ML4Seismic Partners Meeting 2023

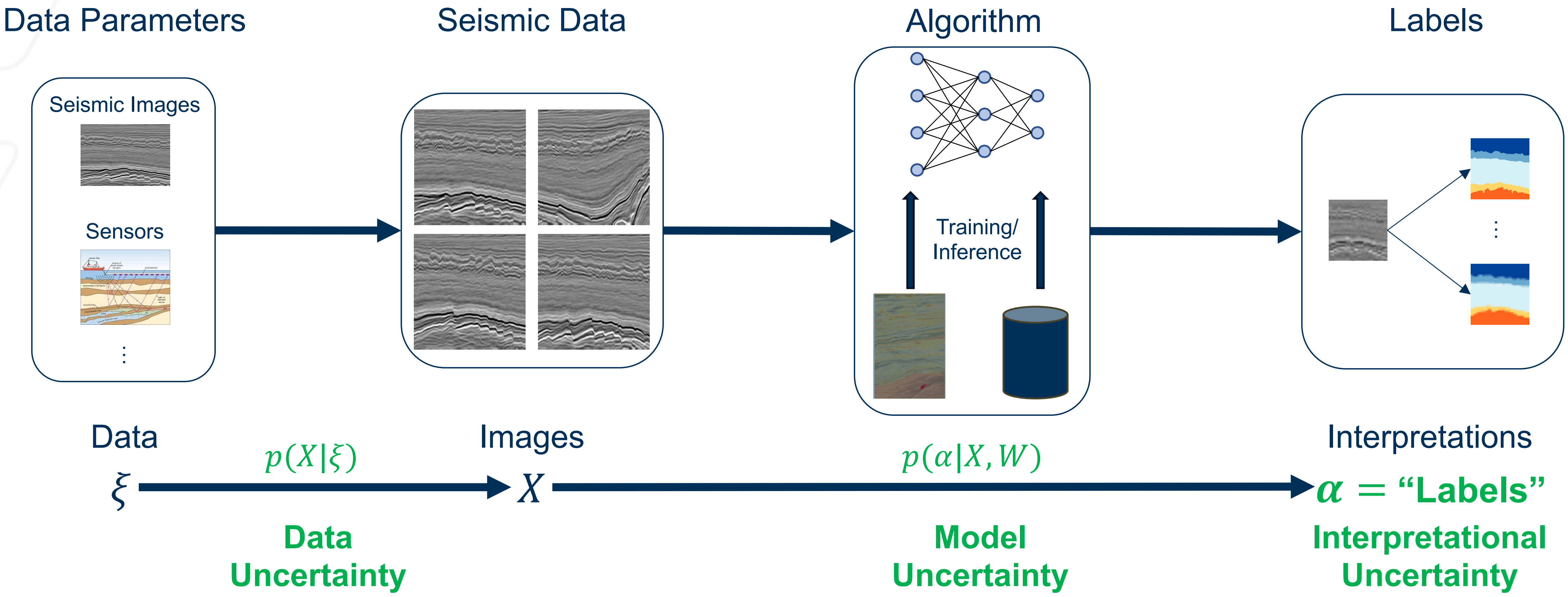
# Interpretation Distribution-aware Sample Selection for Reducing Expert Labeling Efforts

