Supplemental Document

Optics Letters

Learning-based adaptive under-sampling for Fourier single-pixel imaging: supplement

WENXIN HUANG,[1,2,†] FEI WANG,[1,†,*] XIANGYU ZHANG,[1,2] YING JIN,[1] AND GUOHAI SITU[1,2,3,4]

[1] Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China
[2] Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
[3] Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China
[4] ghsitu@siom.ac.cn
[†] These authors contributed equally to this Letter.
[*] wangfei@siom.ac.cn

Supplement DOI: https://doi.org/10.6084/m9.figshare.22684036

Parent Article DOI: https://doi.org/10.1364/OL.486416

Learning-based adaptive under-sampling for Fourier single-pixel imaging: Supplemental Document

This document provides supplementary information to *Learning-based adaptive under-sampling for Fourier single-pixel imaging*. We provide the details of the designing strategy of under-sampling masks, deep neural network (DNN) architecture, and some additional simulation and experimental results.

1. PREPARING OF UNDER-SAMPLING MASKS

A. circle under-sampling mask

Here we explain how we obtain the manually selected circle under-sampling mask used for FSI-circle and FSI-DL. When calculating the spectrum of a color image, we actually divide the process into the following steps shown in Figure S1:

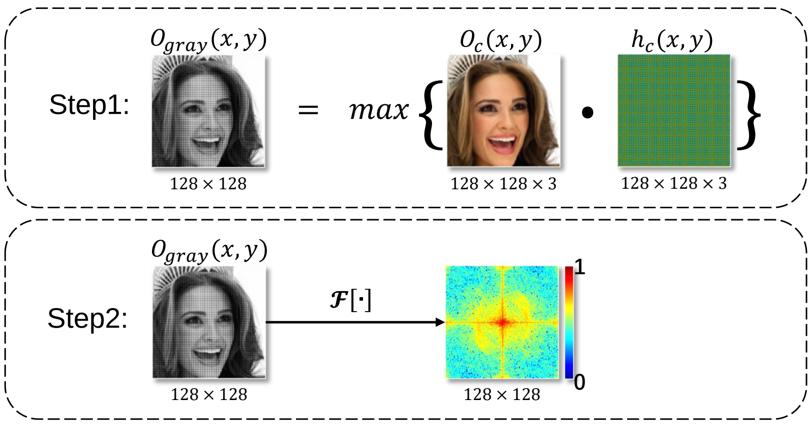

Fig. S1. Spectrum calculation of a color image. Step.1: Modulate the color image $O_c(x,y)$ with a CFA $h_c(x,y)$, then we get $O_{gray}(x,y)$ which is a bayer-grayscale image. Step.2: Calculate the spectrum of $O_{gray}(x,y)$.

As plotted in Figure S1, the energy of the Fourier spectrum of a Bayer grayscale image is mainly concentrated in the center of the image, but there are also notable concentrations of energy in the four corners and midpoints of the spectrum image. The reason for this phenomenon is that we used a CFA $h_c(x,y)$ in Step 1, which is composed of red, green, and blue channels together, and can be expressed in the form of the following matrix:

$$h_R = \begin{bmatrix} 0 & 1 & \dots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & & \vdots & \\ 0 & 1 & \dots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, h_G = \begin{bmatrix} 1 & 0 & \dots & 1 & 0 \\ 0 & 1 & \cdots & 0 & 1 \\ \vdots & & \ddots & & \vdots \\ 1 & 0 & \dots & 1 & 0 \\ 0 & 1 & \cdots & 0 & 1 \end{bmatrix}, h_B = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 1 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \cdots & 1 & 0 \end{bmatrix}, \tag{S1}$$

As shown in Figure S2, each single-spectral filter is constructed in the Fourier domain using three pulse functions that are distributed at the center, as well as at the corners and midpoints of the spectrum image. Therefore, when we use h_c to modulate a color image, the color information of different channels of O_c is also modulated to corner and midpoint regions by these pulse functions. According to the conjugate symmetry of the spectrum (see Figure S3 (a)), the energy of a complete Fourier spectrum of a Bayer grayscale image is concentrated in the center, four corners, and midpoints (see Figure S3 (b)).

