Task-Driven Uncertainty Quantification in Inverse Problems via Conformal Prediction

Jeffrey Wen[1], Rizwan Ahmad[1], Philip Schniter[1]

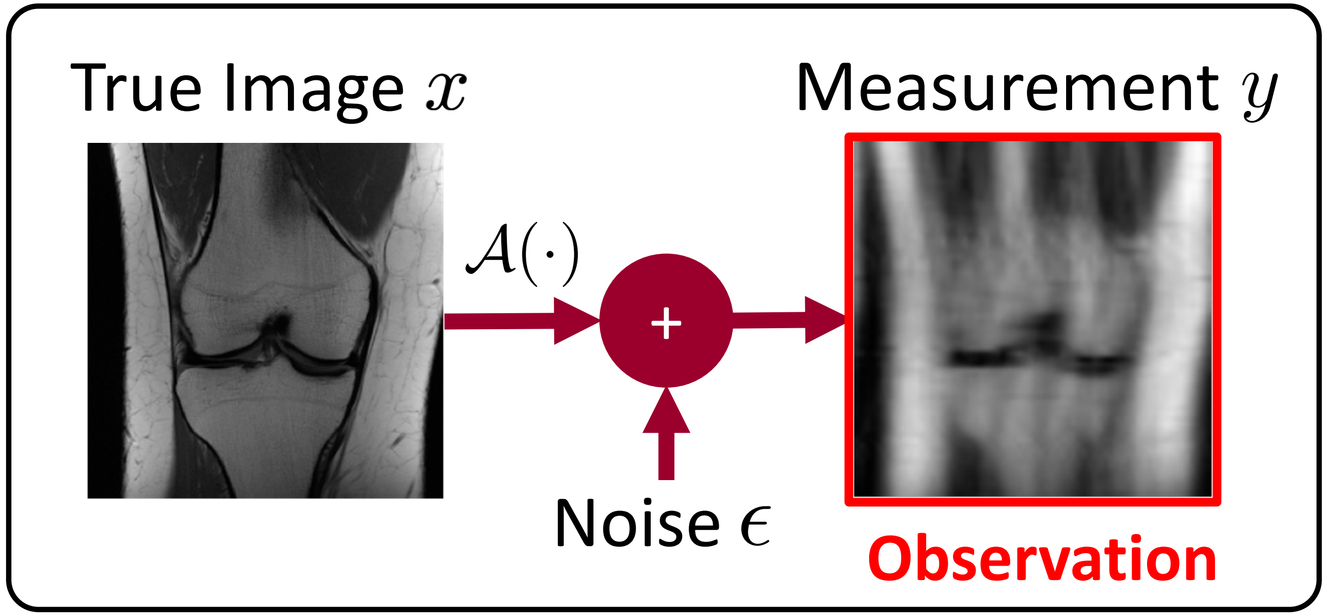
[1]The Ohio State University, Columbus, OH, US

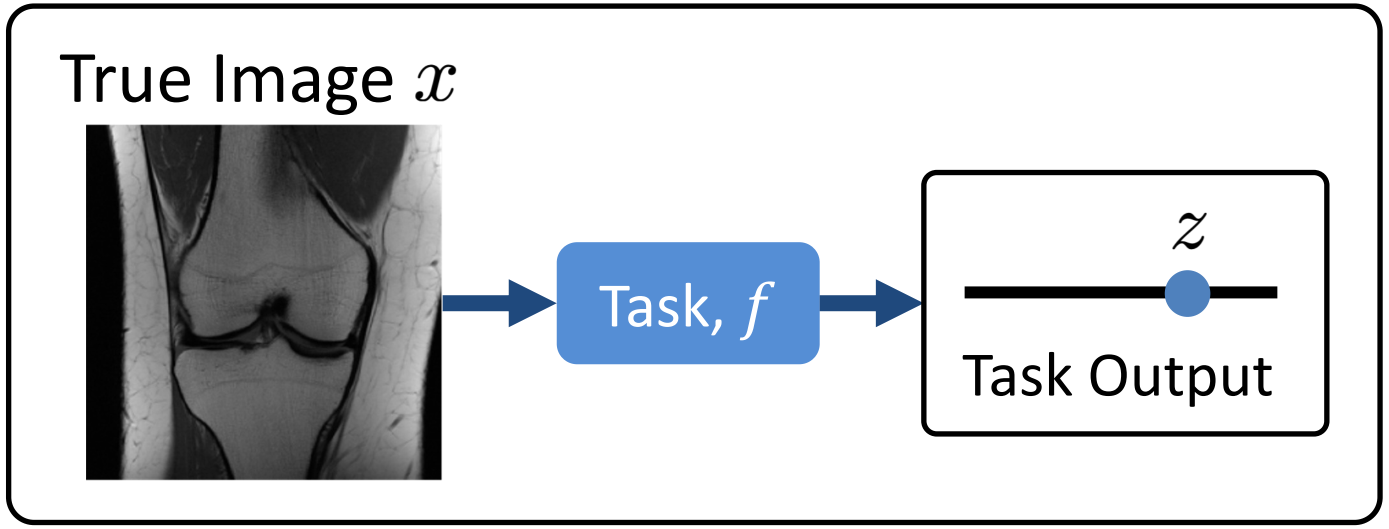
2024 Asilomar Conference on Signals, Systems, and Computers