Chen Zhou, Jorge Quesada, Mohit Prabhushankar, and Ghassan AlRegib

# Uncertainty in Seismic Interpretation

## The Framework Reflects the Processing Steps of the Entire Interpretation Pipeline

### Uncertainty in Seismic Interpretation

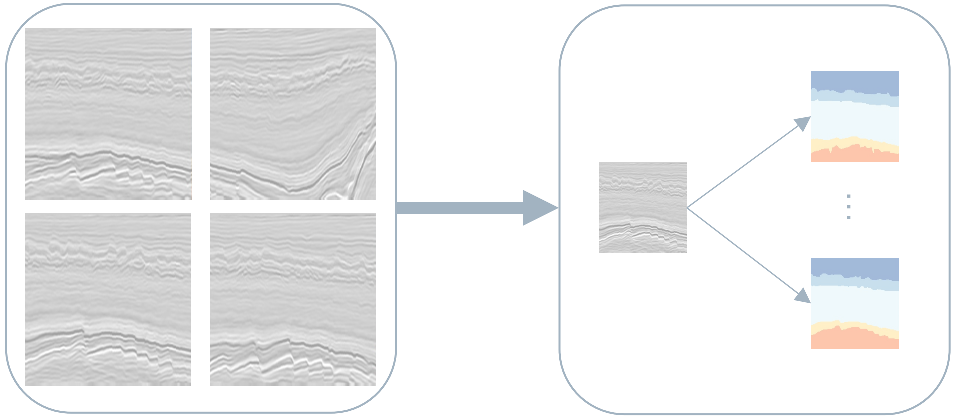


# Uncertainty in Seismic Interpretation

## The Framework Reflects the Processing Steps of the Entire Interpretation Pipeline

### Characterizing Interpretational Uncertainty

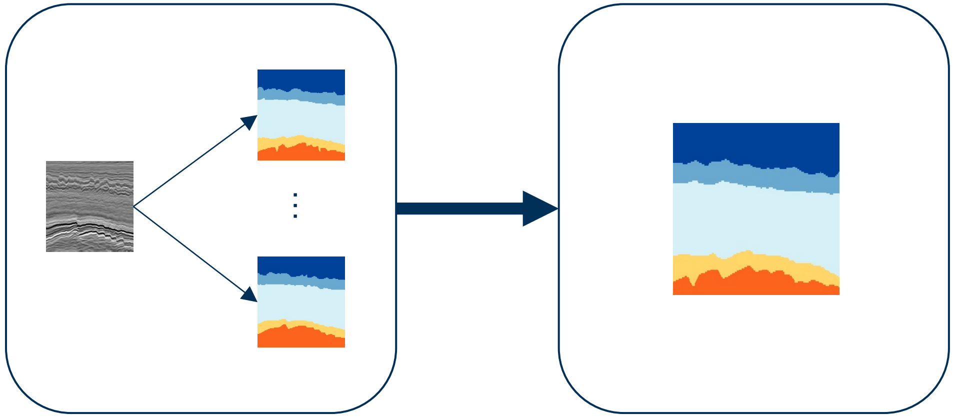
Seismic Data | Labels



Images | Interpretations

$$X \xrightarrow[p(\alpha|X)]{} \alpha$$

Labels | Geology



Interpretations | Ground Truth

$$\alpha \xrightarrow[\boldsymbol{p(Y|\alpha)}]{} Y$$

**Probability of interpretations representing the ground truth**

# Characterizing Interpretational Uncertainty

## Interpretational Uncertainty as the Probability of Interpretations Representing the Ground Truth

### Characterizing Interpretational Uncertainty

- **Uncertainty varies** w.r.t. individual interpretations:

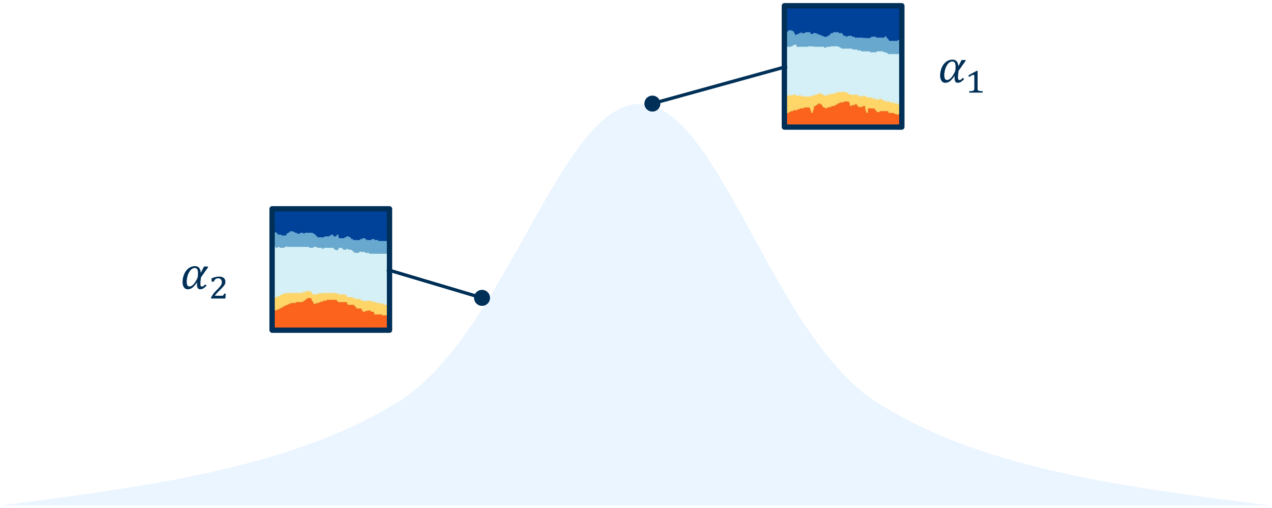
$$p(Y|\alpha_1) > p(Y|\alpha_2)$$

$Y$: ground truth, $\alpha$: interpretations

- Interpretations vary due to **expertise hierarchy**

E.g., Expert**:** $\alpha_1$, Non-expert: $\alpha_2$

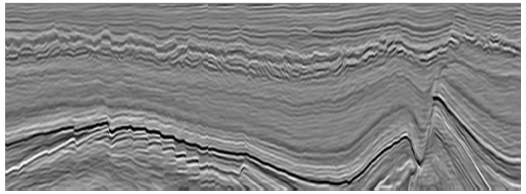
**Probability Distribution** of Interpretations

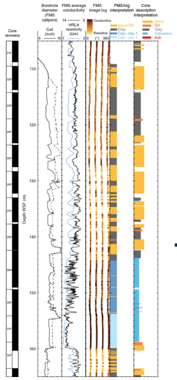


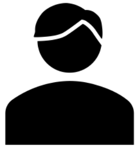
# Comparison between Interpretations from Different Levels of Expertise

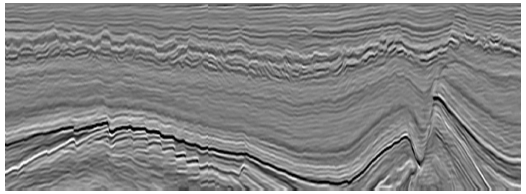
## Interpretation Requires Higher Budgets for Domain Experts than Non-experts

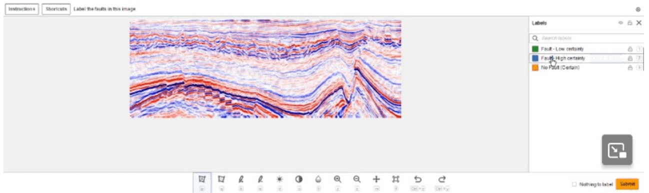
### Expert Interpretation Costs Higher than Non-expert Interpretation