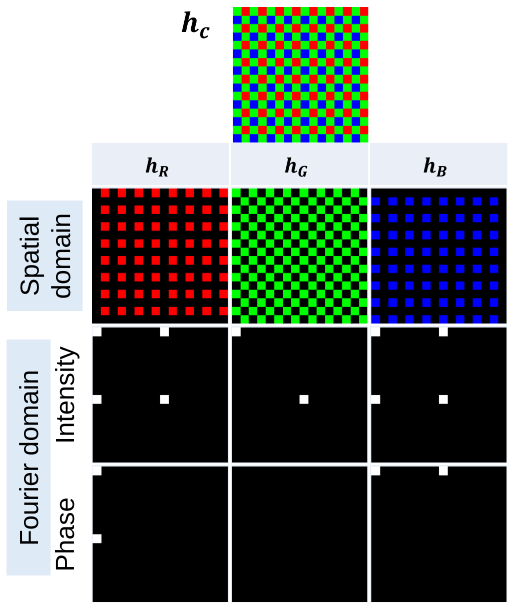

Fig. S2. Spectrum of each single-spectrum filter

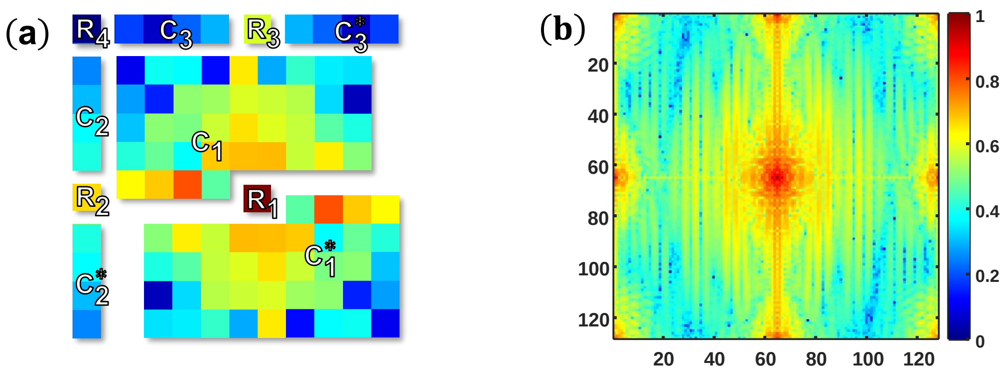

Fig. S3. (a) Example of Conjugate Symmetry Schematic Diagram of Spectrum using 16×16 Pixel Size. For a grayscale image with an even number of rows and columns, the spectrum exhibits conjugate symmetry, and the C_i and C_i^* regions exhibit a symmetry around the point R_i $(i = 1, 2, 3)$. (b) The mean Fourier intensity spectrum of 28,000 Bayer-CFA-sampled images with 128×128 Pixel Size.

This inspires us to design the circle under-sampling mask M_{circle} shown in Figure S4(a) to impose importance sampling on the Fourier spectrum of a Bayer grayscale image. The sampling ratio determined by M_{circle} can be simply defined as

$$\beta \approx \frac{1.5 \times 4\pi r^2}{m \times n} \tag{S2}$$

Here, r represents the radius of the circles. Therefore, we can design circular mask M_{circle} with different sampling ratios using circles with different radius r.

B. learned under-sampling mask

Here we describe details about how we obtain the learned under-sampling mask. Specifically, we use a Monte Carlo-based sample averaging strategy for optimization. We also exploit some

