Supplemental Document

Optics EXPRESS

Scan-less machine-learning-enabled incoherent microscopy for minimally-invasive deep-brain imaging: supplement

RUIPENG GUO,[1] SOREN NELSON,[1] MATTHEW REGIER,[2] M. WAYNE DAVIS,[3] ERIK M. JORGENSEN,[3] JASON SHEPHERD,[2] AND RAJESH MENON[1,*]

[1]*Department of Electrical & Computer Engineering, University of Utah, UT 84112, USA*
[2]*Department of Neurobiology, University of Utah, UT 84112, USA*
[3]*School of Biological Sciences and Howard Hughes Medical Institute, University of Utah, UT 84112, USA*
[*]*rmenon@eng.utah.edu*

Supplement DOI: https://doi.org/10.6084/m9.figshare.17700779

Parent Article DOI: https://doi.org/10.1364/OE.446241

Scan-less machine-learning-enabled incoherent microscopy for minimally-invasive deep-brain imaging : supplemental document

RUIPENG GUO,[1] SOREN NELSON,[1] MATTHEW REGIER,[2] WAYNE DAVIS,[3] ERIK JORGENSEN,[3] JASON SHEPHERD,[2] RAJESH MENON[1,*]

[1] *Department of Electrical & Computer Engineering, University of Utah.*
[2] *Department of Neurobiology, University of Utah.*
[3] *School of Biological Sciences and Howard Hughes Medical Institute, University of Utah.*
**rmenon@eng.utah.edu*

1. Optical Setup

The details of CCM system have been shown in main text. Under stage for sample, we set a reference system, which is a conventional widefield fluorescent microscope, as shown in Fig. S1. Same objective (PLN 20X, Olympus) is utilized to image sample onto a CCD camera (AmScope MU130). During calibration, this reference system can be used to record images of sample as ground truth for training synchronously. Once the dataset is built, the reference system will be removed and brain sample is set under cannula. For thin sample, such as brain slice or phantom, we can also set reference system under it to track the image close to the distal end of cannula, which is an effective method to verify the correctness of reconstructed images.

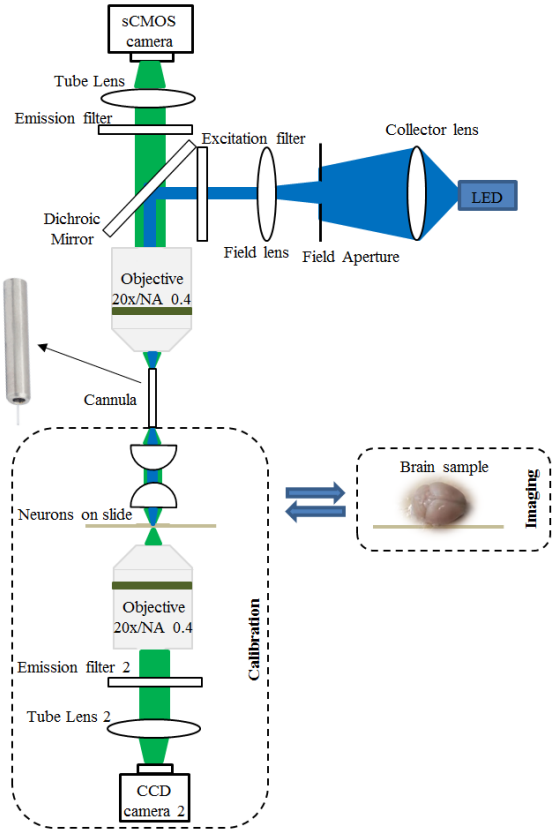


Fig. S1. Sketch of system. A reference system which is utilized for calibration is set under sample. It could be removed once calibration is done

2. U-net artificial neural network

This network is a typical U-net with the skip connections to concatenate the encoder and decoder outputs[1,2], as shown in Fig. S2. Dense blocks, which include 2 convolutional layers with RELU activation function and a batch-normalization layer are used to build each encoder and decoder. Sigmoid function is utilized as the last layer. We set the sizes of input and output as 128*128 to fit the performance of computer.