This work was supported by NIH Grant R01EB029957

Introduction

- Consider
  - Inverse problem $y = \mathcal{A}(x) + \epsilon$
    - Accelerated MRI
    - Inpainting, super-resolution, etc
  - Imaging task $f(x) \in \mathbb{R}$
    - Classification (soft-output)
    - Tumor size estimation

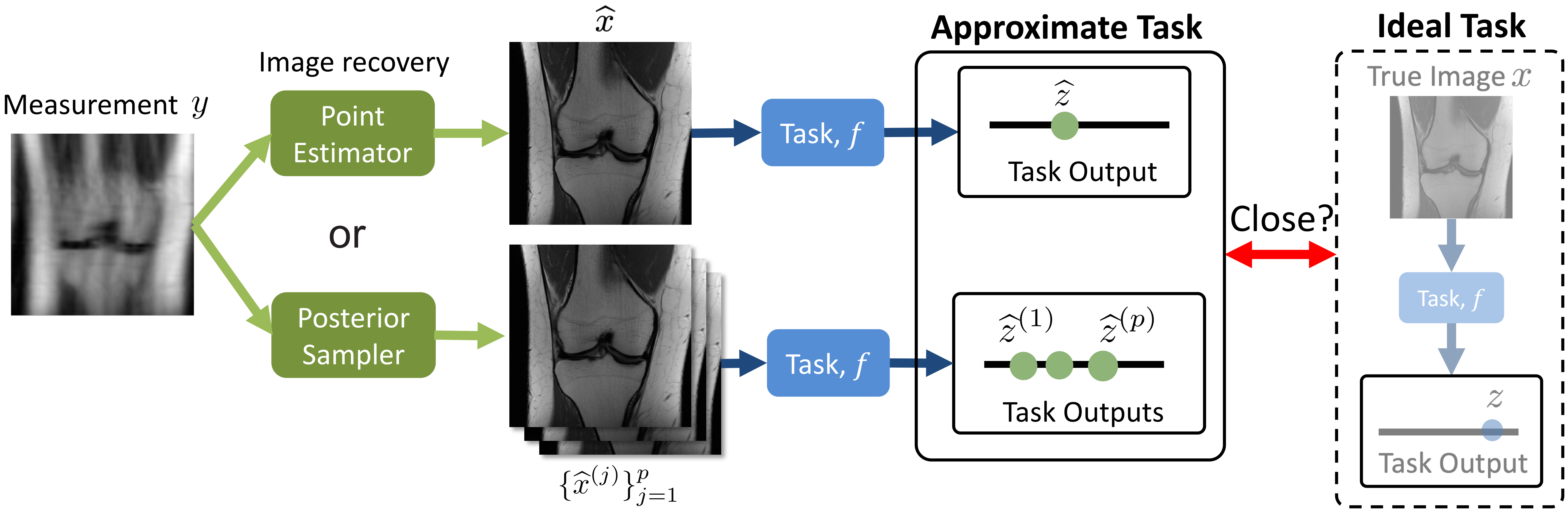




- Goal: Perform the task given only the measurements y

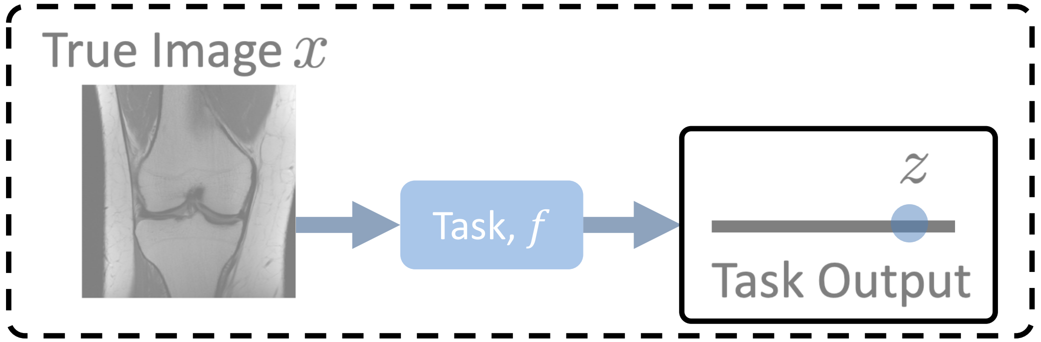
Introduction

- Could approximate the task using reconstructed image(s) $\{\hat{x}^{(j)}\}_{j=1}^{p}$



- Rather than quantifying the accuracy or uncertainty of $\widehat{z}$ or $\widehat{z}^{(j)}$, we compute rigorous bounds on the true value of z itself

Objective



- Given:
  - Measurement y
  - Image reconstruction method/sampler $y \mapsto \{\widehat{x}^{(j)}\}$
  - Calibration set

- Goal: Construct prediction interval $\mathcal{C}_{\widehat{\lambda}}(\{\widehat{x}^{(j)}\}) = [b_l, b_u]$ guaranteed to contain the unknown true task output z with high probability