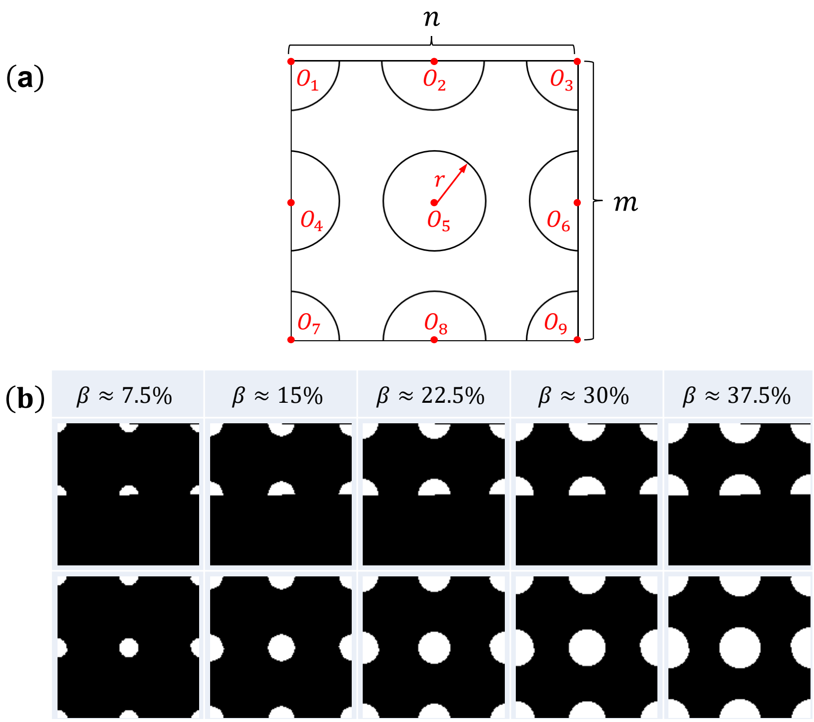

Fig. S4. Illustration of mask selection method in FSI-circle and FSI-DL. (a) A schematic of the circular mask with radius r. (b) The circle masks uesd in FSI-circle and FSI-DL: The second row is the result of the conjugate symmetry of the first row.

constraints to meet the binary and symmetry requirements. Finally, we apply post-processing techniques to ensure that the desired sampling ratio is precisely achieved. More details can be found in the released code: https://github.com/FeiWang0824/AuSamNet.

B.1. Monte Carlo-based sample averaging strategy

Our AuSamNet realizes the co-design of under-sampling mask M and the reconstruction DNN model $\mathcal{R}_\omega$ by solving

$$
\begin{aligned}
M^*, \omega^* &= \arg\min_{M,\omega} \frac{1}{K} \sum_{i=1}^{K} \|\mathcal{R}_\omega \left(\mathcal{F}^{-1} \left(M \cdot \mathcal{F} \left(O^i_{gray} \right) \right) \right) - O^i_c\|_2^2, \\
& s.t. \frac{1.5\|M\|_1}{m \times n} = \beta,
\end{aligned}
\tag{S3}
$$

The major issue in Equation S3 is how to guarantee a given sampling ratio, i.e., $\frac{1.5\|M\|_1}{m\times n} = \beta$ and a binary mask M. For this purpose, we introduce a normalization trick

$$
N_\beta(P) = \begin{cases} \frac{\beta}{\bar{p}} P & \bar{p} \geq \beta \\ 1 - \frac{1-\beta}{1-\bar{p}}(1-P) & others \end{cases}
\tag{S4}
$$

suppose $\bar{p} = \frac{1.5\|P\|_1}{m\times n}$. According to the definition of $N_\beta(P)$, for a given input P that satisfy $\bar{p} \geq \beta$, $N_\beta(P)$ is smaller than P, and vice versa. Thus, the constrained optimization problem can be turned into an unconstrained optimization problem

$$
M^*, \omega^* = \arg\min_{M,\omega} \frac{1}{K} \sum_{i=1}^{K} \|\mathcal{R}_\omega \left(\mathcal{F}^{-1}(\sigma_1 \left(N_\beta \left(\sigma_2 \left(M\right)\right) - U^{(i)} \right) \cdot \mathcal{F}(O^i_{gray})) \right) - O^i_c\|_2^2,
\tag{S5}
$$

where $\sigma_i(x) = \frac{1}{1+e^{-t_i \cdot x}}$ denotes a Sigmoid function with a trainable slope t_i, which leads to near-binary activated results, especially for a large t_i. $U^{(i)} \in \mathbb{R}^{m\times n} \sim \mathcal{U}(0,1)$.

