STRUCT

Learning Continuous Image Representation with Local Implicit Image Function

CVPR 2021 (Oral)

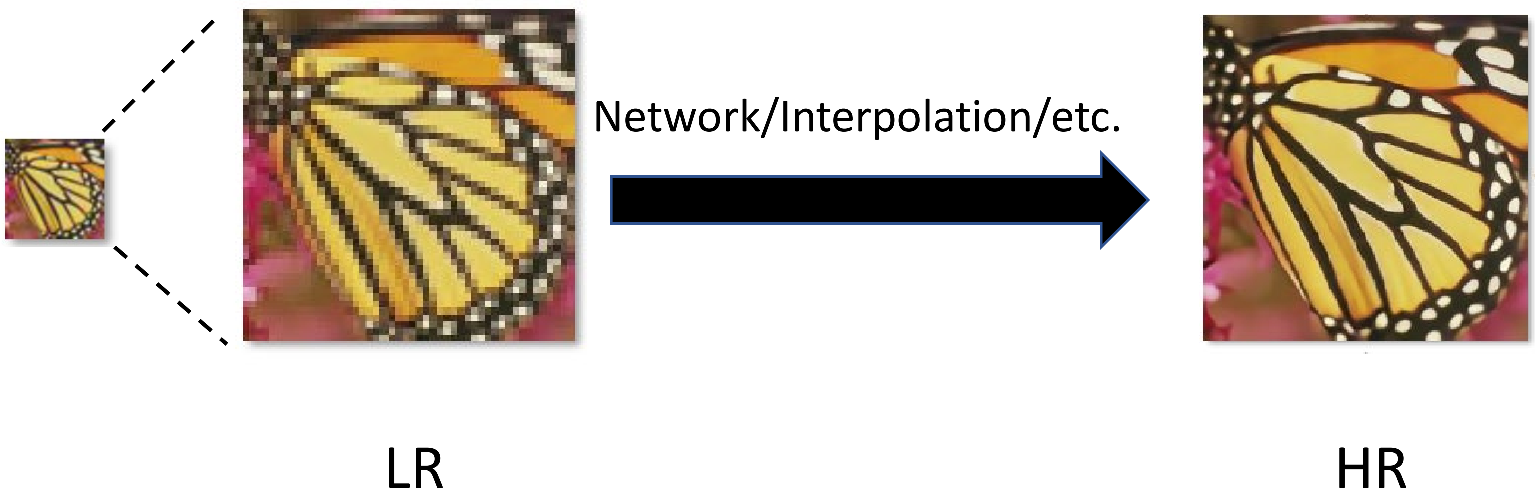
Yinbo Chen, Sifei Liu, Xiaolong Wang

Presented by Yuzhang Hu
2020.03.21

Outline

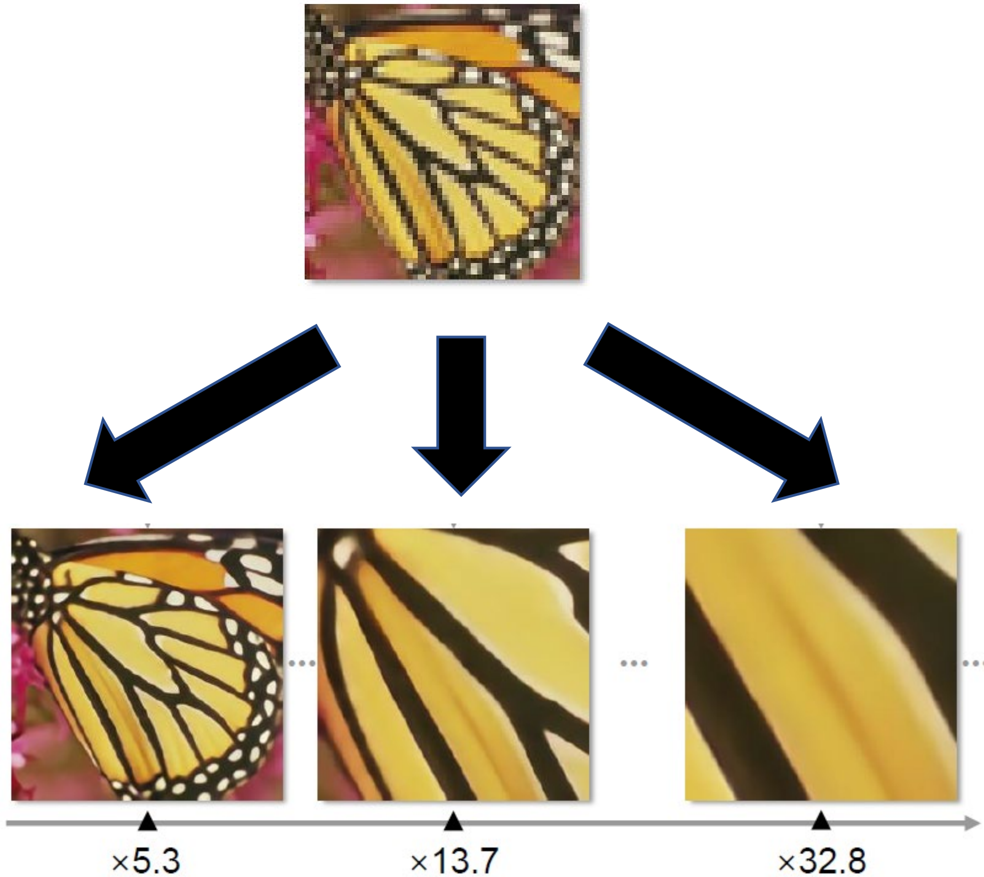
- Authorship
- Background
- Method
- Experiment
- Conclusion

