西安電子科技大學
XIDIAN UNIVERSITY

VIPSL Video & Image Processing System Lab

ACMmultimedia Nice, France 21-25 October 2019

# P4C-11 Lightweight Image Super-Resolution with Information Multi-distillation Network

Zheng Hui, Xinbo Gao, Yunchu Yang, Xiumei Wang
zheng_hui@aliyun.com, xbgao@mail.xidian.edu.cn, yc_yang@aliyun.com, wangxm@xidian.edu.cn

## ABSTRACT

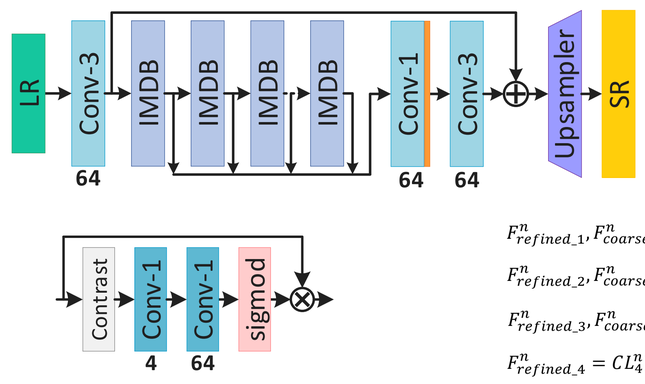

Low-resolution image

Super-Resolution


High-resolution image

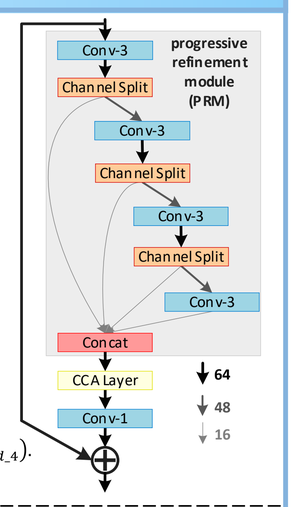
- Our information multi-distillation block (IMDB) with contrast-aware attention (CCA) layer.
- The adaptive cropping strategy (ACS) to achieve the processing image of any arbitrary size (implementing any upscaling factors using one model).
- The exploration of factors affecting actual inference time.

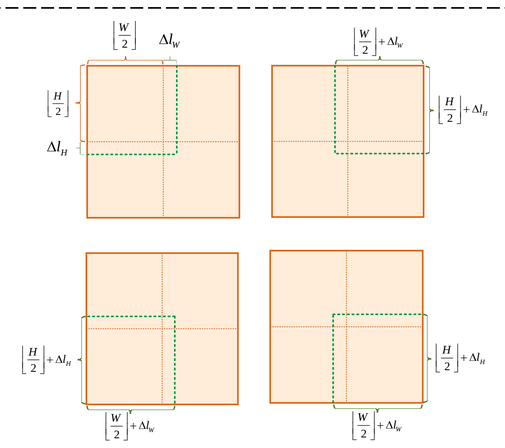
## FRAMEWORK

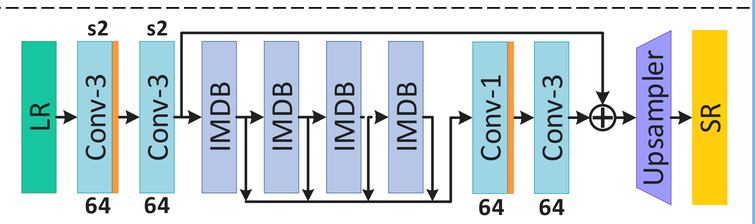



$$Contrast = Mean + Std$$

$$F^n_{refined\_1}, F^n_{coarse\_1} = Split^n_1(CL^n_1(F^n_{in})),$$
$$F^n_{refined\_2}, F^n_{coarse\_2} = Split^n_2\left(CL^n_2(F^n_{coarse\_1})\right),$$
$$F^n_{refined\_3}, F^n_{coarse\_3} = Split^n_3\left(CL^n_3(F^n_{coarse\_2})\right),$$
$$F^n_{refined\_4} = CL^n_4(F^n_{coarse\_3}),$$
$$F^n_{distilled} = Concat(F^n_{refined\_1}, F^n_{refined\_2}, F^n_{refined\_3}, F^n_{refined\_4}).$$