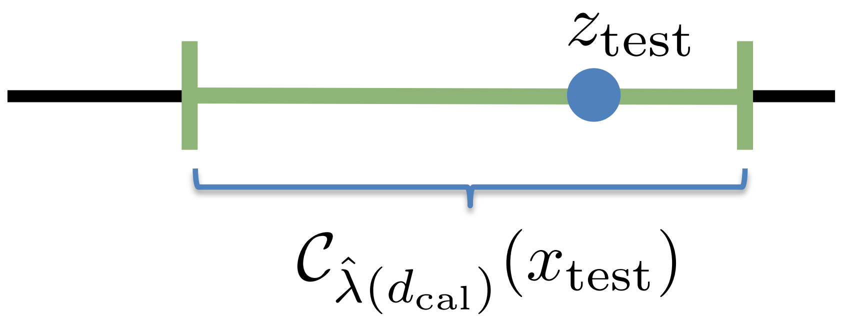
Prediction Interval

Where: $\mathbb{P}(Z \in \mathcal{C}_{\widehat{\lambda}(D_{\text{cal}})}(\{\widehat{X}^{(j)}\})) \geq 1 - \boxed{\alpha}$

User-chosen error rate

Main tool: conformal prediction

- Setup
  - Given a black-box predictor $f : \mathcal{X} \rightarrow \mathcal{Z}$
  - Given a calibration set $d_{\mathrm{cal}} = \{(x_i, z_i)\}_{i=1}^{n}$
  - Choose an error rate α
  - Construct a prediction set $\mathcal{C}_{\lambda}(x) \subset \{\text{subsets of } \mathcal{Z}\}$



- Compute $\hat{\lambda}(d_{\mathrm{cal}})$ such that $\underbrace{\mathbb{P}(Z_{\mathrm{test}} \in \mathcal{C}_{\hat{\lambda}(D_{\mathrm{cal}})}(X_{\mathrm{test}}))}_{\text{on expectation over random calibration set and test sample}} \geq 1 - \alpha$

Holds when $\{(X_1, Z_1), \ldots, (X_n, Z_n), (X_{\mathrm{test}}, Z_{\mathrm{test}})\}$ are exchangeable

V. Vovk, A. Gammerman, and G. Shafer. *Algorithmic Learning in a Random World*. Springer, 2005.
Anastasios N. Angelopoulos and Stephen Bates. Conformal prediction: A gentle introduction. *Foundations and Trends in Machine Learning*, 2023.

Constructing the prediction set

1. Choose a nonconformity score $s(x, z; f) \in \mathbb{R}$
   - Higher values of s correspond to worse predictions of z

2. Compute scores for calibration samples:

$$s_i = s(x_i, z_i; f), \; i = 1, \ldots, n$$

3. Compute empirical quantile

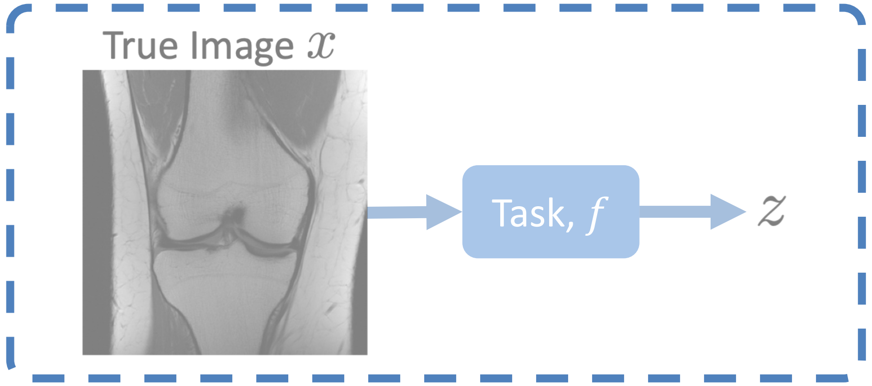
$$\hat{\lambda}(d_{\text{cal}}) = \text{Quantile}\left(\frac{\lceil(1-\alpha)(n+1)\rceil}{n}; s_1, \ldots, s_n\right)$$

