A Notations

Symbol	Description
$\boldsymbol{x}$	The fully-sampled spatial MRI image. (Ground truth for supervision. Inaccessible for self-supervision.)
$\boldsymbol{y}$	Undersampled k-space data computed with Equation 2. (The first-hand acquired data in medical practice.)
$\boldsymbol{U}$	Undersampling mask. (Indicating positions to fill in the acquisition of MRI.)
$\boldsymbol{s}$	The observation, composed of $\boldsymbol{y}$ and $\boldsymbol{U}$. (Given information after MRI acquisition.)
$\hat{\boldsymbol{x}}$	The output of the reconstruction model with Equation 3.
$\cdot^*$	Values from one-time collected sample set containing only one observation.
$\cdot^{(i)}$	Values in the i_{th} observation of sample set containing multiple observations.
$\cdot^B$	Values from the Bootstrap resampled set. $\cdot^{B_k}$ means the k_{th} set.
$\cdot^V$	Values from the virtual sample set.

B Implementation of Algorithms

Algorithm 1: Pseudo Resampling Function

```
Input: Mask $U^*$, Mask probability $P_{U^*}$, Size of virtual sample set $n$
Result: Bootstrap observation mask $U^{B*}$
def main ($U^*$, $P_{U^*}$, $n$) :
    calculate $P_U$ according to Equation 11;
    initialize $P_{U^{B*}}$ as a zeroed tensor with the same shape of $P_{U^*}$;
    fill $P_{U^{B*}}$ with Equation 15;
    draw random variable $V$ with uniform distribution in $[0, 1)$ and the same shape $U^*$;
    return the 0-1 converted result of $U^{B*} = toInteger(V < P_{U^{B*}})$
```

Algorithm 2: Bootstrap Estimation MSE of Reconstruction

```
Input: Observed k-space $y$, Mask $U^*$, Mask probability $P_{U^*}$, Size of virtual sample set $n$,
       Number of Bootstrap samples $m$, Reconstructor $f$
Result: Estimated MSE $l_{mse}$
compute the original estimation $\hat{x} = f(y, U^*)$;
$boot\_mse\_list \leftarrow \emptyset$;
for $k \leftarrow 1$ to $m$ do
    get resampling mask $U^{B_k}$ with Algorithm 1;
    resample the k-space with $y^{B_k} = U^{B_i}$;
    compute the corresponding estimation $\hat{x}^{B_k} = f(y^{B_k}, U^{B_k})$;
    compute the mean square error between the estimation and the original estimation
      $\widehat{mse}^{B_k} = (\hat{x}^{B_k} - \hat{x})^2$;
    append $\widehat{mse}^{B_k}$ to $boot\_mse\_list$ ;
average the values to get the Bootstrap estimation of MSE: $mean(boot\_mse\_list)$;
```

C Derivation of Distribution of Virtual Sample Set Observations

From an intuitive aspect, for a given position i, the size of the virtual sample set n and the corresponding equivalent sample distribution parameter $\boldsymbol{P}_{\boldsymbol{U}_i}$, the times of the position being selected in n observations in the virtual sample set, represented as $\boldsymbol{C}_i$, follows a Binomial distribution $\mathcal{B}(n, \boldsymbol{P}_{\boldsymbol{U}_i})$. Thus, with Bayes' Theorem, the conditional probability can be computed in Equation 17.