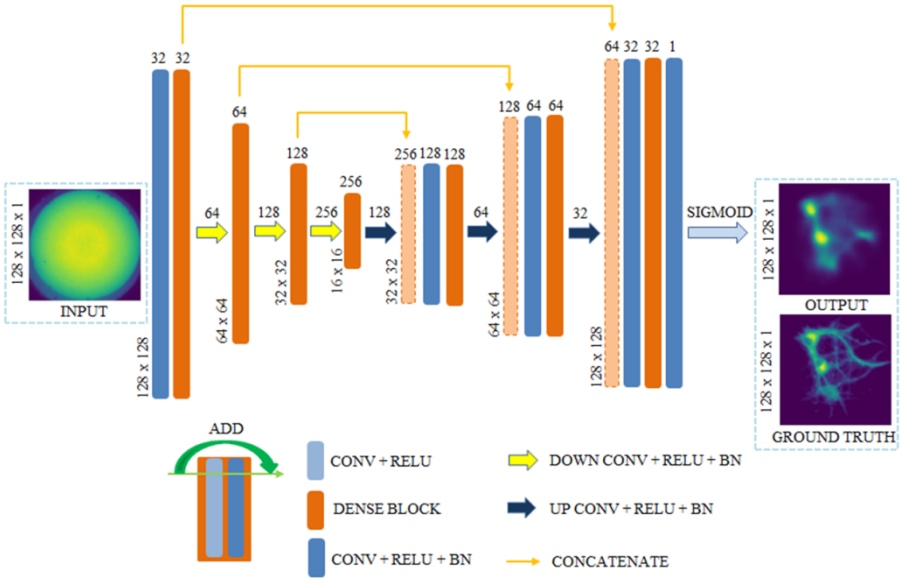


Fig. S2. Sketch of U-net ANN. The architecture of ANN includes 7 dense block, which consists of 2 convolutional layers with RELU activation function and a batch-normalization layer. The input of each block is added to the output of block to prevent loss of information. Camera image is downsampled to 128*128 and used as input of ANN. Ground truth from reference system is used as output.

3. Materials and Methods

Preparation of cultured neurons: Primary neurons were taken from dissociated hippocampi of E18.5 Sprague-Dawley rat pups. Hippocampi were dissociated using 0.01% DNase (Sigma-Aldrich) and 0.067% papain (Worthington Biochemicals) prior to trituration through glass pipettes to obtain a single-cell suspension. Cells were then plated at 8 × 104 cells/ml in the Neurobasal medium (Thermo-Fisher) supplemented with 5% horse serum, 2% GlutaMax (Thermo Fisher), 2% B-27 (Thermo Fisher), and 1% penicillin/streptomycin (Thermo Fisher) on coverslips (No. 1, Bioscience Tools) coated overnight with 0.2 mg/ml poly-L-lysine (Sigma-Aldrich) in 100 mM Tris-base (pH 8). Neurons were grown at 37◦C/5% CO and fed via a half-media exchange every third-day with astrocyte-conditioned Neurobasal media supplemented with 1% horse serum, 1% GlutaMax, 2% B-27, and 1% penicillin/streptomycin, with the first feeding containing 5 µM β-D-arabinofuranoside (Sigma-Aldrich) to limit the overgrowth of the glial cells. The neurons were grown for 12–14 days *in vitro* prior to transfection, fixation, and imaging.

The neurons were transfected after 12 days *in vitro* with 0.5 µg of pCAG-eGFP (Addgene: 89684) using lipofectamine 2000 at a 3:1 ratio when complexed with plasmid DNA. Then neurons were transfected over the course of 1 h at 37◦C in pH 7.4 Minimum Essential Media

(Thermo Fisher) supplemented with 2% GlutaMax, 2% B-27, 15 mM HEPES (Thermo Fisher), 1 mM Sodium Pyruvate (Thermo Fisher), and 33 mM Glucose. After the transfection, the neurons were given 24 hrs in growth media at 37◦C/5% CO2 to allow a sufficient recovery and expression of the plasmid prior to fixation in 4% formaldehyde (thermo fisher)/4% sucrose (VWR) in phosphate buffered saline for 15 min at room temperature. After fixation, the neurons were mounted in a Prolong Gold Aqueous Mounting Medium (Thermo Fisher) and imaged.

Preparation of brain samples from mice: C57/Bl6J mice carrying the Thy1-eGFP gene were used (Jackson Labs stock #007788). Mice were housed on a 12h light:12h dark cycle, with water and food available *ad libitum*. For preparation of brains for imaging, mice were exposed to 5% isoflurane in an anesthesia chamber. After deep anesthetization, mice were transcardially perfused with ice cold PBS followed by 4% paraformaldehyde in PBS. Mouse brains were then removed and placed in PBS for immediate imaging. Animal experiments were approved by the Institutional Animal Care and Use Committee at the University of Utah.