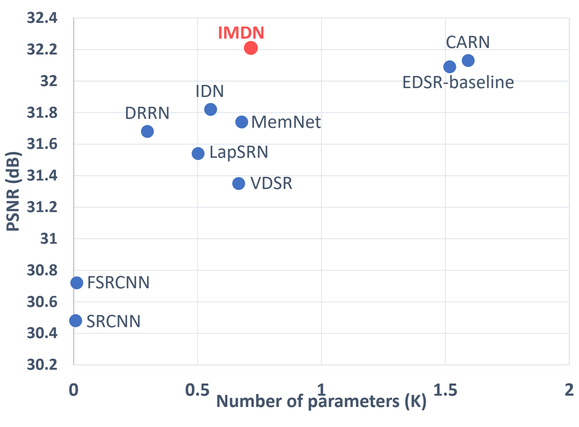


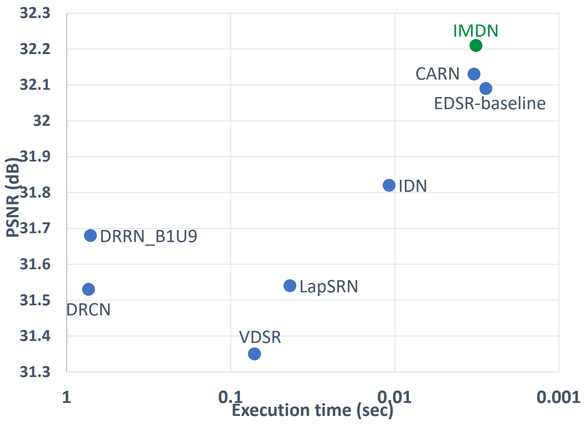






$$\left(\left\lfloor\frac{H}{2}\right\rfloor + \Delta l_H\right)\%4 = 0,$$
$$\left(\left\lfloor\frac{W}{2}\right\rfloor + \Delta l_H\right)\%4 = 0.$$
$$\Delta l_H = padding_H - \left(\left\lfloor\frac{H}{2}\right\rfloor + padding_H\right)\%4,$$
$$\Delta l_W = padding_W - \left(\left\lfloor\frac{W}{2}\right\rfloor + padding_W\right)\%4,$$
$$padding_H = padding_W = 4k, k \geq 1.$$

## PARAMETERS AND EXECUTION TIME




*The parameters and execution time for 4× SR on Set5*

## AVERAGE PSNR/SSIM RESULTS

| Method | Scale | Params | Set5 | Set14 | BSD100 | Urban100 | Manga109 |
|---|---|---|---|---|---|---|---|
| EDSR-baseline | × 2 | 1,370K | 37.99 / 0.9604 | 33.57 / 0.9175 | 32.16 / 0.8994 | 31.98 / 0.9272 | 38.54 / 0.9769 |
| SRMDNF | | 1,511K | 37.79 / 0.9601 | 33.32 / 0.9159 | 32.05 / 0.8985 | 31.33 / 0.9204 | 38.07 / 0.9761 |
| CARN | | 1,592K | 37.76 / 0.9590 | 33.52 / 0.9166 | 32.09 / 0.8978 | 31.92 / 0.9256 | 38.36 / 0.9765 |
| IMDN | | 694K | **38.00** / **0.9605** | **33.63** / **0.9177** | **32.19** / **0.8996** | **32.17** / **0.9283** | **38.88** / **0.9774** |
| EDSR-baseline | × 3 | 1,555K | **34.37** / **0.9270** | 30.28 / **0.8417** | **29.09** / **0.8052** | 28.15 / **0.8527** | 33.45 / 0.9439 |
| SRMDNF | | 1,528K | 34.12 / 0.9254 | 30.04 / 0.8382 | 28.97 / 0.8025 | 27.57 / 0.8398 | 33.00 / 0.9403 |
| CARN | | 1,592K | 34.29 / 0.9255 | 30.29 / 0.8407 | 29.06 / 0.8034 | 28.06 / 0.8493 | 33.50 / 0.9440 |
| IMDN | | 703K | 34.36 / **0.9270** | **30.32** / **0.8417** | **29.09** / 0.8046 | **28.17** / 0.8519 | **33.61** / 0.9445 |
| EDSR-baseline | × 4 | 1,518K | 32.09 / 0.8938 | 28.58 / **0.7813** | 27.57 / 0.7357 | 26.04 / 0.7849 | 30.35 / 0.9067 |
| SRMDNF | | 1,552K | 31.96 / 0.8925 | 28.35 / 0.7787 | 27.49 / 0.7337 | 25.68 / 0.7731 | 30.09 / 0.9024 |
| CARN | | 1,592K | 32.13 / 0.8937 | **28.60** / 0.7806 | **27.58** / 0.7349 | **26.07** / 0.7837 | **30.47** / **0.9084** |
| IMDN | | 715K | **32.21** / **0.8948** | 28.58 / 0.7811 | 27.56 / 0.7353 | 26.04 / 0.7838 | 30.45 / 0.9075 |

