W2S: A Joint Denoising and Super-Resolution Dataset Supplementary Material

## 1 Structured-Illumination Microscopy (SIM)

We provide an introduction to the principle of structured interference acquisition with 1D signals. The extension to higher dimensions follows the same principle [1].

We define $s(t)$ to be the signal we want to acquire, where t represents a certain spatial dimension. In the Fourier domain, the corresponding signal $S(\omega)$ is not necessarily band-limited and can be non-zero for arbitrary frequencies ω. The impulse response of the capturing system is called its point spread function, and its Fourier transform is its optical transfer function (OTF) that we call $F(\omega)$. The resulting visible signal through that imaging system is given by $V(\omega) = F(\omega) \odot S(\omega)$, where $\odot$ is the element-wise multiplication. The OTF limits the captured content to a certain range of frequency components as $F(\omega) = 0 \ \forall \omega > \omega_c$, where ω_c is the cut-off frequency of the OTF. Therefore, only frequency components $\omega < \omega_c$ can be captured. By using a structured-illumination pattern, the frequency content can be manipulated. For instance, if we multiply the signal $s(t)$ with a cosine function of frequency ω_0, the captured signal becomes $s(t) \odot cos(\omega_0 t)$ and the corresponding frequency-domain equivalent is given by $\frac{1}{2}[S(\omega - \omega_0) + S(\omega + \omega_0)]$. Using the same imaging system, the visible signal becomes

$$V(\omega) = F(\omega) \odot \frac{1}{2}[S(\omega - \omega_0) + S(\omega + \omega_0)], \tag{1}$$

with $F(\omega)$ still equal to zero above its cut-off frequency ω_c. However, frequency components such that $\omega - \omega_0 < \omega_c$ can now be acquired, effectively pushing the cut-off to $\omega_c + \omega_0$, where ω_0 can be controlled by modifying the periodicity of the illumination pattern. In other words, higher frequencies that could not be visible to the imaging system can be shifted down to lower ones that lie within the observable range of that system. The shifted components can overlap in the frequency domain, and multiple shifted captures are needed to resolve the ambiguity and recover the true signals. SIM acquires 15 different structured-illumination images to perform an upscaling by a factor of two. In practice, the OTF is also not necessarily an ideal low-pass filter and a deconvolution post-processing step can be required.

Note that theoretically, applying nonlinear SIM can produce images of arbitrary resolution [2], illustrating SIM's potential of producing a higher upscaling-factor SR dataset.

(a) SIM input captures   (b) SIM reconstruction
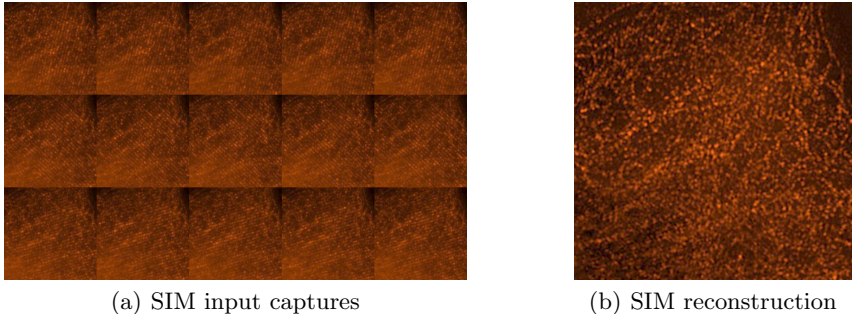
Fig. 1. Example FOV showing the different captured images in (a) that are given as input to the SIM method, and the reconstructed result of SIM in (b). Gamma correction is applied for better visualization.

2   Acquisition Details

Widefield Images To capture the widefield images, the microscope is fitted with a Plan Apochromat TIRF 100X, 1.49NA objective and an electron-multiplying charge-coupled device camera (IXON3; Andor Technology). The acquisition is taken with a 5ms exposure time using a 488nm Coherent sapphire laser at 0.37mW, a 5ms exposure time using a 561nm Cobolt Laser at 0.28mW, and a 5ms exposure using a Coherent 640nm Cobolt Laser at 0.26mW. In total, we capture 120 different fields-of-view (FOVs) on 3 different wavelength values, and each FOV is repeatedly captured 400 consecutive times. All images have a resolution of 512×512.

SIM Imaging The SIM images are captures using the same device (a microscope fitted with a Plan Apochromat TIRF 100X). We use the 3D SIM acquisition mode [3] (15 images per plane; five phases of three rotations) with an 70ms exposure time, using a 488nm Coherent sapphire laser at 0.20mW; 30ms exposure time using a 561nm Cobolt Laser at 0.27mW, and a 100ms exposure using a Coherent 640nm Cobolt Laser at 0.14mW. Image reconstruction and processing are performed using the NIS-Elements software. The example of the acquired images and the reconstructed HR results are shown in Fig. 1.

3   Training Details

Denoisers DnCNN and MemNet use a batch size of 128 and a starting learning rate of 10^{-3}, while RIDNet uses batches of 64 patches and a starting learning rate of 5×10^{-4}, all trained with the Adam optimizer [4] for 50 epochs, and with a ten-fold decrease in the learning rate after the milestone of 30 epochs. The same settings are used when training for the noise levels corresponding to an average of 1, 2, 4, 8, and 16 raw images.