4. Frame rate evaluation

We also tried to image brain sample with various exposure time to evaluate the frame rate of our system. When the cannula was inserted into brain sample, we fixed the cannula and recorded images continuously with varied of exposure time. As shown in Fig. S3(a), when exposure time is decreased to 0.025s, the images could still be reconstructed well, which means our system could achieve as close as 40Hz frame rate. Besides, we also imaged cultured neurons with varied of exposure time, which is shown in Fig. S3(b). Besides exposure time, we also need to consider prediction time. With the dataset, which includes 84538 images, it takes 4.3 hours for complete training that includes 17 epochs. When the training is done, the average computation times for predicting each image is 3.5 ms.

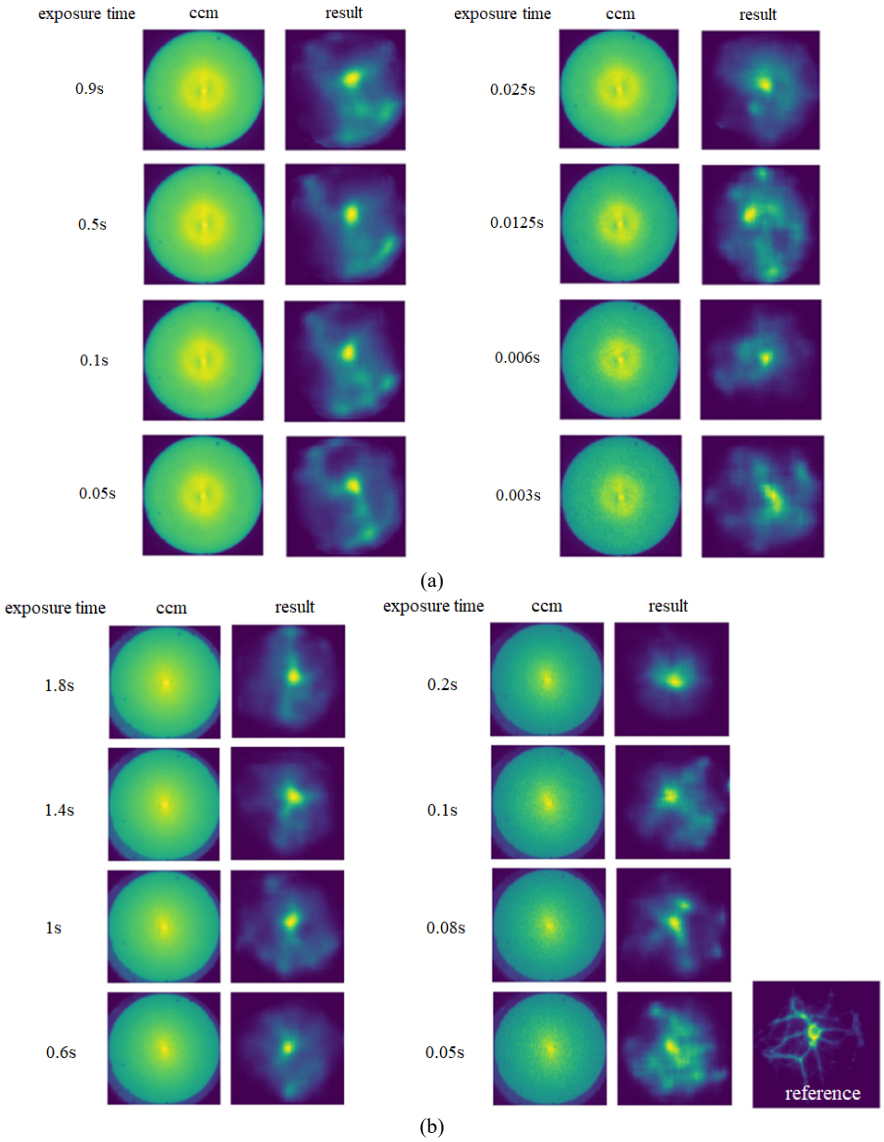


Fig. S3. Imaging with varied of exposure time. (a) Brain sample imaging with different exposure time. The depth of cannula into the brain is approximately 1mm and fixed. Then images were recorded as exposure time decreasing. The reconstructed image is still good when exposure time decrease to 0.025s. (b) Cultured neurons on slide imaging with different exposure time. Images were recorded with different exposure time at same position. Image on right is reference image recorded by reference microscope. Since signal from this cultured neurons on slide is weaker than signal in brain sample, it needs higher exposure time.