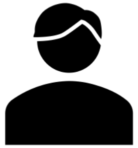

Seismic data


Well log data


Expert


Seismic data


Basic labeling instructions


Non-expert

Interpretation by **domain experts**

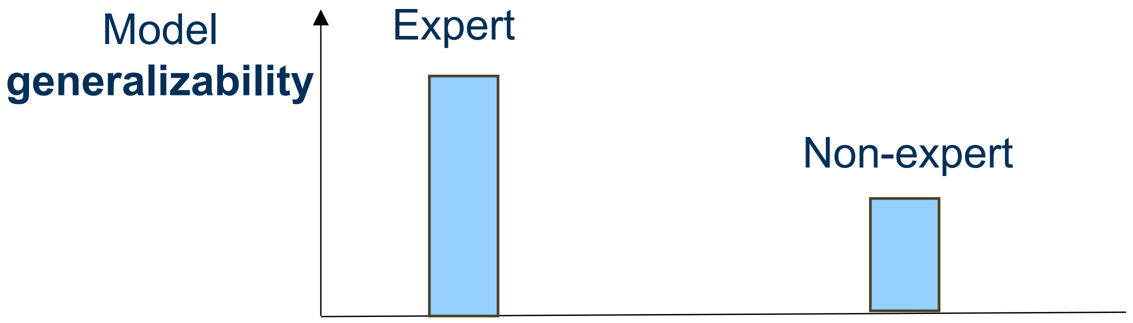
- **Expensive** and **labor intensive**

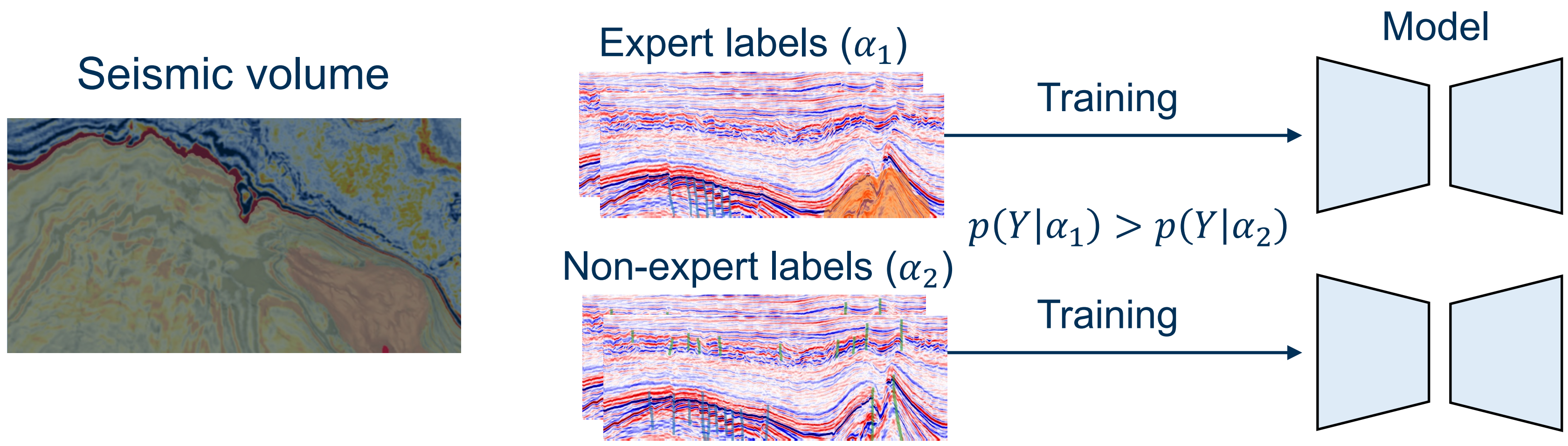
Interpretation by crowd **non-experts**

- **Less expensive**

# Comparison between Interpretations from Different Levels of Expertise
## Training with Non-expert Labels Generalizes Worse than Training with Expert Labels