Background



- Image super-resolution
  - Low resolution->high resolution

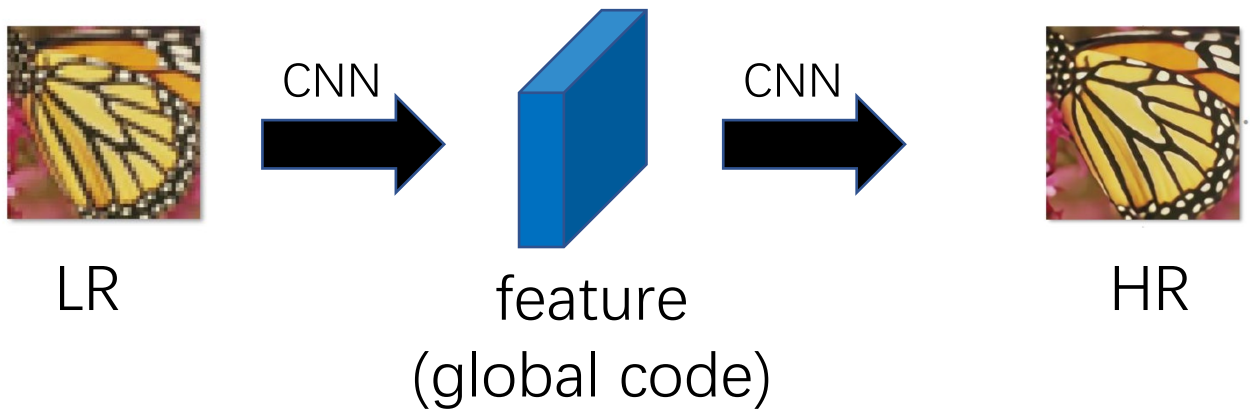
Background



- Difference with other enhancement task?
  - Others: one->one
  - SR: one->many

Background

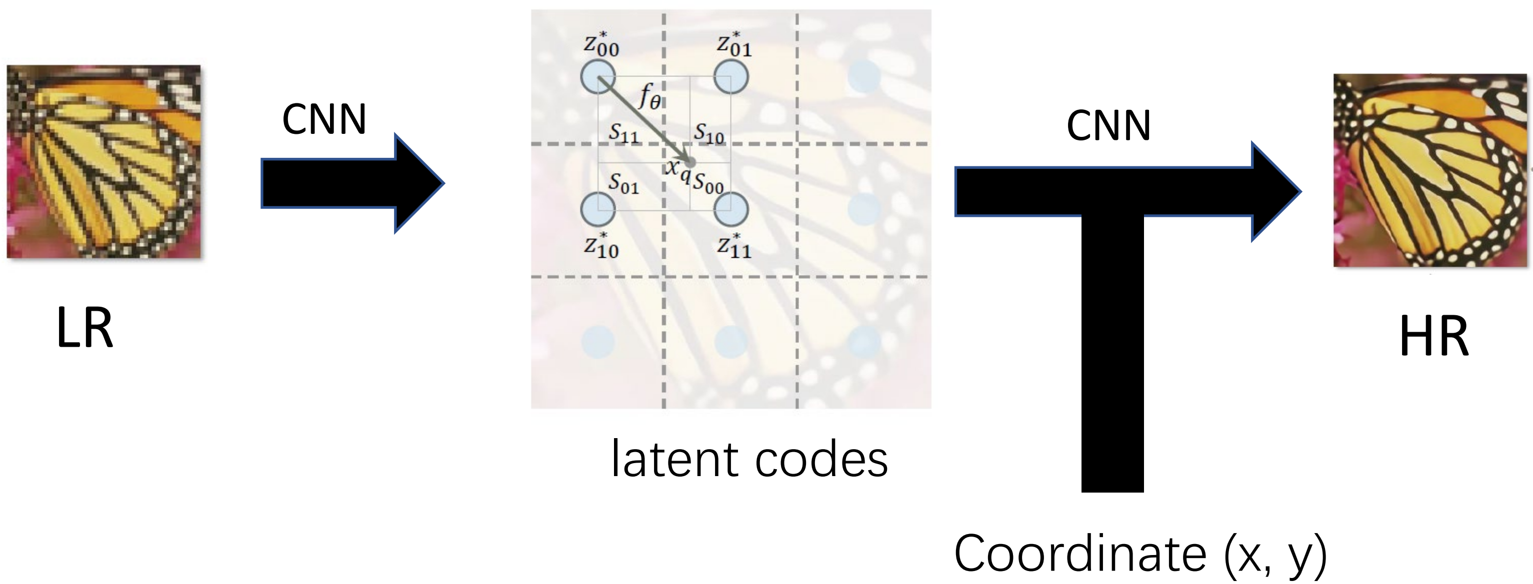
- Traditional methods



Different CNNs for different scales. (e.g. x2/x3/x4)

Method

- Query local latent codes



Method

- Local Implicit Image Function (LIIF)

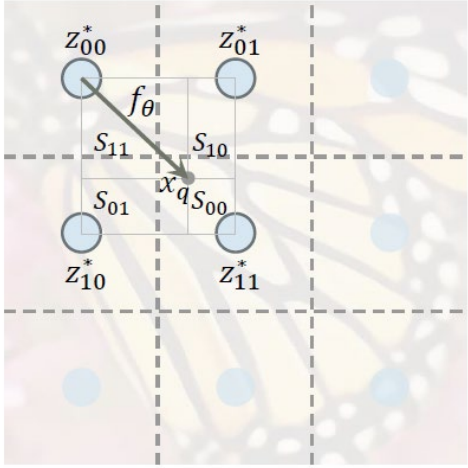
Implicit neural representation (z: vector, x: coordinate)

$$s = f(z, x)$$

$f(z, \cdot)$

The function that maps coordinate to RGB value.