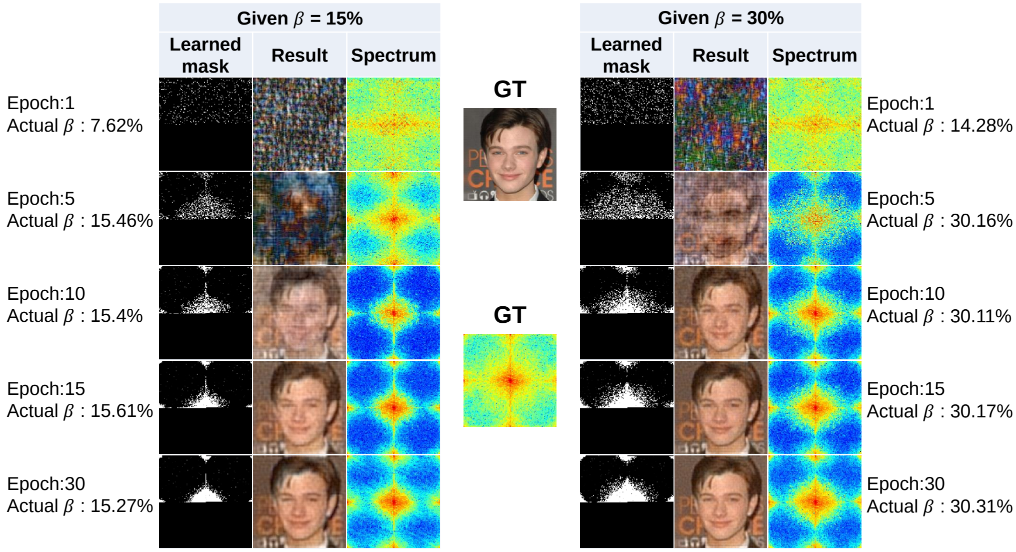

Fig. S5. Learned under-sampling mask, FSI results, and corresponding Fourier spectrum at different training epochs.

We initialize t_1 to 5 (which scales the learned mask to the (0,1) range), t_2 to 200 (which performs the thresholding operation), and M to a randomly generated matrix from a uniform distribution on (0,1) with the same size as the target grayscale image. During training, the M, t_1, t_2 in the encoder and $\mathcal{R}_\omega$ in the decoder are iteratively updated together. As shown in Figure S5, we show the learned mask, FSI results using the learned mask, and corresponding spectrum at different epochs. As the training epochs increase, it is evident that the learned mask distribution gradually becomes more concentrated, resulting in improved FSI results. This clearly demonstrates the adaptability of AuSamNet in designing optimal under-sampling masks. See more results in Visualization 1.

B.2. guarantee of symmetry

According to Equation S5, one can hopefully obtain a binary under-sampling mask that satisfies the constraint of a given sampling ratio. This is actually the method reported in Ref. [1]. However, this learning strategy does not consider the conjugate symmetry property of the Fourier spectrum. The learned under-sampling mask M^* has no guarantee of symmetry, leading to the need to collect some unnecessary information.

To address this limitation, we introduce a fixed matrix T (see Figure S6) to restrict the optimization region of the under-sampling mask to only a half, and the final under-sampling mask is obtained by symmetric calculation $Full[\cdot]$. So the optimization problem of our AuSamNet is defined as

$$M^*, \omega^* = \underset{M,\omega}{\arg\min} \frac{1}{K} \sum_{i=1}^{K} \|\mathcal{R}_\omega \left(\mathcal{F}^{-1}(Full[T \cdot \sigma_1 \left(N_\beta \left(\sigma_2 \left(M \right) \right) - U^{(i)} \right)] \cdot \mathcal{F}(O^i_{gray})) \right) - O^i_c\|_2^2. \tag{S6}$$

After optimization, the learned half under-sampling mask is

$$M_{net} = T \cdot \sigma_1 \left(N_\beta \left(\sigma_2 \left(M^* \right) \right) \right). \tag{S7}$$

B.3. Pruning

In our experiment, we found that the learned mask M_{net} (see Figure S7(a)) obtained directly by AuSamNet can not exactly realize the required sampling ratio. For a fair comparison with other SPI methods, we use some post-processing tricks to ensure the same sampling ratio. The post-processing mainly contains three steps. First, for an expected sampling ratio, we find points $S(x,y)|(x,y) \in M_{net} \cap (x,y) \notin M_{circle}$ in the non-overlapping region between M_{net} and M_{circle}. Second, we calculate the Euclidean distance between $S(x,y)$ and its nearest adjacent center O_i determined by M_{circle} (see Figure S4(a)). Third, we order points in $S(x,y)$ according

Fig. S6. Matrix T used to restrict the optimization region.