$$
\begin{aligned}
&Pr(\mathbf{U}_i^V = 1|\boldsymbol{U}_i^* = 1) \\
&= \frac{Pr(\mathbf{U}_i^* = 1|\boldsymbol{U}_i^V = 1)Pr(\mathbf{U}_i^V = 1)}{Pr(\mathbf{U}_i^* = 1)} \\
&= \frac{Pr(\mathbf{U}_i^* = 1|\boldsymbol{U}_i^V = 1)Pr(\mathbf{U}_i^V = 1)}{1 - (1 - P_{\boldsymbol{U}_i})^n} \\
&= \frac{Pr(\mathbf{U}_i^V = 1)}{1 - (1 - P_{\boldsymbol{U}_i})^n} \\
&= \frac{Pr(\mathbf{U}_i^V = 1|\boldsymbol{C}_i > 0)Pr(\mathbf{C}_i > 0) + Pr(\mathbf{U}_i^V = 1|\boldsymbol{C}_i = 0)Pr(\mathbf{C}_i = 0)}{1 - (1 - P_{\boldsymbol{U}_i})^n} \\
&= \frac{Pr(\mathbf{U}_i^V = 1|\boldsymbol{C}_i = k)Pr(\mathbf{C}_i > 0)}{1 - (1 - P_{\boldsymbol{U}_i})^n} \\
&= \frac{\sum_{k=1}^{+\infty} Pr(\mathbf{U}_i^V = 1|\boldsymbol{C}_i = k)Pr(\mathbf{C}_i = k)}{1 - (1 - P_{\boldsymbol{U}_i})^n} \\
&= \frac{\sum_{k=1}^{n} \frac{k}{n} Pr(\mathbf{C}_i = k)}{1 - (1 - P_{\boldsymbol{U}_i})^n} \\
&= \frac{\sum_{k=1}^{n} \frac{k}{n} \binom{k}{n} P_{\boldsymbol{U}_i}^k P_{\boldsymbol{U}_i}^{n-k}}{1 - (1 - P_{\boldsymbol{U}_i})^n}
\end{aligned}
\tag{17}
$$

The computation is too complicated and hard to use, so we provide a simplified version by analyzing three special cases:

1. $\boldsymbol{C}_i = 0$, which means that the position is not selected in all observations, thus $\boldsymbol{U}^V$ will always be zero.
2. $\boldsymbol{P}_{U_i} = 1$, common in the MRI acquisition to keep more low-frequency information. In this case, every observation in the virtual sample set contains the position.
3. $\boldsymbol{P}_{U_i}$ is very small. In this case, the probability of being selected again ($\boldsymbol{C}_i > 1$) in the observations is small and can be neglected. If $\boldsymbol{C}_i \neq 0$, we can assume $\mathbf{U}^V \sim \mathcal{B}(\boldsymbol{U}^V; 1, 1/n)$.

Note that only $\boldsymbol{P}_{\boldsymbol{U}_i^*}$ is given and $\boldsymbol{P}_{U_i}$ is computed with Equation 11, and n is a hyper-parameter that can be set manually, so for better approximation we can choose larger n when $\boldsymbol{P}_{\boldsymbol{U}_i^*}$ is not small enough. As a result, the third case can actually cover all halfway situations by setting proper n. Some examples of C_i distributions are shown in Figure 9. It can be found that the conditional distributions (golden bars) have a much higher probability in $\boldsymbol{C}_i = 1$, which means nearly all sampled positions are sampled only once.

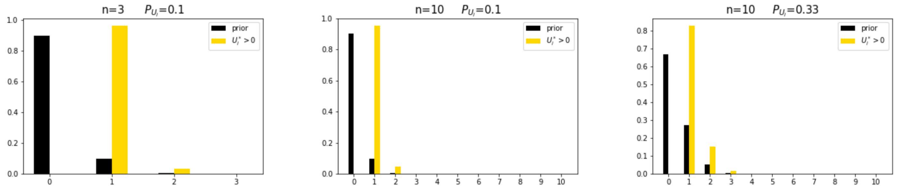

Figure 9: Examples of $\boldsymbol{C}_i$'s distribution with different n and $\boldsymbol{P}_{\boldsymbol{U}_i^*}$. The golden bars show distributions when $\boldsymbol{C}_i > 0$ is conditioned. The corresponding $\boldsymbol{P}_{\boldsymbol{U}_i}$ is around 0.0345, 0.0105 and 0.0393.