5. Supplementary Phantom results

We imaged phantom made with fluorescent beads and reconstructed images via ANN trained with dataset built from beads sample on slide. Phantom was made as following steps. (1) Measure 0.2g of agarose powder (EZ BioResearch S-1020-250) and add it with 10mL of high purity water to a flask. Stir the flask until agarose disperses uniformly in the solution. (2) Place the flask in microwave oven and heat using 100% power for 10 seconds. Repeat 10 seconds interval heating until agarose completely dissolves. Gently stir between intervals to

suspend agarose. (3) Before agarose solidify, add 50uL of 1:500 diluted bead (2% solid, 4μm green FluoSpheres sulfate microsphere) mixture into the agarose. Mix the solution well using vortex. (4) Pour agarose solution into a mold on a glass slide and wait few minutes till agarose solidify. Then glue it on glass slide[3].

The dataset used for training ANN includes 4 depths. Each depth consists of approximately 20000 images. Once the ANN was trained, we put phantom under cannula and inserted it into phantom, then recorded images as cannula inserting. Fig. S4 shows reconstructed images at three different depths. Because the thickness of phantom is 1mm, we can focus the reference objective on close to the distal end of cannula to track the beads distribution in the field of view of our system. The beads can be successfully reconstructed, even though beads in same images are on different depths from cannula.

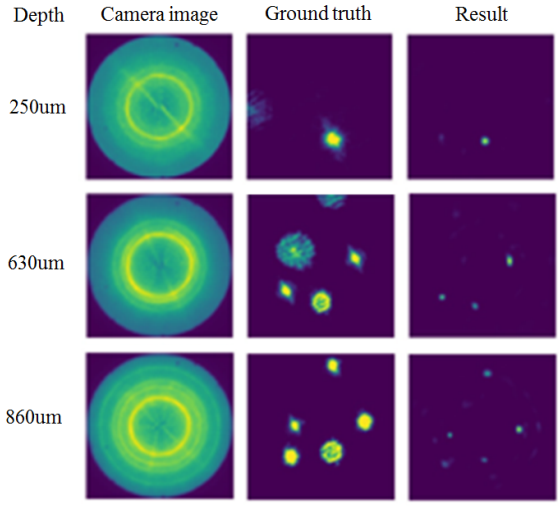


Fig. S4. Phantom results. Ground truth was recorded by reference system. The objective was focused close to the distal end of cannula. The beads these are out of focus are on different depths to the cannula. To find out the beads which are out of focus, we set the exposure time high, that results in over-exposure for some beads.

6. Additional cultured neurons images

Additional images taken from cultured neurons are presented in Fig. S5.

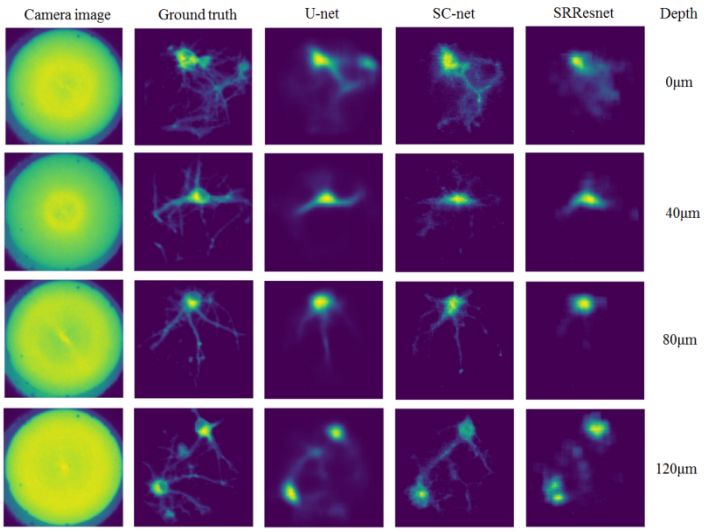


Fig. S5. Additional results from cultured neurons.

7. Additional brain images

Additional images taken from the whole brain specimen are shown in Fig. S6. Each figure shows selected images recorded from the same location of brain sample with various depth.