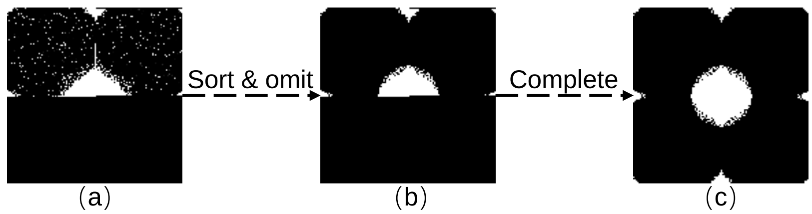

Fig. S7. Illustration of pruning method. (a) The half mask with sampling rate of 15% directly generated by the network; (b) The fine-tuned mask; (c) The completed mask.

to the calculated Euclidean distance and omit the points with higher distance to exactly realize a given sampling ratio. As such, we obtain the final learned half under-sampling mask (see Figure S7(b)) and it satisfies the requirements of the given sampling ratio. The full mask can be simply calculated according to the conjugate symmetry property (see Figure S7(c)). Figure S8(b) shows the whole learned under-sampling masks under different sampling ratio cases discussed in this Letter.

The above correction method also has some drawbacks because it is difficult to artificially determine which frequency components are more desirable for the Decoder. One does not have to use the pruning method for correction when we don't need to satisfy a given sampling rate.

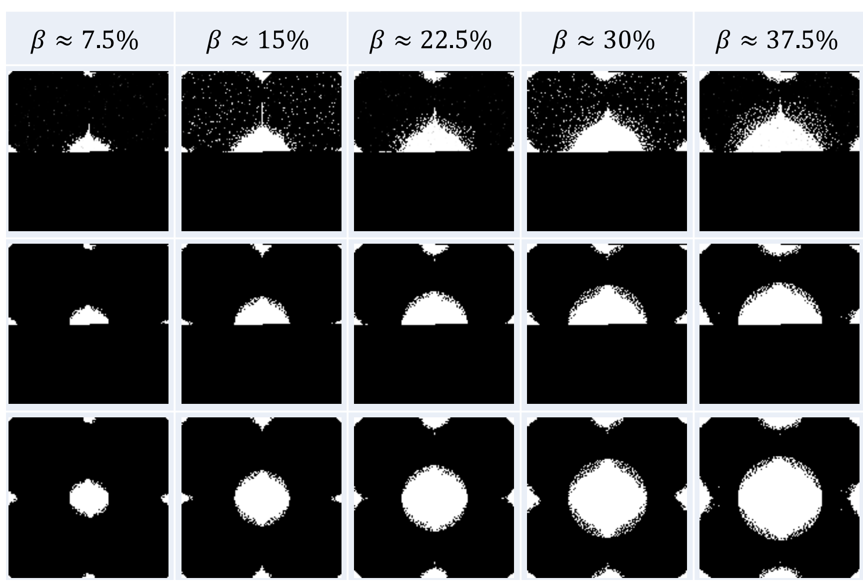

Fig. S8. The learned under-sampling mask for FSI-learned and our AuSamNet. The first row is the mask directly learned by the network. The second row is the result after pruning. The last row is the full masks.

B.4. Compared with FSI-DL using different under-sampling masks designed manually

In FSI-DL, selecting the right sampling mask is crucial, as different masks may affect the algorithm's performance. However, this raises the question of whether AuSamNet can outperform

FSI-DL when different masks are used. To explore this issue, we conducted three simulation experiments using a sampling rate of 7.5%, keeping the network architecture and training approach constant but varying the under-sampling masks. Five types of masks were used for comparison. Among them "Smaller", "Similar", and "Larger" (mask) represent cases where the radius of the central region is smaller, equal to, and larger than that of "Ours" (the mask learned by AuSamNet), respectively. "Average" represents the mask we used for FSI-DL in this work, in which the radius of the central region is equal to that of the corners and midpoints regions.

Figure S9 presents a summary of our findings, demonstrating that our proposed method using a learned under-sampling mask outperforms other methods relying on manually selected masks. It is important to note that the size of the central region in the mask has a significant impact on the quality of the reconstructed image. Specifically, reducing the central region size (as in the "Smaller" scheme) produces a more vibrant color but a blurred contour in the reconstructed image, while increasing the central region size (as in the "Larger" scheme) leads to a clearer contour but may result in inaccurate color information recovery. The "Similar" scheme, with a central region radius identical to our proposed method, produces better results than other manually designed sampling schemes. However, determining the optimal radius of the central region without prior knowledge is challenging. In contrast, AuSamNet provides an efficient approach for designing the under-sampling mask adaptively and delivers superior reconstructions without the need for manual selection of the central mask radius. We also tried different sampling rates (15% and 22.5%), and the results obtained were consistent with those obtained with 7.5% (see Figure S10 and Figure S11). Table S1 presents quantitative results that provide further evidence for our argument.