**Training** with only **non-expert** labels leads to **degraded generalizability**

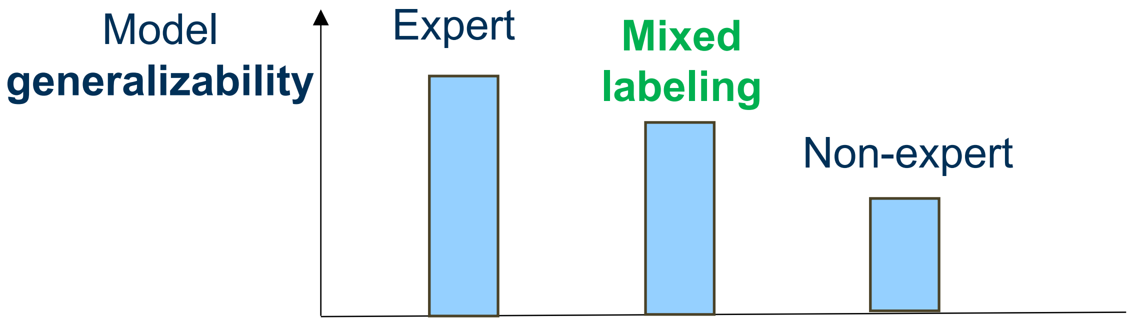


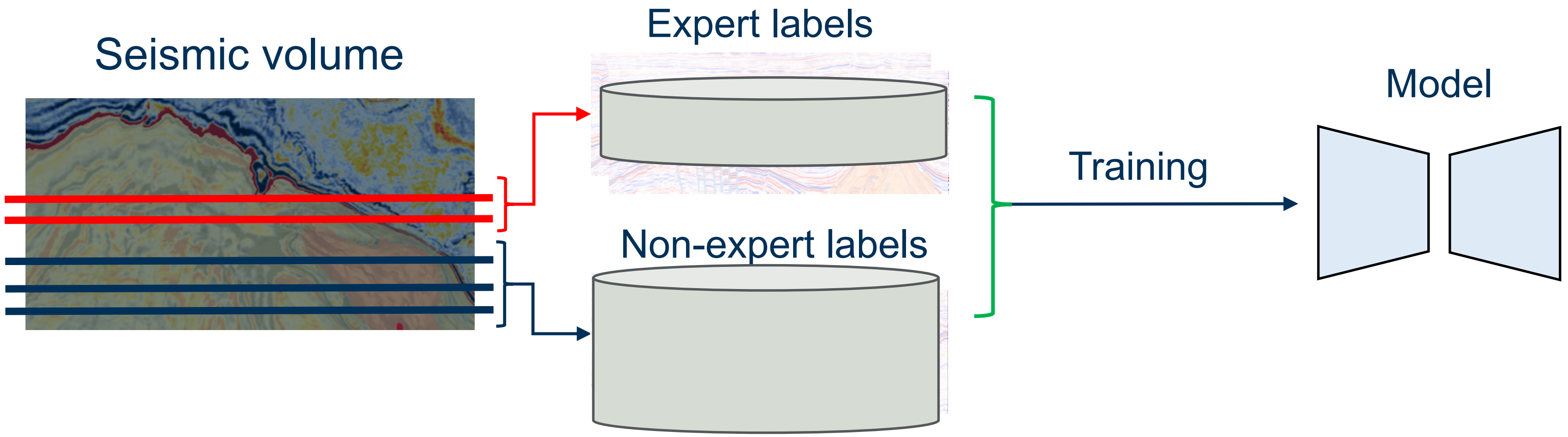



# Sample Selection for Expert Interpretation to Enhance Generalizability

## Bridging the Generalization Gap between Training with all Non-expert Labels versus Expert Labels

**Objective:** To achieve **generalization** like **expert** labels by leveraging **non-experts** on **selectively sampled** data.



## Fault Labeling Dataset with Different Expertise Levels

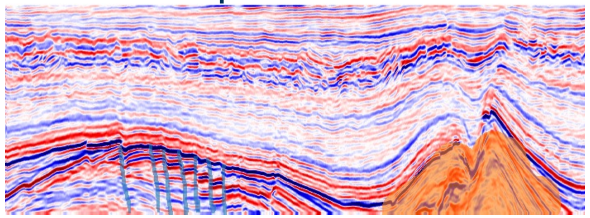
Details:

- 400 sections of F3 block
- 1 expert, 8 non-experts

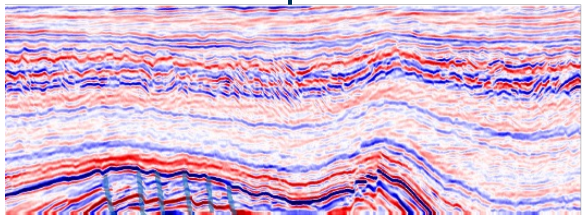
Expertise Hierarchy


Fault Labeling Examples

Expert labels



Non-expert labels



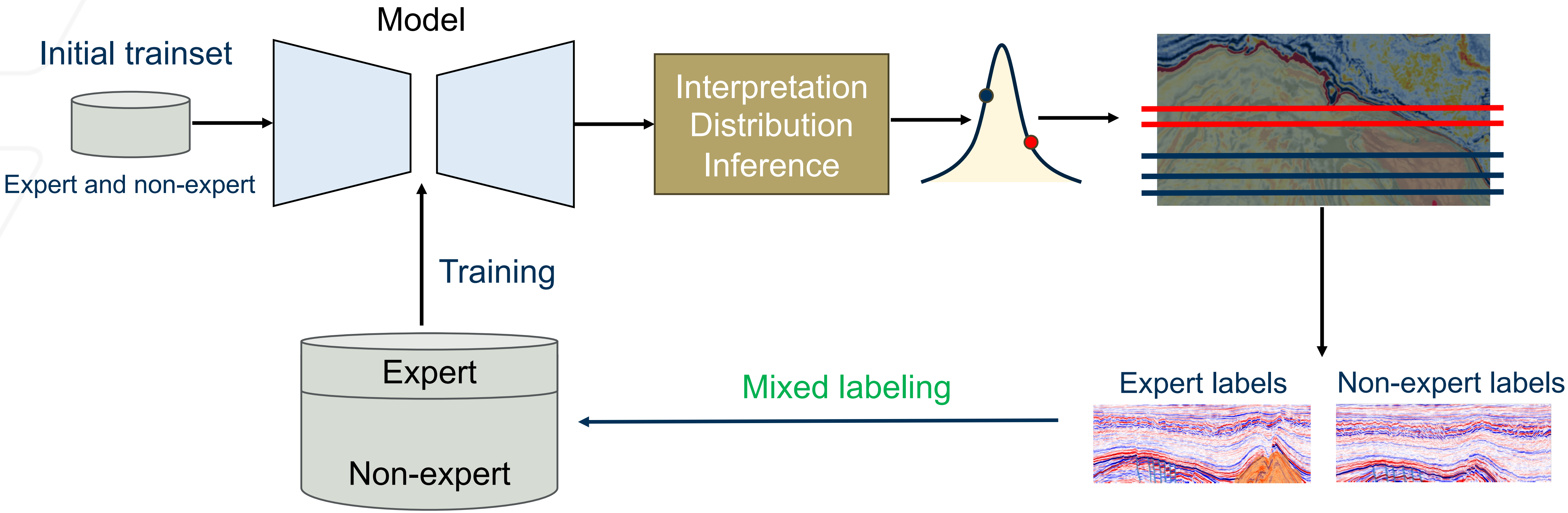


Y. Alaudah, P. Michalowicz, M. Alfarraj, G. AlRegib, "A Machine Learning Benchmark for Facies Classification," in Interpretation 2019