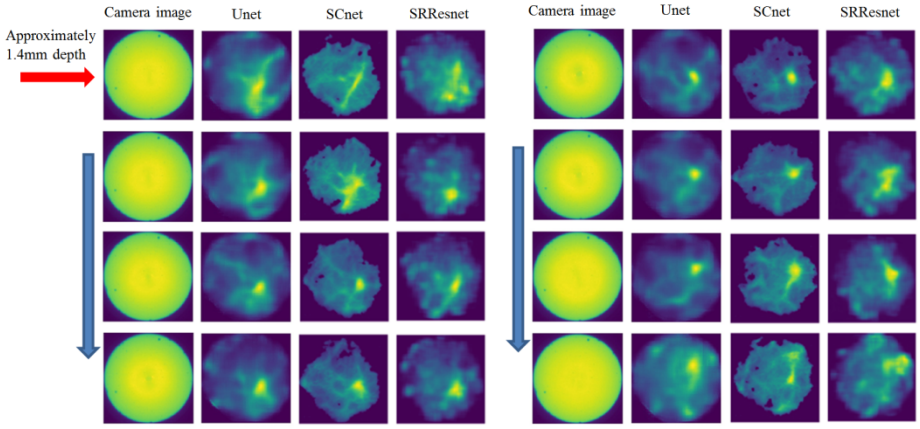
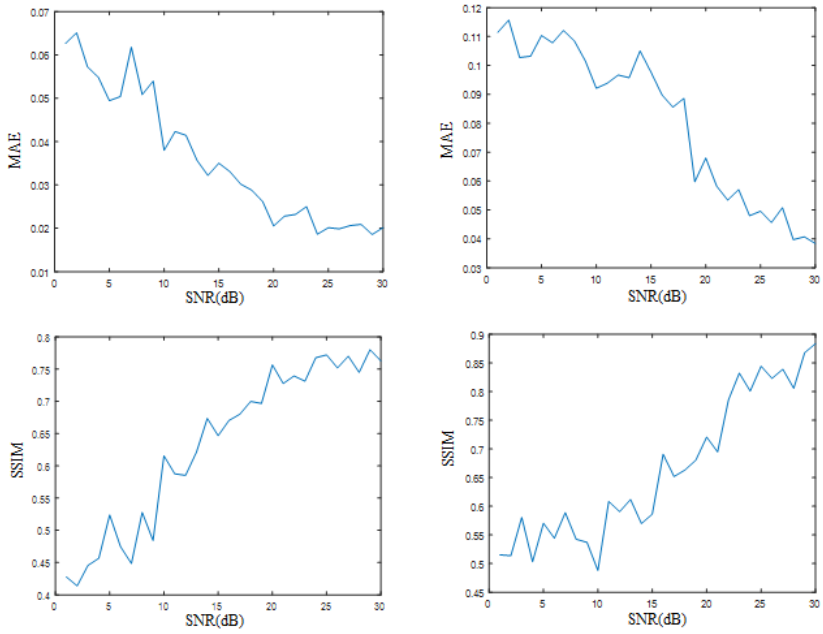


Fig. S6. Additional brain images.

8. Influence of Noise

The influence of noise to the quality of predicted images is tested by adding noise in ccm images with Matlab. We set 30 SNRs from 1dB to 30dB with step 1 dB. For images of cultured neurons, ground truth recorded by reference camera is utilized as reference to calculate SSIM and MAE. For brain images, we don't have ground truth, so we use the predicted image from ccm image without noise as reference. From the curve, we can get similar SSIM values when the SNR are bigger than 20 dB, as shown in Fig. S7.



(a) (b)

Fig. S7. SSIM and MAE vs SNR.(a) Plotted curves from cultured neurons. (b) Plotted curves from Brain images.

9. Discussion on Cycle GAN and ensemble averaging:

Zhu et al. designed the CycleGAN to map unpaired images from one domain to another. We instead apply the Cycle Consistency loss to further constrain our model when reconstructing paired cannula-reference images. However, we found that applying the CycleGAN alone on paired images consistently led to mode collapse. We enable the CycleGAN to work for paired images by applying the L1 reconstruction loss to both models alongside the cycle consistency and adversarial losses. This combination outperformed both training with the L1 loss alone and with the L1 and adversarial losses.

We trained the SC-net for 40 epochs with $\lambda = 100$ and batch size 8. The SRResNet was trained with a batch size of 16 over 45 epochs. We have included further explanation of the novelty over the CycleGAN and the training hyperparameters used in the manuscript as well as made our code public and cited it for reproduction of these results.