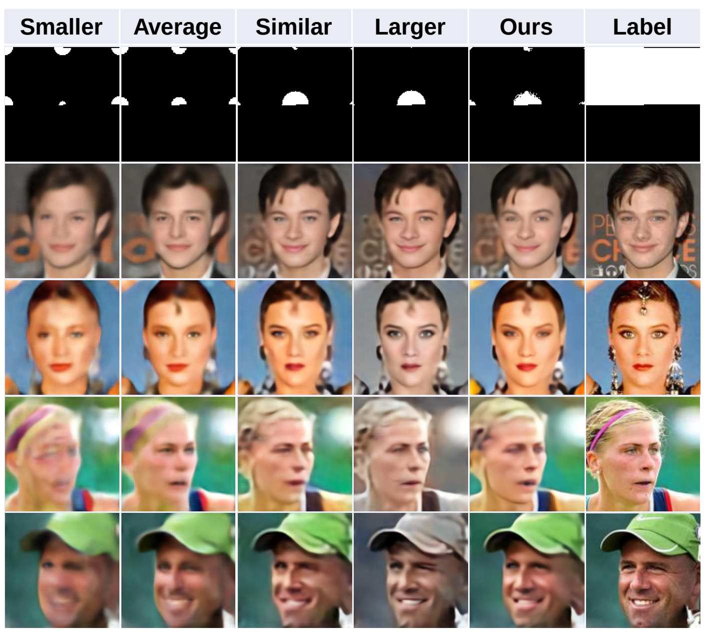

Fig. S9. Comparison of FSI-DL using different under-sampling masks M_{circle} designed manually and our AuSamNet. $\beta \approx 7.5\%$.

Table S1. PSNR/SSIM of each method with $\beta \approx 7.5\%$, $\beta \approx 15\%$, and $\beta \approx 22.5\%$. 1000 unseen test images were used for this evaluation.

β	Smaller	Average	Similar	Lager	Ours
7.5%	21.96/0.81	24.43/0.85	26.06/0.88	24.85/0.81	**26.37/0.89**
15%	25.22/0.87	26.46/0.89	27.89/0.90	27.42/0.89	**28.73/0.93**
22.5%	25.09/0.88	27.28/0.91	29.83/0.94	28.53/0.90	**30.21/0.95**

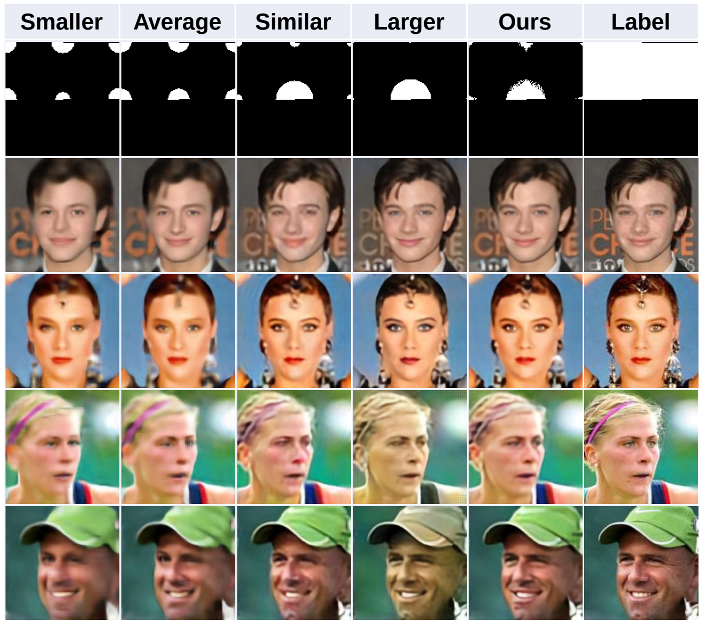

Fig. S10. Comparison of FSI-DL using different under-sampling masks M_{circle} designed manually and our AuSamNet. $\beta \approx 15\%$.