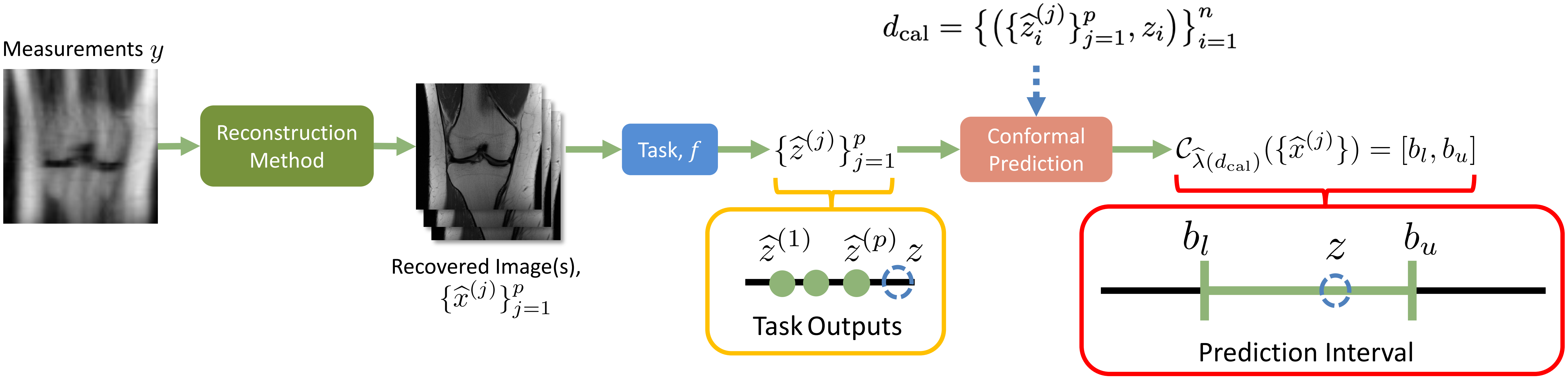
4. Construct prediction set

$$\mathcal{C}_{\hat{\lambda}(d_{\text{cal}})}(x_{\text{test}}) = \{z : s(x_{\text{test}}, z; f) \leq \hat{\lambda}(d_{\text{cal}})\}$$

Our method: Overview

➢ Goal: Bound the true task output z



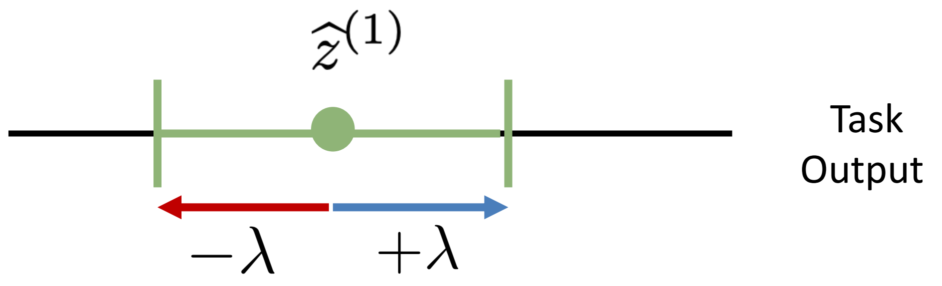


Where: $\mathbb{P}(Z \in \mathcal{C}_{\widehat{\lambda}(D_{\text{cal}})}(\{\widehat{X}^{(j)}\})) \geq 1 - \alpha$

Our method: Three ways to construct prediction sets

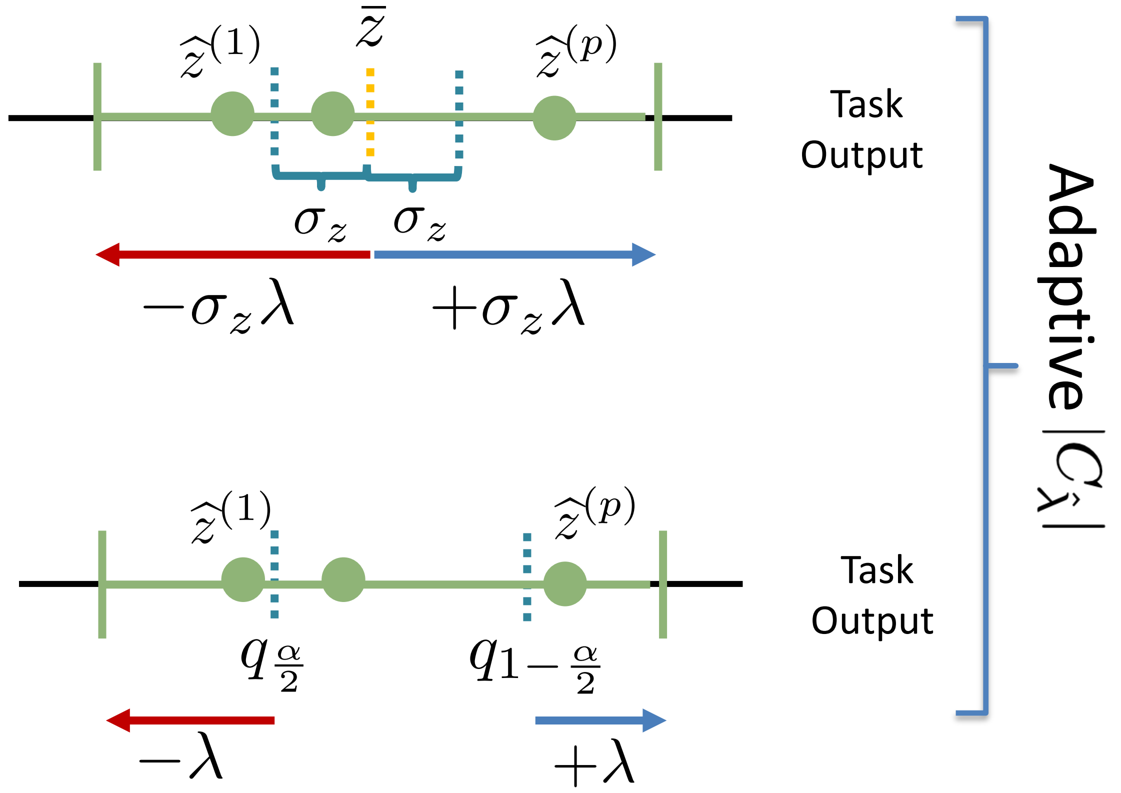
- Absolute Residual (AR)

$$C_\lambda(\widehat{x}^{(1)}) = [\widehat{z}^{(1)} - \lambda, \widehat{z}^{(1)} + \lambda]$$