Combining the situations, we can get the distribution of the mask.

$$P_{\boldsymbol{U}_i^V} = Pr(\mathbf{U}_i^V = 1) = \begin{cases} 1 & \boldsymbol{P}_{U_i} = 1 \\ 1/n & \boldsymbol{U}_i^* \neq 0 \& \boldsymbol{P}_{U_i} \neq 1 \\ 0 & \boldsymbol{U}_i^* = 0 \end{cases} \tag{18}$$

D The Pipeline of Bootstrap Self-supervised Training

By referring to Algorithm 2, the Bootstrap estimated MSE can serve as a loss function for self-supervised training. The PyTorch-style code snippet of the training process is shown below. We use stop-gradient and inverse consistency to enable and stabilize the training.

```
n = 5
m = 20

with torch.no_grad():
    x_orginial=model(y*mask,mask)
    y_original=fft(y_original)
for _ in range(m):
    re_mask=generate_re_mask(mask,p_maskin,n)
    x_bootstrap=model(y*re_mask,re_mask)
    y_bootstrap=fft(x)
    x_bootstrap=ifft(y_bootstrap*(~mask)+y_orginial*mask)
    loss=nn.MSELoss(x_bootstrap,x_orginial)/m
    loss.backward()
optimizer.step()
```

E Briefs of Other Zero-shot Models

Besides the basic SENSE reconstruction that zero-fills the missing k-space, the differences between the models mainly come from the different loss functions. All the models are based on the same backbone structure of 8-layer DC-CNN. Note that due to the memory limit, the weight of the model is shared among the z-axis.

For the DIP models, the loss function is limited to keeping the consistency of all the knowing positions. Due to the data consistency layers in the model, the loss in k-space will always be 0 and nothing can be learned, so the loss is defined in the image domain as Equation 19, where f_{SENSE} is the SENSE reconstruction.

$$\mathcal{L}_{DIP}(\hat{\boldsymbol{x}}, \boldsymbol{y}^*) = mse(f_{SENSE}(\boldsymbol{y}^*), \hat{\boldsymbol{x}}) \tag{19}$$

The SSDU model applies re-undersampling in training and defines a normalized $l_1 - l_2$ loss in k-space as Equation 20. As for the super-parameters, we choose $K = 10$ and the re-undersampling ratio of 40% following the optimal setting of original paper(Yaman et al., 2020), and a uniform re-undersampling strategy is used. Also, Gaussian re-undersampling is conducted as described in the paper instead of deriving re-undersampling masks based on virtual sample set. The settings are modified to follow our basic setting of sharing weights among different slices.

$$\mathcal{L}_{SSDU}(\hat{\boldsymbol{x}}, \boldsymbol{y}^*) = \frac{||\boldsymbol{y}^* - \mathcal{F}\hat{\boldsymbol{x}}||_2}{||\boldsymbol{y}^*||_2} + \frac{||\boldsymbol{y}^* - \mathcal{F}\hat{\boldsymbol{x}}||_1}{||\boldsymbol{y}^*||_1} \tag{20}$$

F Implementation Methods

Given a reconstructor f and an acquisition $\boldsymbol{y}^*$ and $\boldsymbol{U}^*$ of the spatial anatomy $\boldsymbol{x}$, a reconstruction and estimation result can be obtained by $\hat{\boldsymbol{x}} = f(\boldsymbol{y}^*, \boldsymbol{U}^*) = f_{AF}((\boldsymbol{y}^*, \boldsymbol{U}^*))$. MSE and its provide a simple but effective evaluation of the differences between $\hat{\boldsymbol{x}}$ and the target $\boldsymbol{x}$.

To conduct the bootstrapping in reconstructing a single MRI observation, although it may be complicated to state the transformation of the problem, the resulting solution is rather simple and intuitive. The construction of a virtual sample set and the computation of aggregation function become invisible parts of the pipeline. The only extra thing we need to do is to compute a re-undersampling mask based on the selected cardinality n, the actual mask $\boldsymbol{U}$ and its distribution parameter $\boldsymbol{P_U}$. We provide the pseudo code of the computation in Algorithm 1. As bootstrapping enables the computation of various statistic properties without acquiring more sample sets, statistics like MSE can be estimated with only a single observation of partial measurements acquired CS MRI. The pseudo code of unsupervised computation of MSE can be found in Algorithm 2 based on Equation 13.