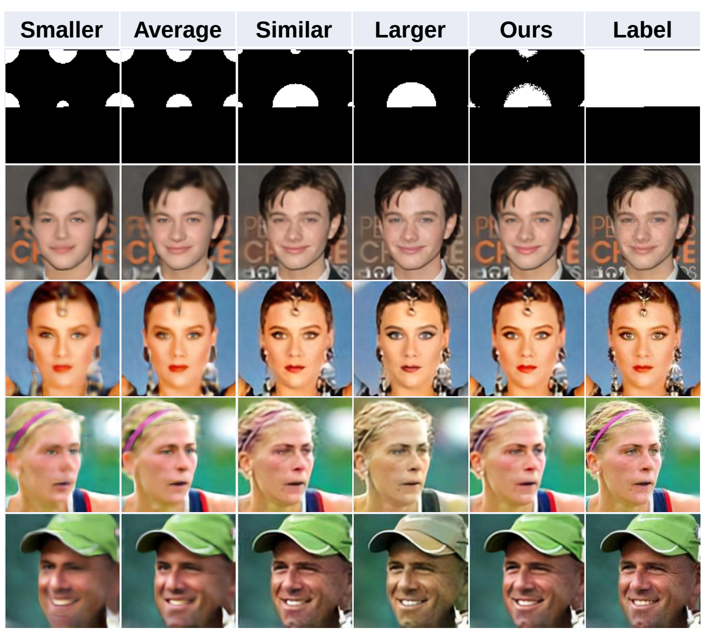

Fig. S11. Comparison of FSI-DL using different under-sampling masks M_{circle} designed manually and our AuSamNet. $\beta \approx 22.5\%$.

2. NEURAL NETWORK STRUCTURE

In the decoding part, we use a residual neural network depicted in Figure S12. Inspired by [2], we use 16 residual blocks with the same layout and each block contains two convolution layers with 3×3 kernels, 64 feature maps, followed by batch-normalization layer and Parametric ReLU. We also use a demosaics operator[3] as the last layer of the network. This allows us to obtain the predicted color image. See more details at https://github.com/FeiWang0824/AuSamNet.

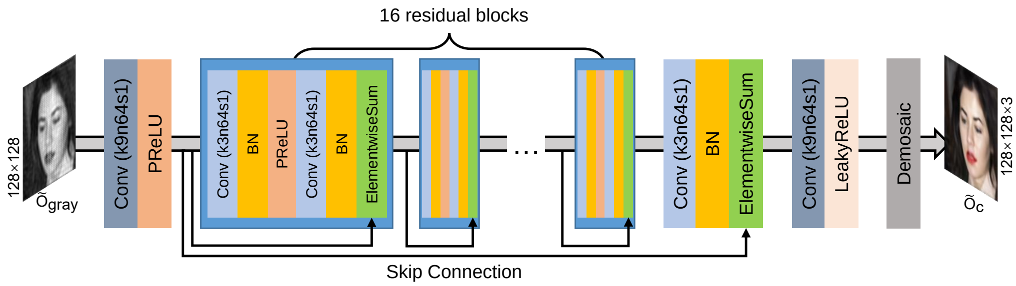

Fig. S12. Structure of the Decoder. Here, (k3n64s1) denotes a convolutional layer with a kernel size of 3, a channel number of 64, and a stride of 1. The same representation style applies to (k9n64s1).

3. SIMULATION RESULTS UNDER DIFFERENT SAMPLING RATIOS

See figure S13.