RGB value definition: $I^{(i)}(x_q) = f(z^*, x_q - v^*),$



latent codes $M^{(i)}$

Method

- Further Improvement
  - Feature Unfolding

$$\hat{M}_{jk}^{(i)} = \mathrm{Concat}(\{M_{j+l,k+m}^{(i)}\}_{l,m\in\{-1,0,1\}})$$

enrich the information contained in each latent code

  - Local Ensemble

$$I^{(i)}(x_q) = \sum_{t\in\{00,01,10,11\}} \frac{S_t}{S} \cdot f(z_t^*, x_q - v_t^*),$$

achieve continuous transition at coordinates when z switches

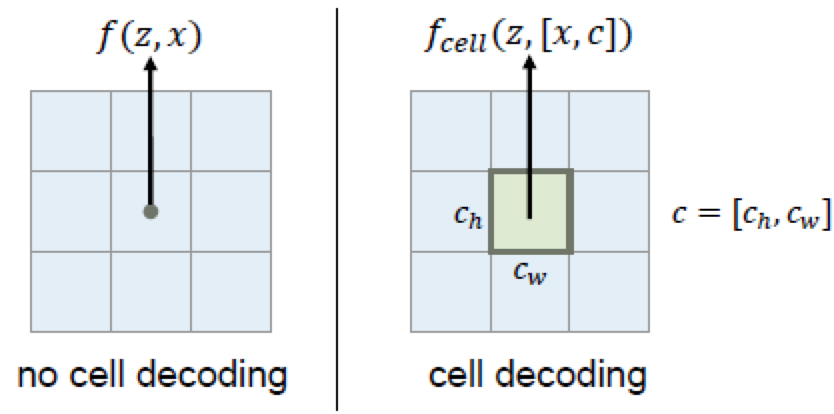
Method

- Further Improvement
  - Cell Decoding

$$s = f(z, x)$$

$$\Downarrow$$

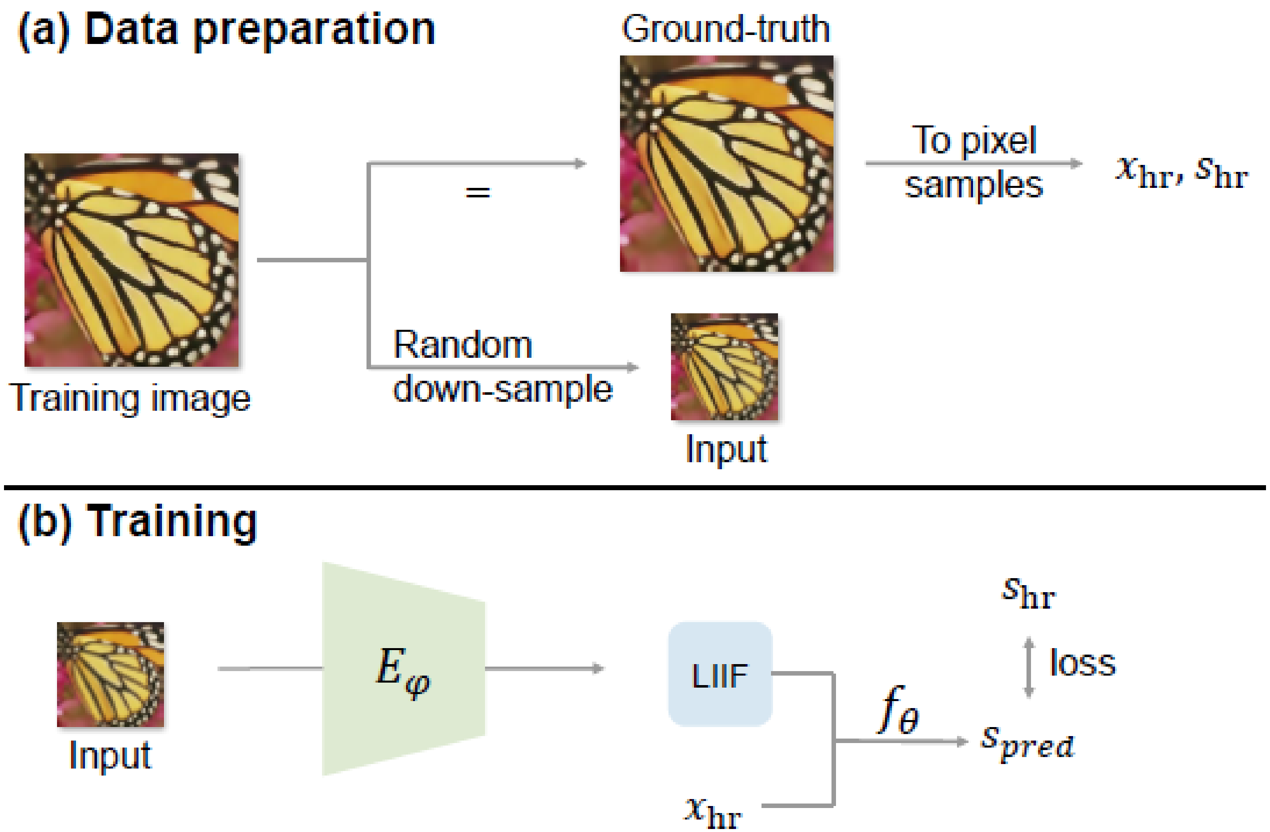
$$s = f_{cell}(z, [x, c])$$



$\Rightarrow$ present the LIIF representation as the pixel-based form in arbitrary resolution.

Method

- Learning Strategy



Experiment

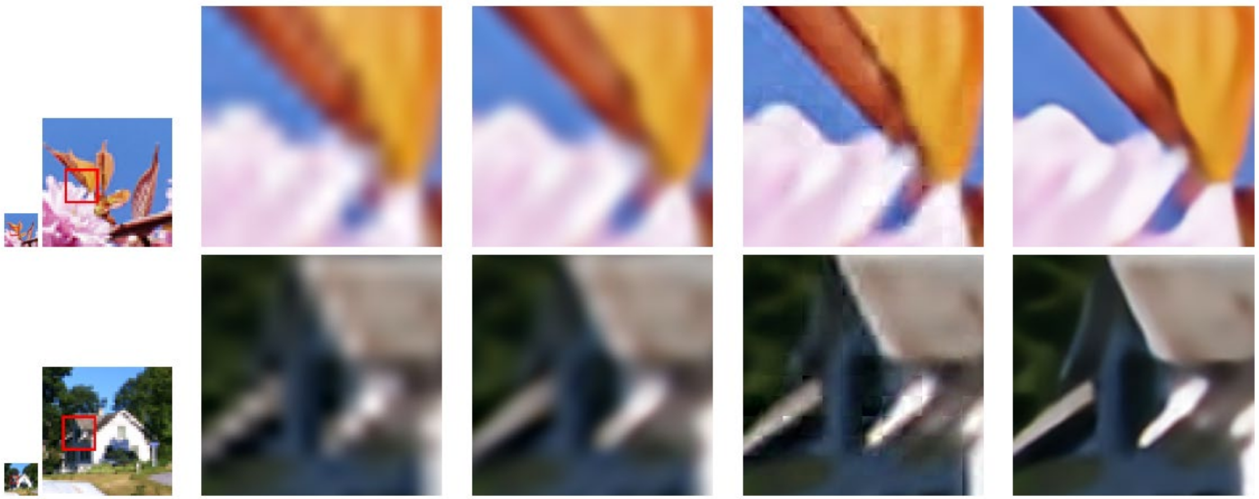