It can be seen that besides the input and the reconstructor, the most important hyper-parameters of the algorithms are the cardinal of the virtual sample set n, which decides the ratio and positions of re-undersampling and the times of Bootstrap sampling m, which increases the accuracy of Bootstrap estimators as it grows.

G Quantitative Comparison of Uncertainty Quantification

With residual errors as UQ notion, we compare our methods with residual magnitude prediction(Angelopoulos et al., 2022), a method that directly enables a separated prediction head to predict the errors and we name it Residual Regression (RR) for simplification. The model is trained on fastMRI single-coil dataset with the same setting from the original paper except for a smaller batch size (48 v.s. 78) due to the limited GPU memory. After training, we directly apply Algorithm 2 of BootRec in the trained model to get our quantification, and square the predictions from RR. We also apply Stein's Unbiased Risk Estimator (SURE) to estimate the trained model's MSE, with settings from (Edupuganti et al., 2020).

Evaluation of UQ is still an open problem, with various metrics focusing on different notions and aspects. Considering the task, we conduct inter-image evaluations and intra-image evaluations. For inter-image evaluation, we average the pixels' estimated MSE for every image and collect the ground truth MSE (GT MSE) of the reconstruction. The visualization can be found in Figure 10 and the quantitative results of correlation coefficients are summarized in Table 1. It can be seen that BootRec estimated MSE continuously presents better coefficients compared with predictions from RR head and has the highest values in Spearman and Kendall Coefficients. For intra-image evaluation, the distribution of the errors is more important than the absolute values as it indicates areas of hallucination. For the inter-image evaluation, we display an example from validation set for comparison of error distributions in Figure 11. The GT MSE shows high error levels in border areas, which is reasonable as the pixel values change rapidly, and the Bootstrap estimated error shows a similar distribution. The RR and SURE error maps put too much uncertainty inside the tissue, especially for the RR results. SURE estimations performed well in inter-image correlation but failed in inter-image distribution of errors, with one possible explanation that the residual variance of specific position is hard to compute.

	Pearson	Spearman	Kendall
RR	0.866	0.864	0.692
SURE	**0.941**	0.752	0.561
Bootstrap	0.888	**0.920**	**0.756**

Table 1: Correlation Coefficients computed corresponding to Figure 10. All values have a negligible p-value (we get zero in computation). The Bootstrap estimator achieves best correlation in Spearman and Kendall Coefficients.

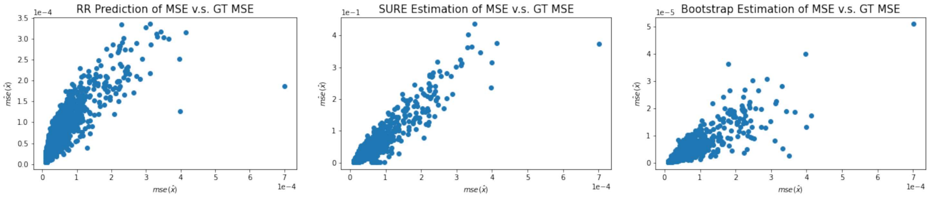

Figure 10: Inter-image Correlation of predicted/estimated MSE in the validation set.

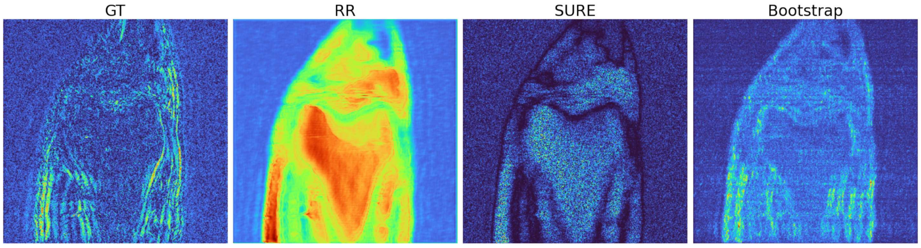

Figure 11: Intra-image distribution of predicted/estimated errors in the validation set (square roots are computed for MSE results). Note that the values are normalized across the image.