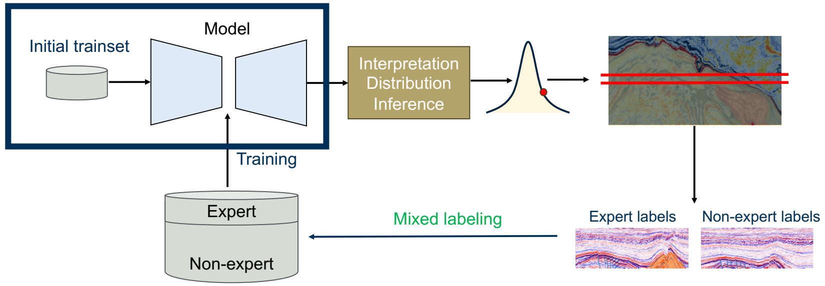
# Sample Selection for Expert Interpretation to Enhance Generalizability

## We Integrate Interpretation Distribution Estimation in Sample Selection

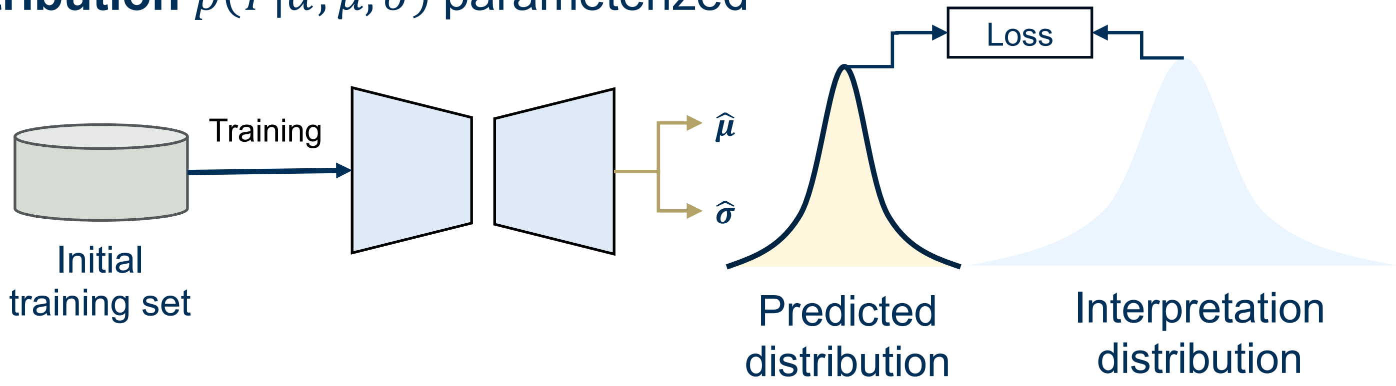
**Contributions: Integrating Interpretation Distribution Estimation in Sample Selection to Reduce Expert Efforts**



## Label Distribution Training with the Expert and Non-experts




- **Interpretation distribution** $p(Y|\alpha)$ parameterized by **labels**:
  - 1 expert, 8 non-experts
  - Assume $p(Y|\alpha_{expert})$ is the highest
- **Predicted distribution** $p(Y|\alpha;\hat{\mu},\hat{\sigma})$ parameterized by the **model**:
  - $\hat{\mu}$: mean
  - $\hat{\sigma}$: std.



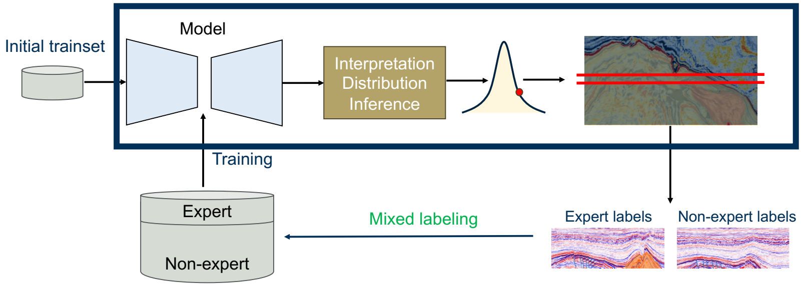

We learn a model that **estimates interpretation distribution** to characterize **interpretational uncertainty**

