Reviewer 1: - Unfair Experiments: There has been a misunderstanding! *We use only one measurement (one partition with no overlap for CS) at test time*. The two measurements (two overlapping partitions for CS) are used *only for training*. The underlying network is setup to estimate the image from a single measurement. During training, we apply the network separately to the two measurements to get two estimates of each training image, and use these estimates jointly to compute our losses. But for evaluation on the test set, our network is given only one measurement as input. Thus, *all methods are provided identical inputs (one measurement, no overlap)* that matches the noted ratios for CS.

- Other Recent Methods: We will cite & discuss these concurrent works (note they were only on arXiv until recently: the BASP workshop Feb'2019 for Metzler, and CVPR Jun'2019 for Zhussip). Both papers propose approaches based on SURE (i.e., very different from ours) for unsupervised training from single measurements. But they are specific to D-AMP estimation: they train denoiser networks for use in unrolled AMP iterations for CS recovery. Our method needs pairs of measurements, but is a more general framework that can be used to train generic neural network estimators.

Comparisons: We can't directly compare to Zhussip since there is no code available yet. Instead, we'll add comparisons to BM3D-AMP which was their main baseline. Moreover, BM3D-AMP was also the best performing CS method in Metzler's evaluation (see their Tbl.2: their proposed methods were faster but had lower avg PSNRs than BM3D-AMP).

Below, we show results of BM3D-AMP on Set11 with a "patch-wise" evaluation from the ISTANet paper, as well as a "full-image" evaluation suggested by R1 (calling BM3D-AMP on the full image with measurements from all patches as input). We see that, even with full image restoration, BM3D-AMP performs worse than our unsupervised method by 3.26 dB at 10% (gaps are even wider at lower rates, as also noted in the ISTANet paper). As reference, BM3D-AMP is better than Metzler, and worse than Zhussip by only 1-1.5 dB, in their own CS evaluations.

Method	1%	4%	10%
BM3D-AMP Patch-wise (from ISTANet Paper)	5.21 dB	18.40 dB	22.64 dB
BM3D-AMP Full Image (our evaluation)	5.59 dB	17.18 dB	23.07 dB
Our Method (unsupervised)	17.84 dB	22.20 dB	26.33 dB

- Only Patch-based Restoration: Nothing in our framework restricts it to patch-based networks: the entire second half of our evaluation (on deblurring) is on a full-image restoration network. For CS, we chose patch-based restoration only to follow the protocol of the recent CVPR'18 ISTANet paper (also, patch-wise is just a experimental choice: ISTANet and our CS network could be easily adapted to treating the whole image as a large "patch"). For completeness, we will also add the full-image BM3D-AMP results above to our evaluation.

Reviewer 2: Thanks for your positive comments! Our main results are really the comparisons between the supervised and unsupervised versions of the same (our) network. In terms of architecture, we actually use concatenated U-Nets only for CS, inspired by ISTANet+ that also concatenated two networks. We had also tried a single U-Net for CS and got similar gaps between unsupervised and supervised (both did slightly worse than concatenated U-Nets). For deblurring, we actually have a single U-Net (with a second decoder path for kernel estimation used only in blind training). Also, rather than claim a contribution on a specific application (where the improvements from the architecture are relatively minor: 0.1-0.2dB for 10% CS and deblurring), we wanted to keep the focus on our general unsupervised framework.

Reviewer 3: Thanks for the detailed feedback! **- Q matrix / swap-loss:** We'll further discuss the full-rank requirement of Q to give readers intuition (individual θ can be low-rank, but must be diverse in a way that Q is full-rank). With motion-blur, the decay is along different orientations for different kernels. And as they are thin splines, most of the 'low-rankness' comes from a sinc-like pattern of zeros in frequency. Orientations and zero-frequencies vary across kernels, making Q full-rank. We'll add images of individual kernel and average magnitude spectra to show this.

- Self-loss: The self-loss provides extra supervision per sample (although the swap loss is full-rank in expectation, it is low-rank per-sample). Note we use it in conjunction with the swap-loss, and any noise present will be independent in the two measurements. So for the identity case, the swap- and self-loss together will promote convergence to the mean of the two noisy measurements. We'll add an ablation to Table 1 w/o self-loss: performance drops by 0.04-0.17 dB.

- prox:θ Loss: Blind-training can indeed only be used when estimating θ is feasible. We'll clarify this assumption in the revision. For deblurring, kernel estimation is roughly as hard as deblurring (early deblurring methods first estimated the kernel & ran non-blind deconvolution), and is often reasonably successful (with supervised training).

- Experiments: *Face:* Class-specific deblurring (for faces, text, etc.) is a popular problem because it can perform better on that class compared to generic deblurring, by relying more on class-specific image priors (with supervised training). Since the CS experiments were already on general images, we felt this was an interesting second evaluation for our unsupervised method. *Noise:* The gap between supervised and unsupervised indeed grows with more noise, as supervised benefits from always training on noise-free GT. The deblurring noise level was set by the benchmark. *L1 Loss:* We used L1 because it is a common choice in deblurring networks (gives sharper results). (4) doesn't hold as is, but expected loss is still only minimized by an ideal estimate. *CS Measure:* We just use two randomly shifted partitions: each dividing the image into non-overlapping patches (chosen as it could be realized practically by moving the sensor).