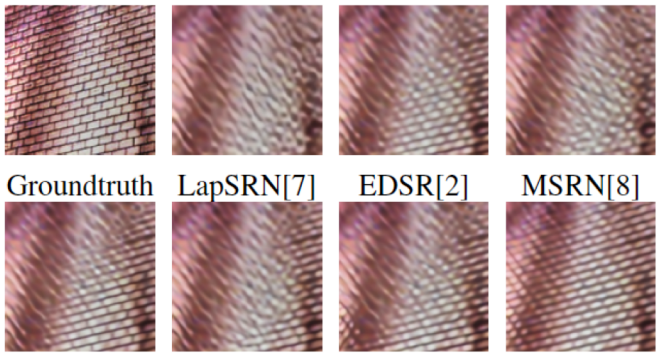


Benchmarks PSNR

Bicubic	x4	28.42/0.8104	26/0.7027	25.96/0.6675	23.14/0.6577
SRCNN[1]	x4	30.48/0.8628	27.50/0.7513	26.90/0.7101	24.52/0.7221
VDSR[6]	x4	31.35/0.8838	28.01/0.7674	27.29/0.7251	25.18/0.7524
LapSRN[7]	x4	31.54/0.8852	28.09/0.7700	27.32/0.7275	25.21/0.7562
MemNet[13]	x4	31.74/0.8893	28.26/0.7723	27.40/0.7281	25.50/0.7630
EDSR[2]	x4	32.46/0.8968	28.80/0.7876	27.71/0.7420	26.64/0.8033
MSRN[8]	x4	32.07/0.8903	28.60/0.7751	27.52/0.7273	26.04/0.7896
D-DBPN[17]	x4	32.47/0.8980	28.82/0.7860	27.72/0.7400	26.38/0.7946
RDN[3]	x4	32.47/0.8990	28.81/0.7871	27.72/0.7419	26.61/0.8028
SRFBN[18]	x4	32.47/0.8983	28.81/0.7868	27.72/0.7409	26.60/0.8015
OISR[10]	x4	32.53/0.8992	28.86/0.7878	27.75/0.7428	26.79/0.8068
HPBN[29]	x4	32.55/0.900	28.67/0.785	27.77/0.743	-
EDRN[19]	x4	32.24/0.8951	28.53/0.7811	27.54/0.7355	25.92/0.7831
HARTnet	x4	**32.71/0.900**	**28.93/0.790**	**27.80/0.745**	**26.91/0.809**

Visualization

Urban100:Img076

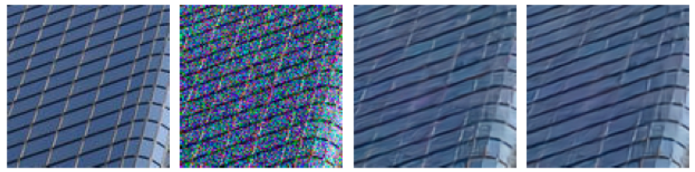


Other image restoration tasks

Denoising

Dataset	Noise Level	DnCNN [33]	RED [34]	MemNet [13]	SGN [35]	RNAN [36]	HARTnet
Kodak24	$\sigma = 30$	31.39	31.43	31.75	31.58	31.79	**31.84**
	$\sigma = 50$	29.16	29.10	29.38	29.36	29.52	**29.57**
	$\sigma = 70$	27.64	27.70	28.00	27.99	**28.12**	28.09
BSD68	$\sigma = 30$	30.40	30.33	30.45	30.45	30.57	**30.63**
	$\sigma = 50$	28.01	27.95	28.08	28.18	28.22	**28.28**
	$\sigma = 70$	26.56	26.50	26.59	26.79	26.79	**26.80**
Urban100	$\sigma = 30$	30.28	30.52	30.88	30.75	31.50	**31.62**
	$\sigma = 50$	28.16	27.98	28.60	28.36	29.08	**29.27**
	$\sigma = 70$	26.17	26.40	27.11	26.85	27.45	**27.56**

Urban100:Img067



Groundtruth | Noisy input | DnCNN[33] | RED[34]

MemNet[13] | SGN[35] | RNAN[36] | **HARTnet**

Low-light image enhancement

Method	CAN [37]	U-Net [38]	SGN [35]	HARTnet
SID-Sony	27.40 / 0.792	28.88 / 0.787	29.06 / -	**29.91 / 0.830**

Groundtruth

Traditional pipeline [39]

U-Net [38]

HARTnet

Summary

- A multi-scale HARTnet is proposed to deal with SISR task. By adopting hierarchically aggregated residual transformation blocks, HARTnet achieves superior SR performance
- The same architecture can handle various image restoration tasks: image denoising, low-light image enhancement
- Experiments and ablation studies show HARTnet achieves state-of-the-art performance