We note that deep learning methods, in particular GANs, although extremely performant can often predict images with artifacts, and this is present in some of our reconstructions where there are dendrites which are not present in the ground truth. We employ ensemble methods to mitigate this issue which have long been used as a variance reduction technique. Neural networks ensembles have shown to be robust to dataset shift [4] (our transfer task) for which diversity in predictions is a critical component [5] and have also shown to reduce artifacts from GANs [6]. We use a combination of diverse architectures and training schemes to arrive at diverse predictions, some of which are more conservative and predictable like the U-net, and some which are produce higher quality images at the cost of potential artifacts like the SC-net. To ensemble the networks we perform a simple pixel-wise average of the predictions from each model. While this does not remove all artifacts it does reduce them and at a minimum, reduce overconfident mispredictions. We note that there are

potentially other methods to handle this such as giving a confidence estimation for the ensemble prediction [7] however we feel this is out of the scope of this paper.

10. Imaging worms

In order to clarify the location of the worm, we have shown an illustration below of an estimate of the worm. Note that our FOV is only 0.22mm X 0.22mm so it's very likely that we are only seeing a portion of a worm.

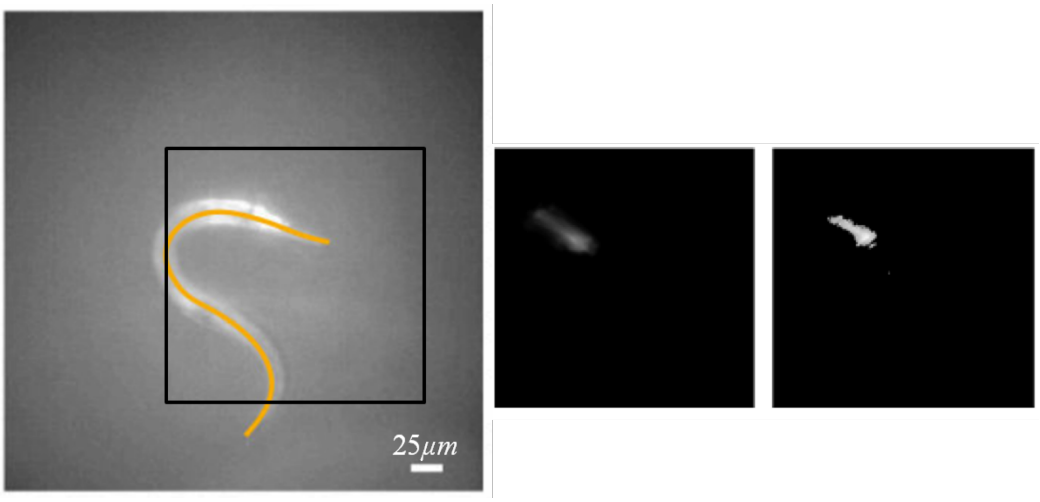


Fig. S8. Left panel is taken from ref [PNAS 113 (8) E1074 (2016)] and right panels are from Fig. 5 of our manuscript. This is just an estimate as we don't have exact brightfield images of the worm for this frame.

11. Movie S1 (separate file).

Video as cannula inserting into brain sample**.** A video is generated with reconstructed images from brain sample, which is shown in attached file 3D.mp4.

12. Movie S2 (separate file).

Video of worms is generated with linear algorithm**.** The frame rate of video is 2 frames per second. It is shown in attached file worms.mp4. Top-left: raw cannula video. Top-right: Ground truth. Bottom-right: reconstructed. Bottom-left: Centroid tracking from reconstructed video.

References

1. R. Guo, *et al.*, *Opt. Lett.* 45 (7) 2111-2114 (2020).
2. R. Guo, *et al.*, *Opt. Exp.* 28(22) 32342-32348 (2020).
3. G. Kim, N. Nagarajan, E. Pastuzyn, K. Jenks, M. Capecchi, J. Shepherd, and R. Menon, Sci. Rep. **7**. 44791 (2017).
4. Y. Ovadia, E. Fertig, J. Ren, Z. Nado, D. Sculley, S. Nowozin, J.V. Dillon, B. Lakshminarayanan, and J. Snoek, "Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift," arXiv:1906.02530 (2019).
5. S. Fort, H. Hu, and B. Lakshminarayanan, "Deep ensembles: A loss landscape perspective," arXiv:1912.02757 (2019).
6. Q. Lyu, H. Shan, and G. Wang, "MRI super-resolution with ensemble learning and complementary priors," in IEEE Transactions on Computational Imaging (2020), pp. 615-624.

7. A. Filos, S. Farquhar, A.N. Gomez, T.G. Rudner, Z. Kenton, L. Smith, M. Alizadeh, A. De Kroon, and Y. Gal, “A systematic comparison of bayesian deep learning robustness in diabetic retinopathy tasks,” arXiv:1912.10481 (2019).