Quantitative comparison on benchmark datasets

Dataset	Method	In-distribution			Out-of-distribution	
		×2	×3	×4	×6	×8
Set5	RDN [47]	38.24	34.71	32.47	-	-
	RDN-MetaSR$^\sharp$ [13]	38.22	34.63	32.38	29.04	26.96
	RDN-LIIF (ours)	38.17	34.68	32.50	**29.15**	**27.14**
Set14	RDN [47]	34.01	30.57	28.81	-	-
	RDN-MetaSR$^\sharp$ [13]	33.98	30.54	28.78	26.51	24.97
	RDN-LIIF (ours)	33.97	30.53	28.80	**26.64**	**25.15**
B100	RDN [47]	32.34	29.26	27.72	-	-
	RDN-MetaSR$^\sharp$ [13]	32.33	29.26	27.71	25.90	24.83
	RDN-LIIF (ours)	32.32	29.26	27.74	**25.98**	**24.91**
Urban100	RDN [47]	32.89	28.80	26.61	-	-
	RDN-MetaSR$^\sharp$ [13]	32.92	28.82	26.55	23.99	22.59
	RDN-LIIF (ours)	32.87	28.82	**26.68**	**24.20**	**22.79**

- Competitive performance for in-distribution scales.
- Significantly outperforms MetaSR for out-distribution scales.