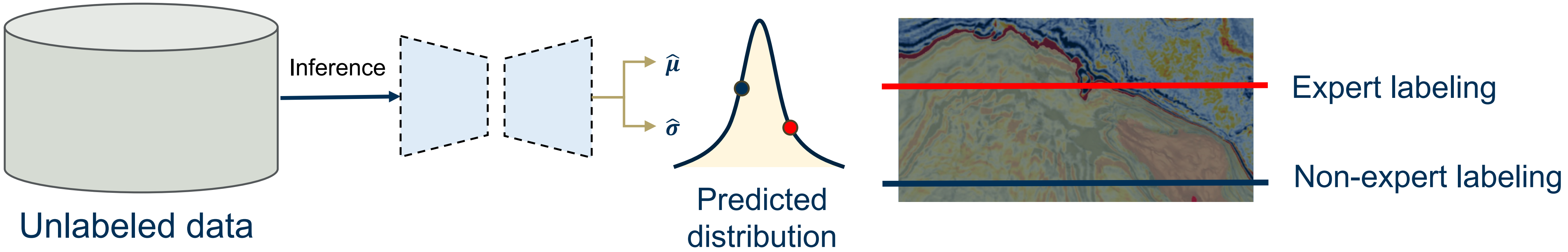
# Our Framework for Interpretation Distribution Estimation and Fault Detection

Estimating Interpretation Distribution during Sample Selection for Expert Labeling

## Estimating Interpretation Distribution during Sample Selection for Expert Labeling

- **Select** samples of which **probabilities** $p(Y|\alpha; \hat{\mu}, \hat{\sigma})$ of **non-expert's** interpretations are **low**.
- **Label** these samples by the **expert**.



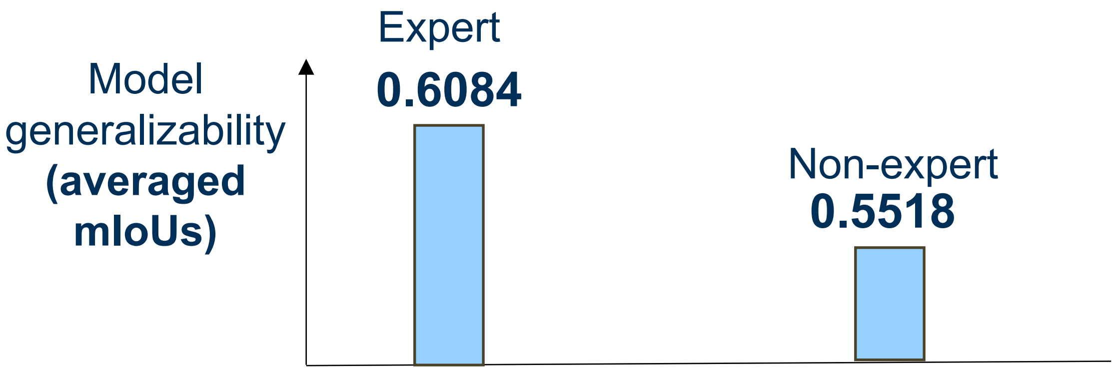




We utilize the **predicted distribution** to **select samples** for **expert labeling**

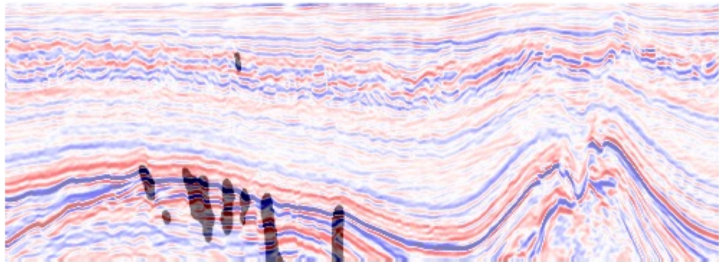
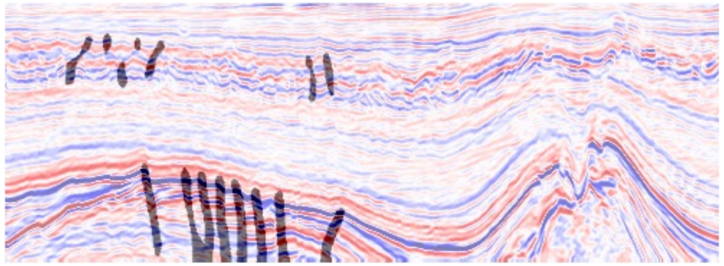
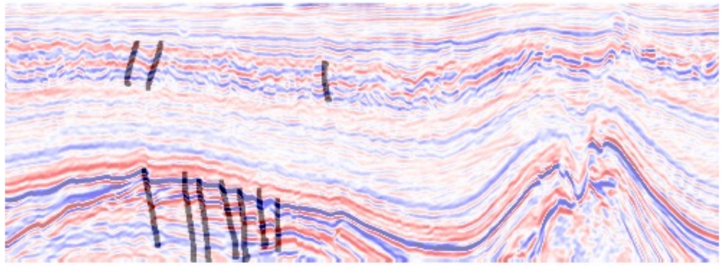
# Results