## MORE EXPLORATIONS

*Memory Consumption (MB) and average inference time (second)*

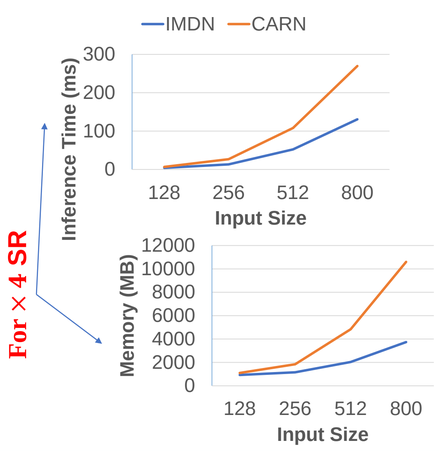
| Method | Scale | Params | Depth | BSD100 Memory / Time | Urban100 Memory / Time | Manga109 Memory / Time |
|---|---|---|---|---|---|---|
| EDSR-baseline | × 4 | 1.6M | 37 | 665 / 0.00295 | 2,511 / 0.00242 | 1,219 / 0.00232 |
| EDSR | | 43M | 69 | 1,531 / 0.00580 | 8,863 / 0.00416 | 3,703 / 0.00380 |
| RDN | | 22M | 150 | 1,123 / 0.01626 | 3,335 / 0.01325 | 2,257 / 0.01300 |
| RCAN | | 16M | 415 | 777 / 0.09174 | 2,631 / 0.55280 | 1,343 / 0.72250 |
| CARN | | 1.6M | 34 | 945 / 0.00278 | 3,761 / 0.00305 | 2,803 / 0.00383 |
| IMDN | | 0.7M | 34 | 671 / 0.00285 | 1,155 / 0.00284 | 895 / 0.00279 |

** Note: the inference time is tested by time.time() in Python*

*Pressure test*

| Scale | Input size | IMDN | CARN |
|---|---|---|---|
| × 2 | 128 × 128 | 4.1396 ms / 931 MB | 4.1963 ms / 975 MB |
| | 256 × 256 | 12.9499 ms / 1,153 MB | 16.7976 ms / 1,333 MB |
| | 512 × 512 | 50.9001 ms / 2,035 MB | 67.6479 ms / 2,789 MB |
| | 1024 × 1024 | 207.6811 ms / 5,547 MB | 282.8919 ms / 8,581 MB |
| × 3 | 128 × 128 | 4.1336 ms / 929 MB | 4.9804 ms / 1,029 MB |
| | 256 × 256 | 12.9492 ms / 1,153 MB | 19.8745 ms / 1,557MB |
| | 512 × 512 | 50.8829 ms / 2,035 MB | 81.0804 ms / 3,685 MB |
| | 1024 × 1024 | 207.3989 ms / 5,547 MB | N.A / > 11G |

** Note: the inference time is tested by torch.cuda.Event()*



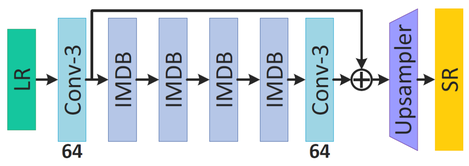

## ABLATION STUDY

*Investigations of PRM, CCA module and IIC scheme*

| Scale | PRM | CCA | IIC | Params | Set5 | Set14 | BSD100 | Urban100 | Manga109 |
|---|---|---|---|---|---|---|---|---|---|
| × 4 | ✗ | ✗ | ✗ | 510K | 31.86 / 0.8901 | 28.43 / 0.7775 | 27.45 / 0.7320 | 25.63 / 0.7711 | 29.92 / 0.9003 |
| | ✓ | ✗ | ✗ | 480K | 32.01 / 0.8927 | 28.49 / 0.7792 | 27.50 / 0.7338 | 25.81 / 0.7773 | 30.16 / 0.9038 |
| | ✓ | ✓ | ✗ | 482K | 32.10 / **0.8934** | 28.51 / 0.7794 | 27.50 / 0.7338 | 25.89 / 0.7793 | 30.25 / 0.9050 |
| | ✓ | ✓ | ✓ | 499K | **32.11** / **0.8934** | **28.52** / **0.7797** | **27.53** / **0.7342** | **25.90** / **0.7797** | **30.28** / **0.9054** |