- Locally-Weighted Residual (LWR)

$$C_\lambda(\{\widehat{x}^{(j)}\}) = [\bar{z} - \sigma_z\lambda, \bar{z} + \sigma_z\lambda] \begin{cases} \bar{z} = \frac{1}{p}\sum_{j=1}^{p} \widehat{z}^{(j)} \\ \sigma_z^2 = \frac{1}{p}\sum_{j=1}^{p} (\widehat{z}^{(j)} - \bar{z})^2 \end{cases}$$



- Conformalized Quantile Regression (CQR)

$$C_\lambda(\{\widehat{x}^{(j)}\}) = [q_{\frac{\alpha}{2}}(\{\widehat{z}^{(j)}\}) - \lambda,\ q_{1-\frac{\alpha}{2}}(\{\widehat{z}^{(j)}\}) + \lambda]$$

Anastasios N. Angelopoulos and Stephen Bates. Conformal prediction: A gentle introduction. *Foundations and Trends in Machine Learning*, 2023.
Romano, Y., Patterson, E., Candès, E.J.: Conformalized quantile regression. *Proc. NeurIPS*, 2019.

Experiments

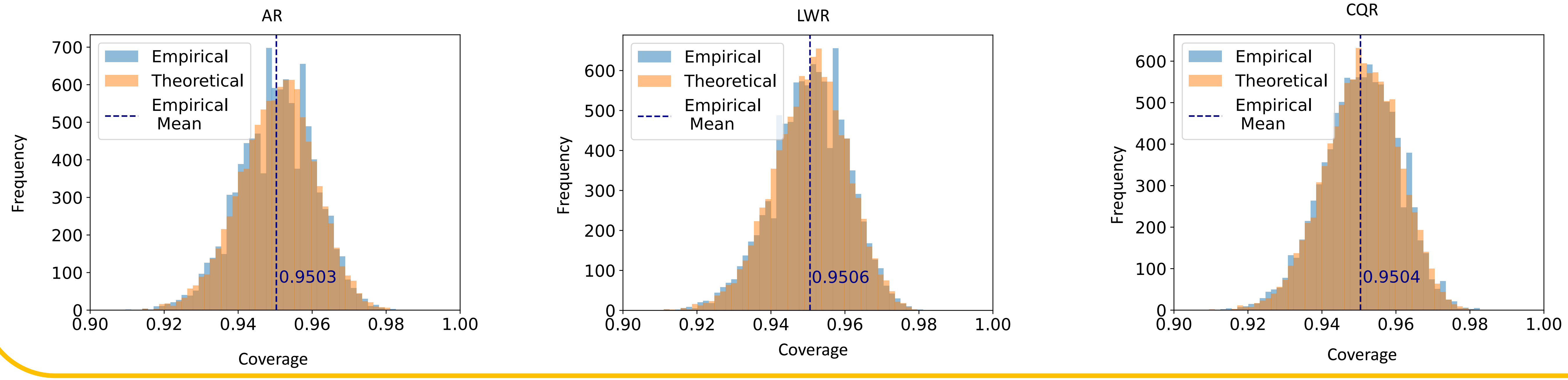
- Dataset: fastMRI multicoil knee, non-fat-suppressed
- Task: Soft-output meniscus-tear classification
- Reconstruction models
  - Non-adaptive: E2E-VarNet point estimator
  - Adaptive: Conditional Normalizing Flow (CNF) posterior sampler

Zbontar et al., fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv:1811.08839,* 2018.
Sriram, Zbontar, Murrell, Defazio, Zitnick, Yakubova, Knoll, and Johnson. End-to-end variational networks for accelerated MRI reconstruction. *Proc. MICCAI*, 2020.
Wen, Ahmad, and Schniter. A conditional normalizing flow for accelerated multi-coil MR imaging. *Proc. ICML*, 2023.

