SR-CACO-2: A Dataset for Confocal Fluorescence Microscopy Image Super-Resolution

Soufiane Belharbi[1*] & Mara KM Whitford[2,3] & Phuong Hoang[2] & Shakeeb Murtaza[1] & Luke McCaffrey[2,3,4] & Eric Granger[1]

[1]LIVIA, ILLS, Dept. of Systems Engineering, ETS Montreal, Canada
[2]Goodman Cancer Institute, McGill University, Montreal, Canada
[3]Dept. of Biochemistry, McGill University, Montreal, Canada
[4]Gerald Bronfman Dept. of Oncology, McGill University, Montreal, Canada

Context

Why microscopy image super-resolution via machine learning?

- Fast low resolution imaging
- Reduce phototoxicity: cell damage/death due to long exposure to light to acquire high resolution images
- Reduce photobleaching
- Allow for the observation of instantaneous inter-cellular events with less damage to the cells
- Live imaging videos with low quality under reduced light exposure

Existing microscopy super-resolution datasets are:

- Very few
- Private
- May focus on high-end imaging techniques, like SIM (Structured Illumination Microscopy), BioSR dataset.

This limits research in machine learning-based super-resolution field.

Why confocal microscopy? ☞Widely accessible

SR-CACO-2: Capture

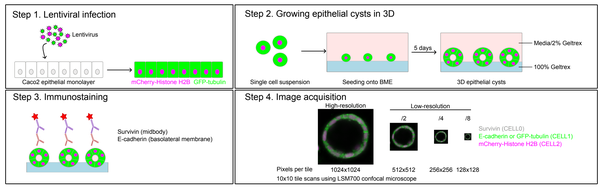


Figure 1. Methodology for capturing the SR-CACO-2 dataset.

Preparation for SISR task: patching

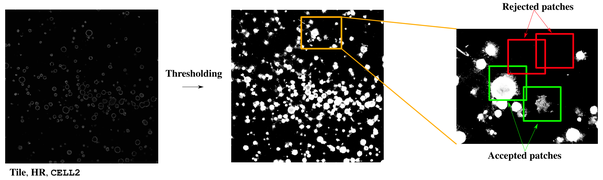


Figure 2. Pre-processing of tiles to patches.

SR-CACO-2: Diversity

- SR-CACO-2 contains the human epithelial cell line Caco-2 (ATCC HTB-37)
- Real pairs: LR-HR
- 4 scales: HR, LR (/2, /4, /8)
- 3 protein markers: Survivin (CELL0), E-cadherin or GFP-tubulin (CELL1), mCherry-Histone H2B (CELL2)
- 16,800 multi-cellular objects
- Per-scale: 22,00 unique images (22 tiles), +9k patches (512×512)
- License: Freely accessible under CC BY-NC-SA 4.0

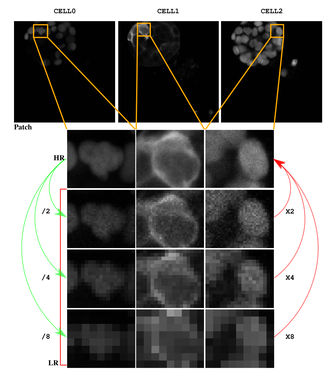


Figure 3. Illustration of SR-CACO-2 patch content for cells CELL0, CELL1, and CELL2, and for the HR patch and its corresponding three LR patches (/2, /4, /8).

	Tiles				Patches			
Data subsets	Train	Validation	Test	Total	Train	Validation	Test	Total
Number	15	3	4	22	7,349	1,117	1,471	9,937
Image size:								
HR	$\sim 9,318 \times 9,318$				512×512			
LR /2	$\sim 4,658 \times 4,658$				256×256			
LR /4	$\sim 2,328 \times 2,328$				128×128			
LR /8	$\sim 1,164 \times 1,164$				64×64			
File size:								
HR	260.5 MB				262.4 KB			
LR /2	65.1 MB				65.8 KB			
LR /4	16.3 MB				16.6 KB			
LR /8	4.1 MB				4.4 KB			

Table 1. Split of SR-CACO-2 into the train, validation and test subsets, along with relevant statistics. Numbers are defined per scale (X2, X4, X8, and HR), and per cell type (CELL0, CELL1, and CELL2).

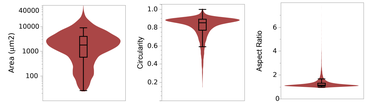


Figure 4. Object-based analysis of cellular structures captured of all the 22,00 high-resolution images ($22 \times 10 \times 10$).

Benchmarking: Results

16 SISR methods.

- Super-resolution task

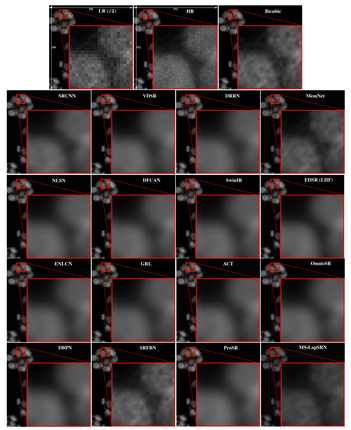


Figure 5. Illustrative super-resolution visual results for X2, CELL2 across all SISR models.