*Comparison with channel attention (CA) and the contrast-aware channel attention (CCA)*

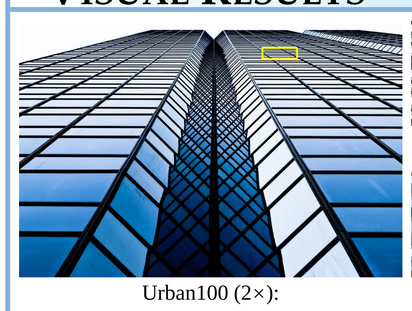
| Module | Set5 | Set14 | BSD100 | Urban100 |
|---|---|---|---|---|
| IMDN_basic_B4 + CA | 32.0821 | 28.5086 | 27.5124 | 25.8829 |
| IMDN_basic_B4 + CCA | **32.0964** | **28.5118** | **27.5185** | **25.8916** |

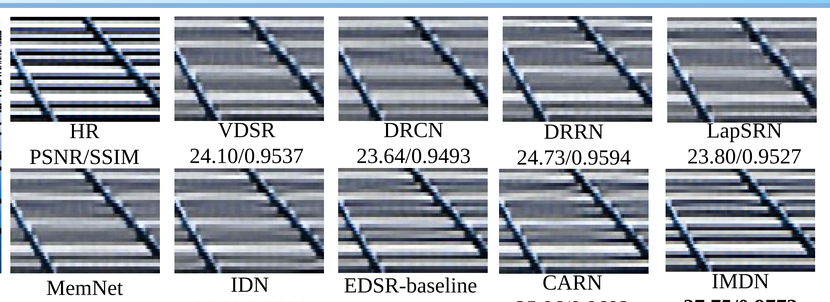


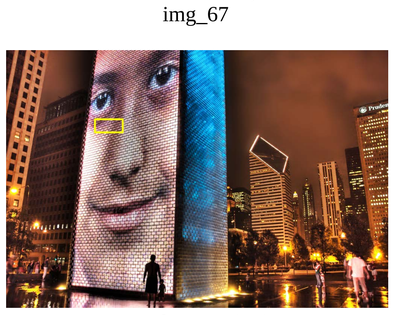

*Study of adaptive cropping strategy (ACS)*

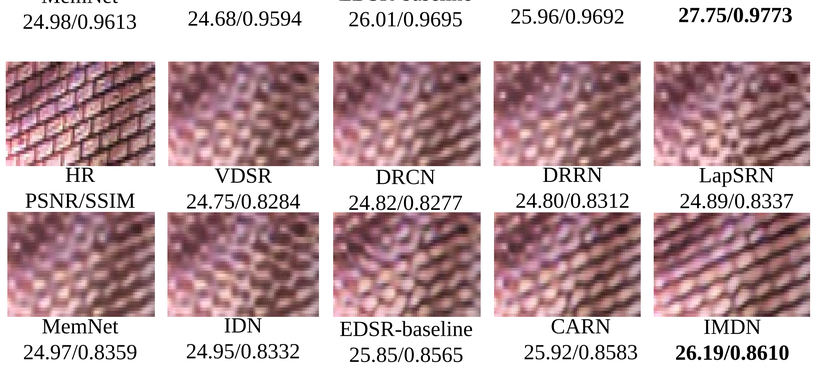
| Method | PSNR | SSIM | LPIPS | Time | Memory |
|---|---|---|---|---|---|
| VDSR | 28.75 | 0.8439 | 0.2417 | 0.0290 | 7,855M |
| IMDN_AS | **29.35** | **0.8595** | **0.2147** | **0.0041** | **3,597M** |

## VISUAL RESULTS


Urban100 (2×): img_67





Urban100 (3×): img_76




*Visual comparisons of IMDN with other SR methods on Urban100 datasets*

## FLOPS

*The computational costs*

| Scale | LapSRN (CVPR'17) | IDN (CVPR'18) | EDSR-b (CVPRW'17) | CARN (ECCV'18) | IMDN (Our) |
|---|---|---|---|---|---|
| × 2 | **112K** | 175K | 341K | 157K | 173K |
| × 3 | 76K | **75K** | 172K | 90K | 78K |
| × 4 | 76K | 51K | 122K | 76K | **45K** |

## CODE