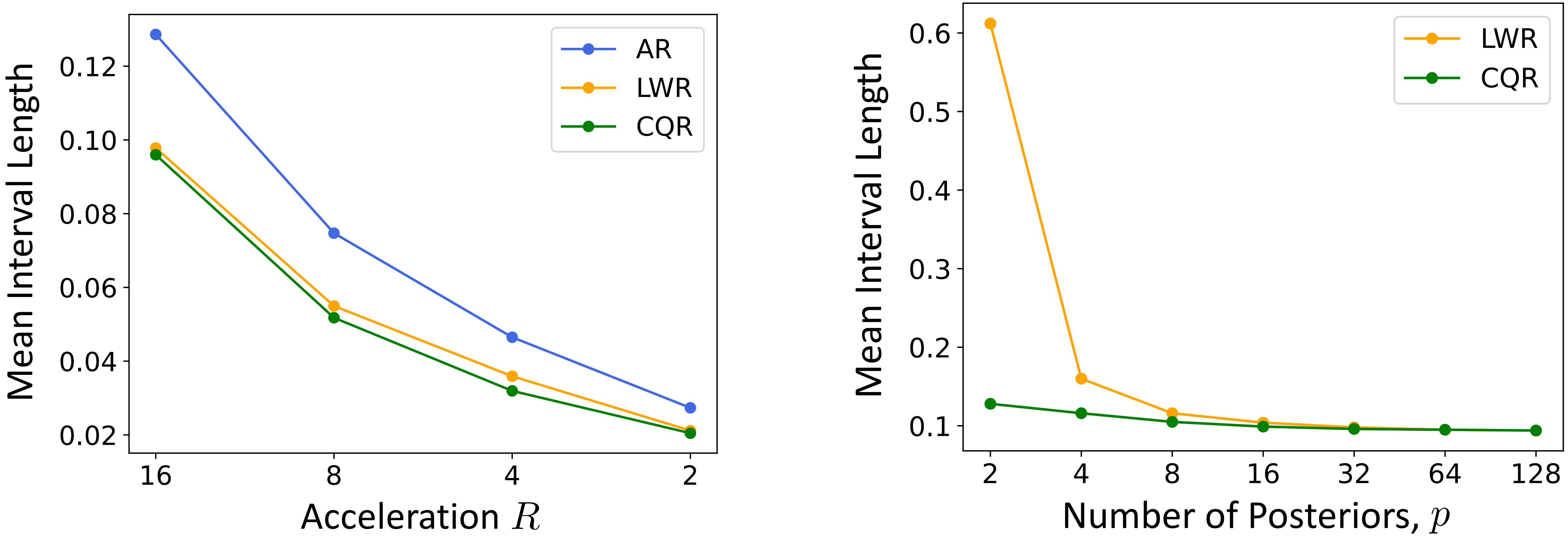
Empirical verification of coverage $\mathbb{P}(Z_{\text{test}} \in C_{\hat{\lambda}(D_{\text{cal}})}) \geq 1 - \alpha$

- Perform $T = 10000$ Monte Carlo trials:
  - Randomly split validation set
    - 70% for calibration
    - 30% for testing
  - Compute empirical coverage $$\mathrm{EC}[t] = \frac{1}{|\mathcal{I}_{\text{test}}[t]|} \sum_{i \in \mathcal{I}_{\text{test}}[t]} \mathbb{1}\left\{z_i \in C_{\hat{\lambda}(d_{\text{cal}}[t])}(\{\hat{x}_i^{(j)}\})\right\}$$

