Supplementary Experiments On Testing Sub-Images Cropped From *Cuprite* Site Of AVIRIS Data

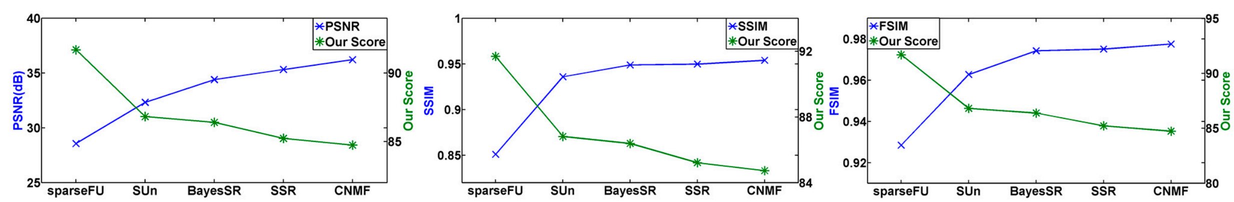
Other than the testing images in the manuscript, we cropped two sub-images from *Cuprite* of AVIRIS data for testing (denoted as *Cuprite*-1 and *Cuprite*-2). The indices are given in Table S1 and S2. In order to show the trend of different indices clearly, we also plot the different indices as curves in Figure S1 and S2. The reconstructed HSIs are shown in Figure S3 and S4.

Table S1. Comparison among PSNR, SSIM, FSIM, and our score on *Cuprite*-1 of AVIRIS data

	sparseFU	**SUn**	**BayesSR**	**SSR**	**CNMF**
PSNR	28.5608dB	32.3148dB	34.3908dB	35.3152dB	36.2089dB
SSIM	0.8511	0.9360	0.9490	0.9499	0.9541
FSIM	0.9285	0.9628	0.9743	0.9751	0.9776
Our score	91.6915	86.8168	86.3921	85.2228	84.7357

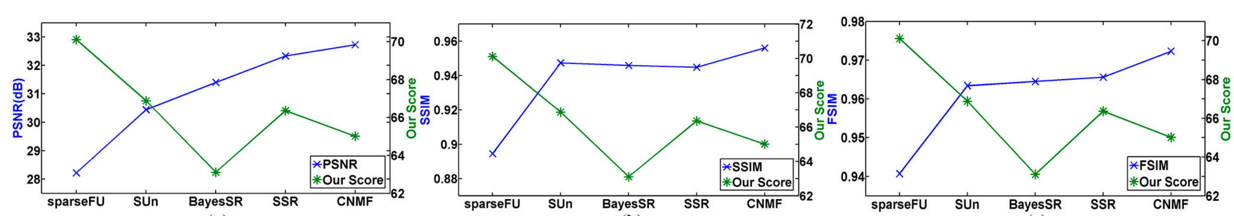
Table S2. Comparison among PSNR, SSIM, FSIM, and our score on *Cuprite*-2 of AVIRIS data

	sparseFU	**SUn**	**BayesSR**	**SSR**	**CNMF**
PSNR	28.2177dB	30.4406dB	31.4005dB	32.3294dB	32.7278dB
SSIM	0.8945	0.9474	0.9459	0.9448	0.9561
FSIM	0.9407	0.9634	0.9645	0.9656	0.9723
Our score	70.1129	66.8792	63.1085	66.3626	65.0143



(**a**) (**b**) (**c**)

Figure S1. Consistency of our score and reference-based indices on *Cuprite*-1 of AVIRIS data. (**a**) our score and PSNR; (**b**) our score and SSIM; (**c**) our score and FSIM.



(**a**) (**b**) (**c**)

Figure S2. Consistency of our score and reference-based indices on *Cuprite*-2 of AVIRIS data. (**a**) our score and PSNR; (**b**) our score and SSIM; (**c**) our score and FSIM.

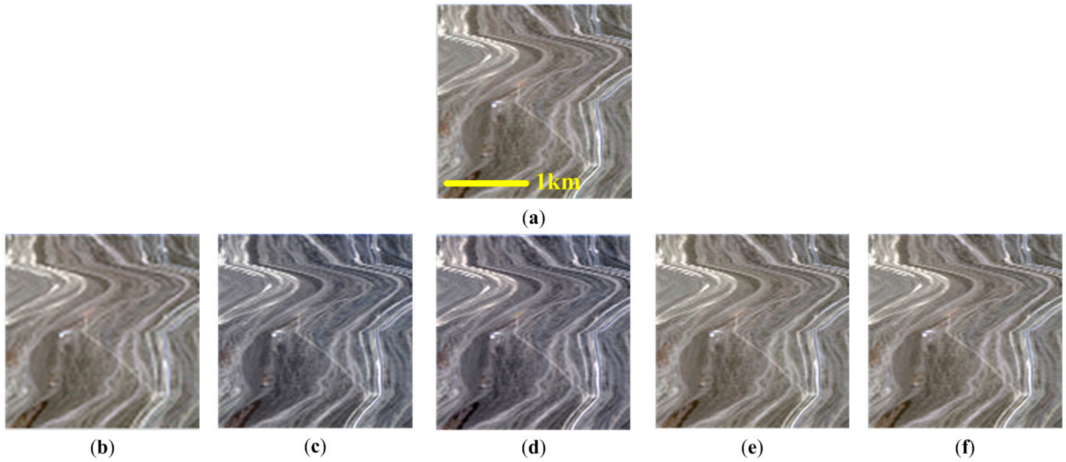


Figure S3. Reconstructed HSI of different super-resolution methods, the images are shown in RGB (band 35, 25, 15). The sub-image with size 128×128×162 is cropped from *Cuprite*-1 of AVIRIS data. (**a**) Original sub-image, (**b**) result of sparseFU, (**c**) result of SUn, (**d**) result of BayesSR, (**e**) result of SSR, (**f**) result of CNMF.