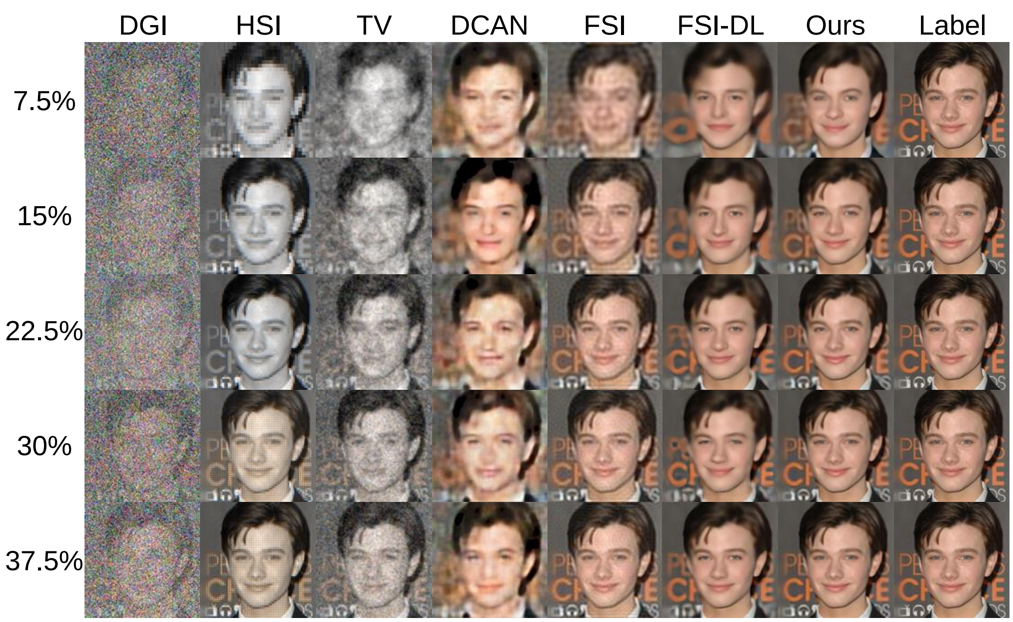

Fig. S13. Simulation results. From top to bottom are the results under different sampling ratios, and from left to right are the results under different methods.

4. OPTICAL EXPERIMENTAL RESULTS UNDER DIFFERENT SAMPLING RATIOS

See Figure S14. The image quality is evaluated using the structural similarity index (SSIM) and the peak signal-to-noise ratio (PSNR), which are expressed as follows

$$SSIM(X,Y) = \frac{(2\mu_X\mu_Y + c_1) + (2\sigma_{X,Y} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\nu_X^2 + \nu_Y^2 + c_2)} \tag{S8}$$

$$PSNR(X, Y) = 10 \log_{10} \frac{max(X)^2}{\|X - Y\|_2^2 / d} \tag{S9}$$

Where X and Y represent the actual value and reference value respectively, μ_* is the mean value of $*$, ν_* is the variance of $*$, $\nu_{X,Y}$is the covariance of X and Y, and d represents the total number of pixels in a single image.

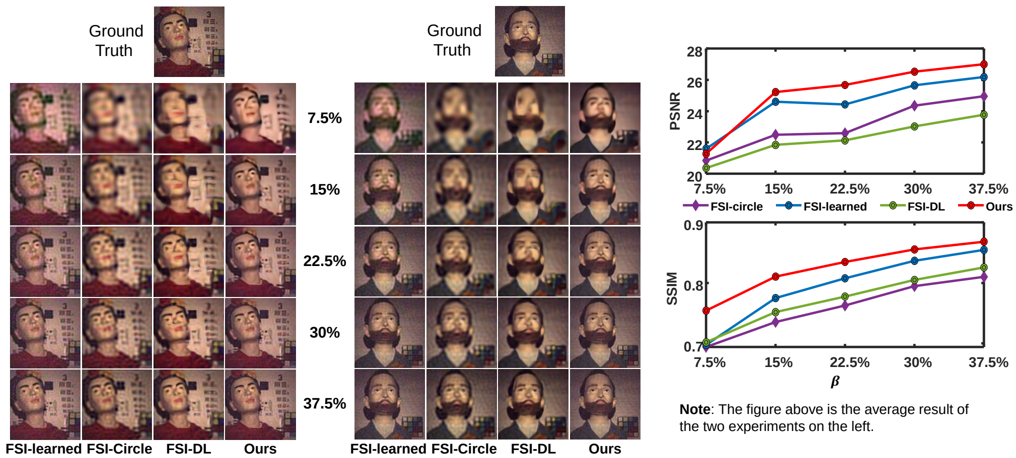

Fig. S14. Experimental results. We show the results of two different targets using FSI-learned, FSI-circle, FSI-DL, and our AuSamNet under different sampling ratio cases. SSIM and PSNR of these results are summarized in the right plots. FSI results under full sampling were used as the ground truth.

REFERENCES

1. C. D. Bahadir, A. Q. Wang, A. V. Dalca, and M. R. Sabuncu, "Deep-learning-based optimization of the under-sampling pattern in mri," IEEE Transactions on Comput. Imaging **6**, 1139–1152 (2020).
2. C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition,* (2017), pp. 4681–4690.
3. H. Malvar, L. wei He, and R. Cutler, "High-quality linear interpolation for demosaicing of bayer-patterned color images," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing,* vol. 3 (2004), pp. iii–485.