Histograms of Empirical Coverage for $\alpha = 0.05$



Comparison of prediction methods

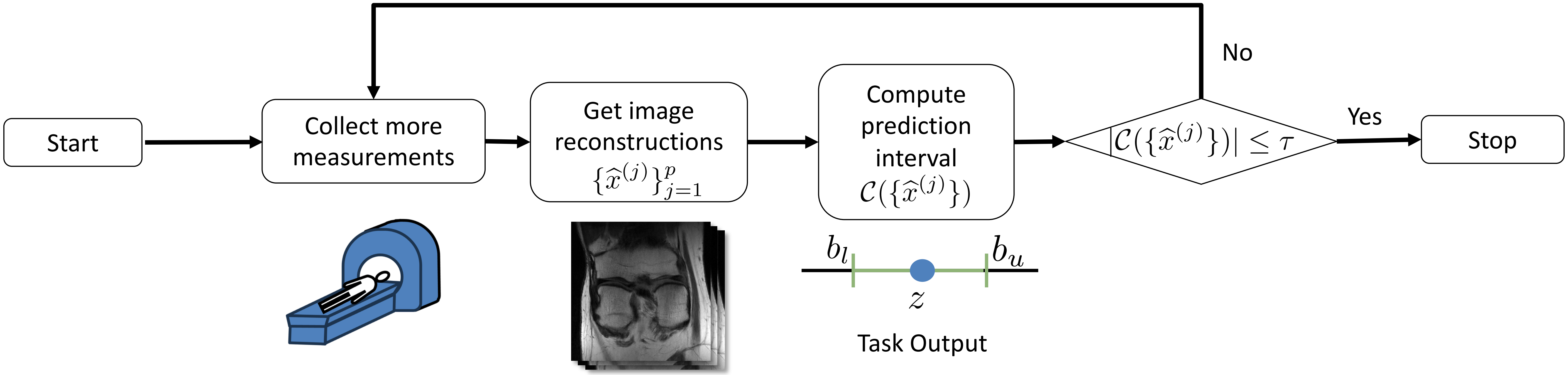


(a) Mean Interval Length vs. R

(b) Mean Interval Length vs p

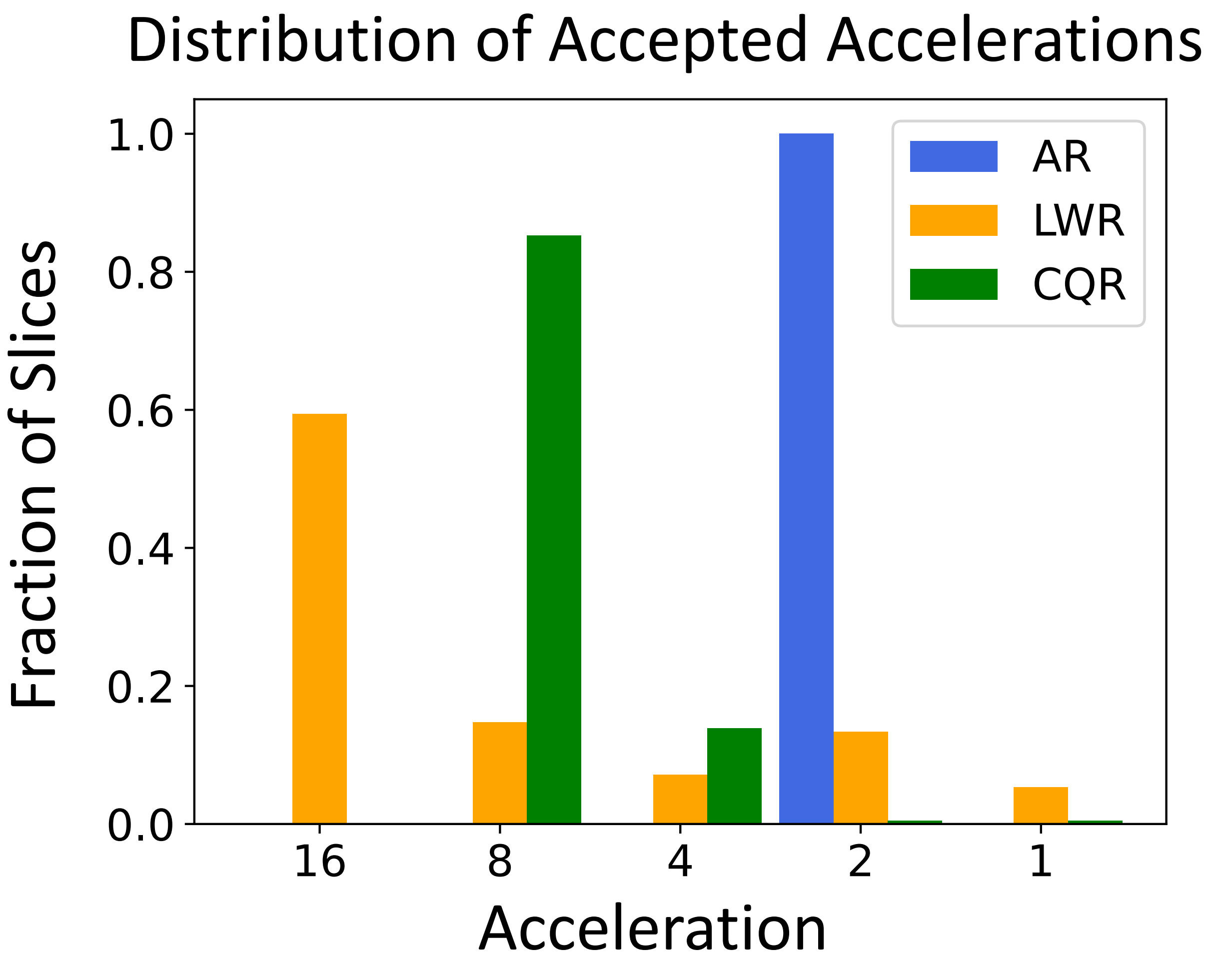
- Adaptive schemes yield smaller uncertainty intervals $|C_{\hat{\lambda}(d_{\mathrm{cal}})}|$
- CQR is robust to the use of very few posterior samples p

A multi-round measurement scheme with uncertainty guarantees



Quantitative results for the multi-round measurement scheme

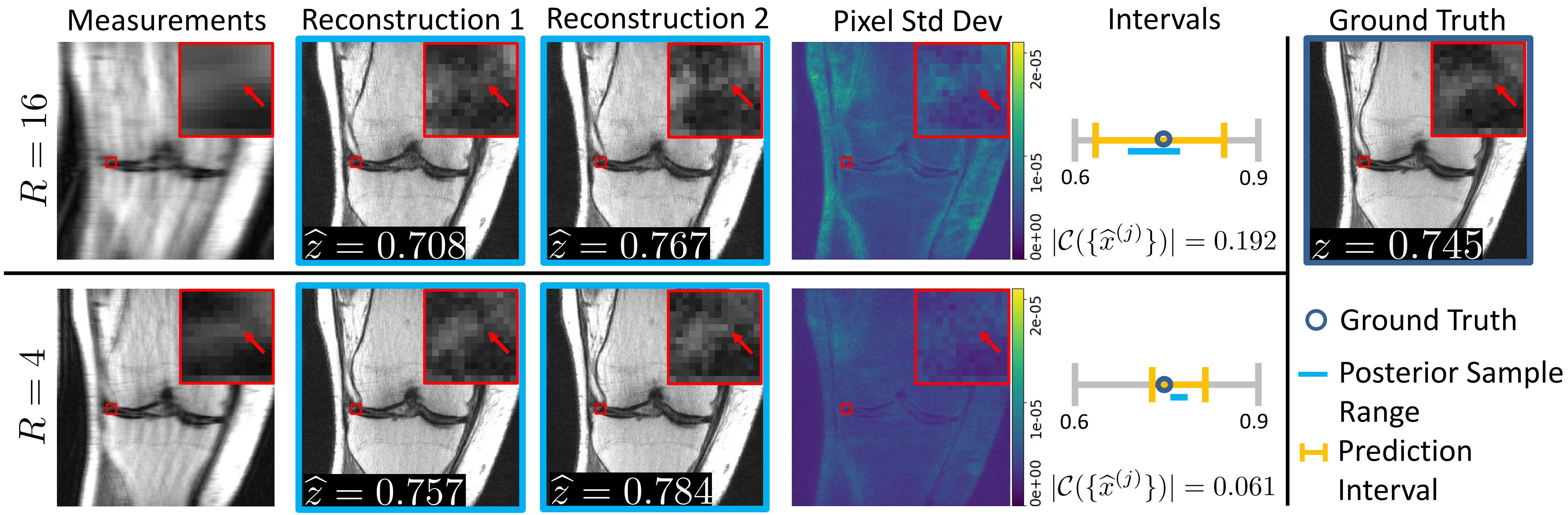
- Error rate $\alpha = 0.01$
- Threshold $\tau = 0.1$