Super-resolution networks For the super-resolution(SR) networks we benchmark, we train them with the initial learning rate and loss function as described

		Super-resolution networks					
		RCAN	RDN	SAN	SRFBN	EPSR	ESRGAN
Denoisers on 2× average	PURE-LET	22.04/0.66	22.97/0.67	22.84/0.67	22.70/0.66	22.54/0.60	23.01/0.67
	VST+EPLL	24.56/0.72	25.04/0.73	24.84/0.72	24.57/0.71	23.92/0.64	24.46/0.71
	VST+BM3D	24.82/0.72	25.09/0.73	24.82/0.72	24.78/0.72	24.01/0.63	24.58/0.71
	DnCNN†	24.77/0.72	25.04/0.72	24.60/0.70	24.86/0.71	24.13/0.64	24.95/0.72
	MemNet†	24.78/0.72	25.03/0.72	24.65/0.71	24.87/0.71	24.11/0.64	24.96/0.71
	RIDNet†	25.01/0.73	25.08/0.73	25.01/0.73	25.06/0.73	24.11/0.64	24.72/0.71
Denoisers on 4× average	PURE-LET	23.21/0.70	24.07/0.71	23.93/0.71	23.62/0.70	23.28/0.63	23.60/0.69
	VST+EPLL	25.01/0.73	25.41/0.74	25.12/0.73	25.00/0.73	24.15/0.65	24.81/0.72
	VST+BM3D	25.21/0.74	25.46/0.74	25.16/0.73	25.19/0.73	24.23/0.64	24.91/0.73
	DnCNN†	25.18/0.73	25.39/0.74	25.02/0.72	25.29/0.73	24.31/0.64	25.23/0.73
	MemNet†	25.19/0.73	25.38/0.74	25.10/0.72	25.25/0.73	24.36/0.64	25.22/0.72
	RIDNet†	25.06/0.73	25.29/0.73	24.92/0.72	25.20/0.72	24.24/0.63	25.13/0.72
Denoisers on 8× average	PURE-LET	24.06/0.72	24.88/0.73	24.75/0.74	24.75/0.73	23.91/0.65	23.99/0.72
	VST+EPLL	25.40/0.74	25.74/0.75	25.39/0.74	25.40/0.74	24.35/0.65	25.14/0.73
	VST+BM3D	25.56/0.75	25.77/0.75	25.45/0.73	25.54/0.74	24.42/0.65	25.22/0.73
	DnCNN†	25.29/0.74	25.56/0.74	25.07/0.72	25.42/0.73	24.47/0.64	25.46/0.73
	MemNet†	25.46/0.74	25.65/0.74	25.33/0.73	25.57/0.73	24.52/0.65	25.42/0.73
	RIDNet†	25.45/0.74	25.63/0.74	25.35/0.73	25.58/0.74	24.59/0.65	25.34/0.73
Denoisers on 16× average	PURE-LET	24.90/0.74	25.52/0.75	25.40/0.75	25.49/0.75	24.34/0.67	24.63/0.74
	VST+EPLL	25.64/0.75	25.96/0.75	25.66/0.74	25.76/0.75	24.50/0.66	25.32/0.74
	VST+BM3D	25.72/0.75	25.96/0.75	25.67/0.74	25.83/0.75	24.56/0.65	25.36/0.74
	DnCNN†	25.64/0.75	25.87/0.75	25.62/0.75	25.85/0.75	24.72/0.66	25.60/0.74
	MemNet†	25.77/0.75	25.90/0.75	25.67/0.74	25.87/0.75	24.75/0.66	25.60/0.74
	RIDNet†	25.67/0.75	25.85/0.75	25.52/0.74	25.84/0.74	24.63/0.65	25.54/0.74

Table 1. PSNR (dB)/SSIM results on the sequential application of denoising and SR methods on the W2S test images for different noise levels. †The learning-based denoising methods are retrained for each noise level, and the SR networks are trained to map the noise-free LR images to the high-quality HR images. For each SR method, we highlight the best PSNR and SSIM value in red. For each denoising method, we underline the best PSNR and SSIM value.

in their papers. For fair comparison, we use the same training setup for all models. For each training batch, 16 LR patches of size 64×64 are extracted. All models are trained using the Adam optimizer [4] for 50 epochs. The learning rate decreases by half every 10 epochs. Data augmentation is performed on the training images with a possibility of 0.5, which are randomly rotated by $90°$, flipped horizontally, flipped vertically.

4 Additional Benchmark Results

We present additional benchmark results of the sequential application of state-of-the-art denoisers and SR methods on low-resolution (LR) images with different noise levels on W2S. The different noise levels correspond to a different number of averaged raw images.

We note that there is no consistent and significantly-better denoiser across all SR methods and noise levels. Between SR models, the best is RDN but not with a large margin. We only observe a consistent and significant improvement when we train networks end-to-end on our JDSR data, and further when we leverage the novel loss presented in the main paper (Table 4 of the paper).

References

1. Gustafsson, M.G.: Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. Journal of microscopy (2000)
2. Gustafsson, M.G.: Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution. Proceedings of the National Academy of Sciences (2005)
3. Gustafsson, M.G., Shao, L., Carlton, P.M., Wang, C.R., Golubovskaya, I.N., Cande, W.Z., Agard, D.A., Sedat, J.W.: Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. Biophysical journal (2008)
4. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint (2014)