Experiment

Qualitative comparison for scale x30



Input (48px) crop=12px | Bicubic | 1-SIREN [37] | MetaSR [13] | LIIF (ours)

Other visualizations

Input (32px) | LIIF (presented in 640px)

Experiment

Ablation Study

	In-distribution			Out-of-distribution				
	$\times 2$	$\times 3$	$\times 4$	$\times 6$	$\times 12$	$\times 18$	$\times 24$	$\times 30$
LIIF	34.67	30.96	29.00	26.75	23.71	22.17	21.18	20.48
LIIF(-c)	34.53	30.92	28.97	26.73	23.72	22.19	21.19	20.51
LIIF(-u)	34.64	30.94	28.98	26.73	23.69	22.16	21.17	20.47
LIIF(-e)	34.63	30.95	28.97	26.72	23.66	22.13	21.14	20.45
LIIF(-d)	34.65	30.94	28.98	26.71	23.64	22.10	21.12	20.42

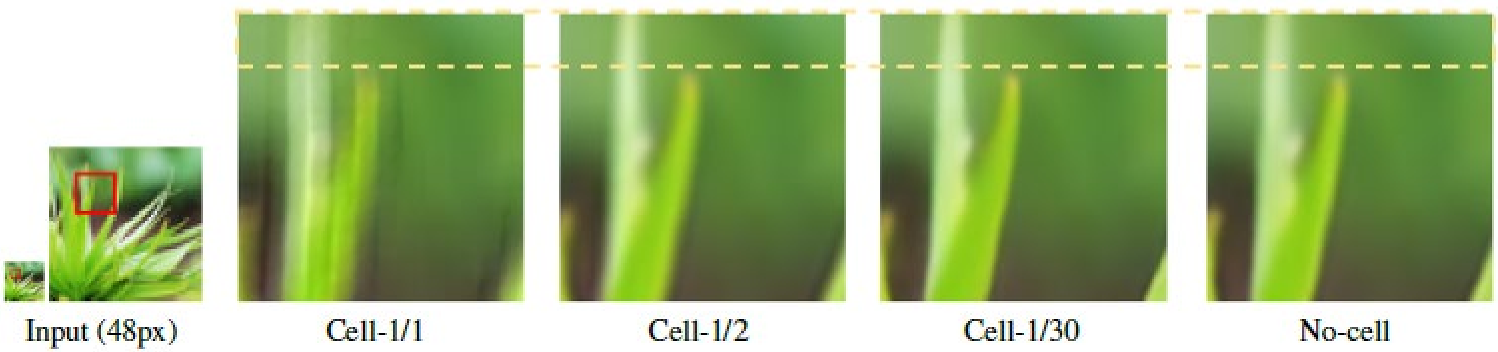
- -c: w/o cell decoding.
- -u: w/o feature unfolding.
- -e: w/o local ensemble
- -d: reducing the depth of neural implicit function

Experiment

Further Analysis on Cell Decoding

	In-distribution			Out-of-distribution				
	$\times 2$	$\times 3$	$\times 4$	$\times 6$	$\times 12$	$\times 18$	$\times 24$	$\times 30$
LIIF	34.67	30.96	29.00	26.75	23.71	22.17	21.18	20.48
LIIF(-c)	34.53	30.92	28.97	26.73	23.72	22.19	21.19	20.51

Cell decoding can hurt the performance for out-of-distribution scale



- For out-of-distribution large scales, there can be much more uncertainty.
- Being blurry may help a bit when uncertainty exists for PSNR computing.

Conclusion

- Neural implicit function for image representation in arbitrary resolution.
- Self-supervised training and better performance for out-of-distribution scales.

Thanks!