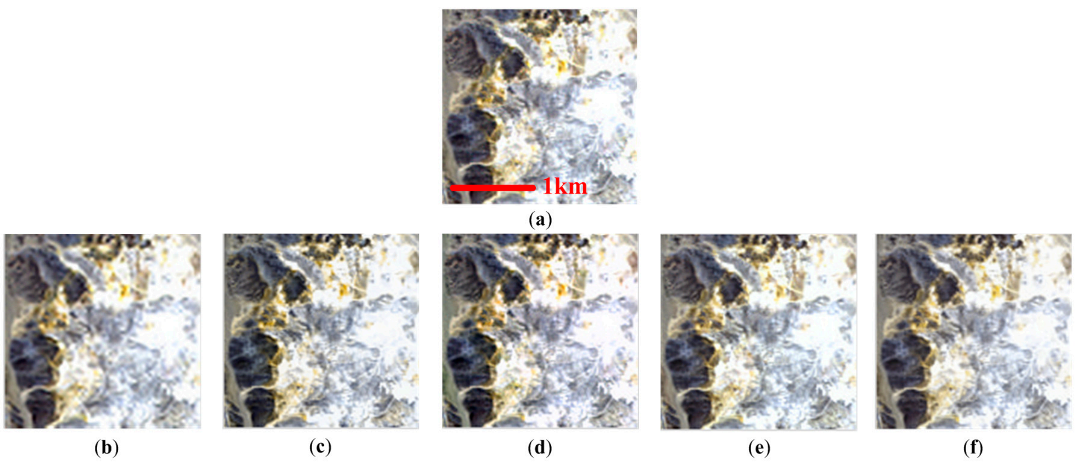


Figure S4. Reconstructed HSI of different super-resolution methods, the images are shown in RGB (band 35, 25, 15). The sub-image with size 128×128×162 is cropped from *Cuprite*-2 of AVIRIS data. (**a**) Original sub-image, (**b**) result of sparseFU, (**c**) result of SUn, (**d**) result of BayesSR, (**e**) result of SSR, (**f**) result of CNMF.

As shown in the tables and figures, except of BayesSR on *Cuprite*-2, the scores of our proposed method are generally consistent with PSNR, SSIM, and FSIM in assessing the reconstructed HSIs. Compared with *Cuprite*-1, the less textural information on *Cuprite*-2 may lead to the failure of our method in assessing BayesSR.

We also assess the spectral distortion by extract the spectral features only, the spectral scores are reported in Table S3 and S4, the indices are also plotted as curves in Figure S5. It is clear that our method can assess the spectral distortion correctly on *Cuprite*-1, all the spectral scores are consistent with SAM. On *Cuprite*-2, the spectral scores of BayesSR and SUn are not consistent with SAM, the spectral scores of SSR and sparseFU are also not consistent.

Table S3. Comparison between SAM and spectral quality score on *Cuprite*-1 of AVIRIS data

	SUn	**BayesSR**	**SSR**	**sparseFU**	**CNMF**
SAM	3.2842°	2.9892°	2.1189°	1.9906°	1.8291°
Spectral score	1.8234	1.8057	1.6508	1.5462	1.2446

Table S4. Comparison between SAM and spectral quality score on *Cuprite*-2 of AVIRIS data

	BayesSR	**SUn**	**SSR**	**sparseFU**	**CNMF**
SAM	2.9789°	2.4808°	2.2109°	1.9010°	1.7015°
Spectral score	1.7158	1.7567	1.5832	1.6175	1.0342

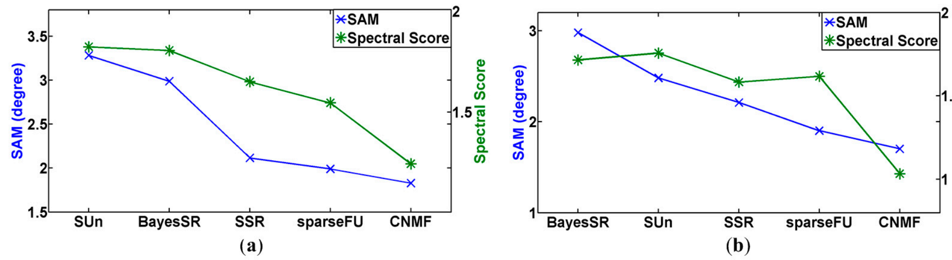


Figure S5. Comparison between SAM and spectral score, (**a**) on *Cuprite*-1; (**b**) on *Cuprite*-2.

It is noted that in the experiment, we crop the sub-images from AVIRIS and HyperspecVC dataset for testing, the rest of each dataset is treated as pristine data and used as training data. Large number of 3D blocks can be extracted from the training data. We only select part of these blocks for training according to the contrast of each block. Standard variance of each training block is computed, a block would be selected for training if its standard variance is higher than a pre-defined threshold. The threshold is empirically set as 50%~70% of the largest standard variance. According to our experience, if the threshold is lower than 50% of the largest standard variance, the consistency may not be guaranteed. If the threshold is higher than 70% of the largest standard variance, the number of selected blocks may be too scarce to learn the benchmark MVG.