Method	Average Acceleration	Empirical Coverage
AR	2.000	0.991 ± 0.008
LWR	5.157	0.992 ± 0.005
CQR	6.762	0.987 ± 0.008

$\pm$ Standard Error

Visual examples for the multi-round measurement scheme



➢ If the threshold was $\tau = 0.1$ then $R = 4$ would suffice, but not $R = 16$

Conclusion

- Goal: Construct an interval guaranteed to contain the true task output z with probability $\geq 1-\alpha$

- Accomplished this using conformal prediction
  - Investigated three prediction methods: AR, LWR, CQR
  - LWR and CQR leverage a posterior sampler to adapt $|C_{\hat{\lambda}}|$ to y
  - CQR is robust to very few posterior samples

- Proposed a multi-round protocol with uncertainty guarantees
  - Adaptive schemes offer large acceleration gains over non-adaptive

Intuition behind conformal prediction

- Suppose $D_{\text{cal}} = \{Z_i\}_{i=1}^n$ with i.i.d $\{Z_1, \ldots, Z_n, Z_{\text{test}}\}$
  - We're ignoring the feature X to keep things simple

- Suppose we want to find $C_\lambda = (-\infty, \lambda]$ and $\hat{\lambda}(D_{\text{cal}})$ such that

$$\mathbb{P}(Z_{\text{test}} \in C_{\hat{\lambda}(D_{\text{cal}})}) \geq 1 - \alpha$$

- If we set $\hat{\lambda}(D_{\text{cal}}) = \text{Quantile}(1-\alpha; Z_1, \ldots, Z_n)$ then

$$\mathbb{P}(Z_{\text{test}} \leq \hat{\lambda}(D_{\text{cal}})) = \mathbb{P}(Z_{\text{test}} \in C_{\hat{\lambda}(D_{\text{cal}})}) = 1 - \alpha \quad \text{as} \quad n \to \infty$$

- Can we do something similar that holds with finite n?

Tibshirani R., Conformal Prediction. *Advanced Topics in Statistical Learning*. Spring 2023.

Intuition behind conformal prediction (cont.)

- Note that, since $\{Z_1, \ldots, Z_n, Z_{\text{test}}\}$ are i.i.d, the rank of Z_{test} is uniform in $\{1, 2, \ldots, n+1\}$

- This implies that

$$
\begin{aligned}
&\mathbb{P}(Z_{\text{test}} \leq \text{the } \lceil (1-\alpha)(n+1) \rceil \text{ smallest of } Z_1, \ldots, Z_n, Z_{\text{test}}) \geq 1-\alpha \\
\Leftrightarrow\ &\mathbb{P}(Z_{\text{test}} > \text{the } \lceil (1-\alpha)(n+1) \rceil \text{ smallest of } Z_1, \ldots, Z_n, Z_{\text{test}}) < \alpha \\
\Leftrightarrow\ &\mathbb{P}(Z_{\text{test}} > \text{the } \lceil (1-\alpha)(n+1) \rceil \text{ smallest of } Z_1, \ldots, Z_n) < \alpha \\
\Leftrightarrow\ &\mathbb{P}(Z_{\text{test}} \leq \text{the } \lceil (1-\alpha)(n+1) \rceil \text{ smallest of } Z_1, \ldots, Z_n) \geq 1-\alpha
\end{aligned}
$$

- Equivalently,

$$\mathbb{P}(Z_{\text{test}} \leq \hat{\lambda}(D_{\text{cal}})) = \mathbb{P}(Z_{\text{test}} \in C_{\hat{\lambda}(D_{\text{cal}})}) \geq 1-\alpha$$

when $\hat{\lambda}(D_{\text{cal}}) = \text{Quantile}\left(\frac{\lceil (1-\alpha)(n+1) \rceil}{n}; Z_1, \ldots, Z_n\right)$