SISR Methods	Scale	PSNR ↑ CELL0	CELL1	CELL2	Mean	NRMSE ↓ CELL0	CELL1	CELL2	Mean	SSIM ↑ CELL0	CELL1	CELL2	Mean
Bicubic	X2	35.02	32.15	30.38	32.52	0.1085	0.0601	0.0724	0.0803	0.7618	0.7658	0.6891	0.7389
	X4	35.46	32.03	31.10	32.86	0.0985	0.0586	0.0660	0.0744	0.8206	0.8002	0.7673	0.7960
	X8	31.88	27.50	26.10	28.49	0.1655	0.1139	0.1349	0.1381	0.6683	0.6266	0.6511	0.6487
Pre-upsampling SR													
SRCNN *(eccv,2014)*	X2	37.54	34.27	33.42	35.08	0.0710	0.0450	0.0500	0.0553	0.8517	0.8524	0.8210	0.8417
	X4	36.14	32.73	32.25	33.71	0.0817	0.0528	0.0572	0.0639	0.8522	0.8216	0.8079	0.8272
	X8	33.05	28.04	26.49	**29.19**	0.1265	0.0967	0.1220	**0.1151**	0.7711	0.7085	0.7092	**0.7296**
Post-upsampling SR													
Omni-SR *(cvpr,2023)*	X2	37.70	34.11	33.51	35.11	0.0759	0.0461	0.0496	0.0572	0.8744	0.8539	0.8313	0.8532
	X4	36.44	32.59	32.34	33.79	0.0849	0.0536	0.0563	0.0649	0.8592	0.8203	0.8111	0.8302
	X8	30.75	27.16	25.30	27.74	0.1713	0.1098	0.1352	0.1387	0.6715	0.6419	0.6591	0.6575
Iterative up-and-down sampling SR													
SRFBN *(cvpr,2019)*	X2	36.13	33.15	31.61	33.63	0.0955	0.0531	0.0625	0.0704	0.8078	0.8091	0.7470	0.7880
	X4	36.08	32.52	31.79	33.46	0.0911	0.0545	0.0605	0.0687	0.8405	0.8147	0.7889	0.8147
	X8	32.27	27.78	26.47	28.84	0.1560	0.1091	0.1278	0.1310	0.7022	0.6549	0.6904	0.6825
Progressive upsampling SR													
MS-LapSRN *(tpami,2019)*	X2	33.88	32.36	29.34	31.86	0.1130	0.0535	0.0791	0.0819	0.7652	0.8164	0.7695	0.7837
	X4	30.80	30.99	31.08	30.96	0.1192	0.0615	0.0626	0.0811	0.7885	0.7837	0.7806	0.7843
	X8	31.83	27.14	25.06	28.01	0.1404	0.0982	0.1323	0.1236	0.7478	0.6933	0.6640	0.7017

Table 2. The performance of SISR methods on the SR-CACO-2 test set of ROI only, *i.e.*, cells. See paper for the full 16 methods.

Access, code, arXiv

Figure 6. Access, code: github.com/sbelharbi/sr-caco-2

Figure 7. Full arXiv paper: arxiv.org/pdf/2406.09168

Benchmarking: Results

- Downstream task: cell detection/segmentation

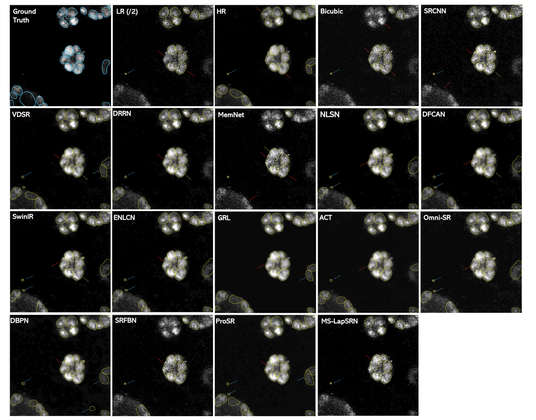
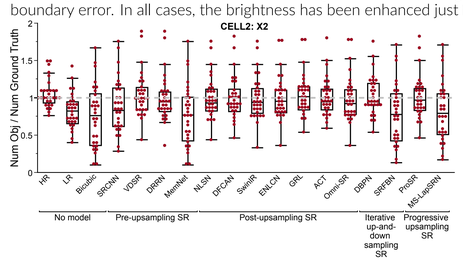


Figure 8. Cell segmentation example using different methods: CELL2, X2. Red arrow for undersegmented errors; Blue arrow for oversegmented error; and green arrow for boundary error. In all cases, the brightness has been enhanced just for visualization.



Model versus HR

Level	Score Mean Diff	p-Value
LR	-18	0.001
Bicubic	-21.3667	<.0001
SRCNN	-16.5667	0.0034
VDSR	7.0333	0.6868
DRRN	2.3333	1
MemNet	-18.7667	0.0005
NLSN	-4.1333	0.992
DFCAN	-3.3667	0.9992
SwinIR	-29.9667	<.0001
ENLCN	-3.7	0.9975
GRL	-14.7667	0.014
ACT	-7.7	0.5677
Omni-SR	-4.0667	0.9932
DBPN	-0.6333	1
SRFBN	-19.3667	0.0003
ProSR	-0.8333	1
MS-LapSRN	-17.8333	0.0011

Figure 9. Analysis of cell detection performance for CELL2, X2 over 30 random test samples (red dots).

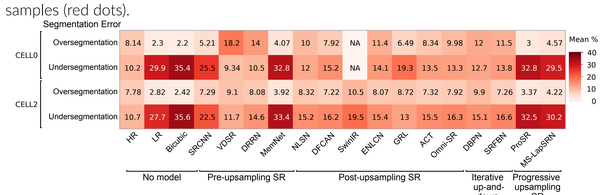


Figure 10. Analysis of cell segmentation performance: CELL0, CELL2, X2.