## Training with all Non-expert Labels Sows Degraded Generalizability Compared to all Expert Labels

### Training with all Expert Labels versus all Non-expert Labels



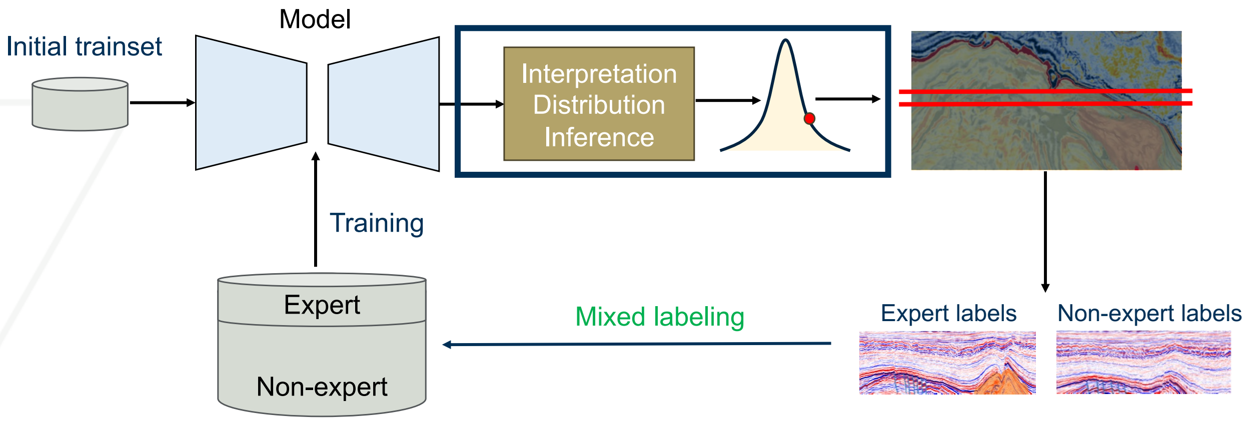

| Training with all non-expert labels | Training with all expert labels | Manual interpretation |
| --- | --- | --- |

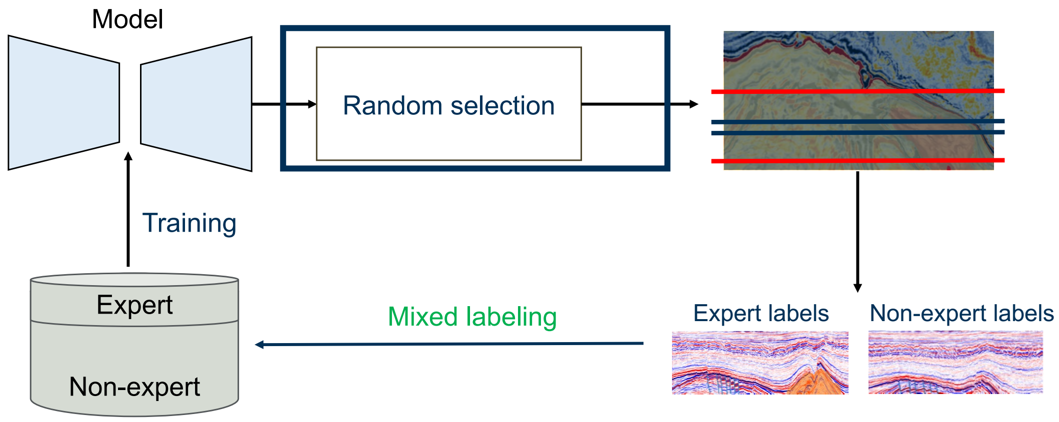




# Results

## Our Method Enhances Generalization and Outperforms the Baseline

### Comparison between Interpretation Distribution-aware and Random-based Selection



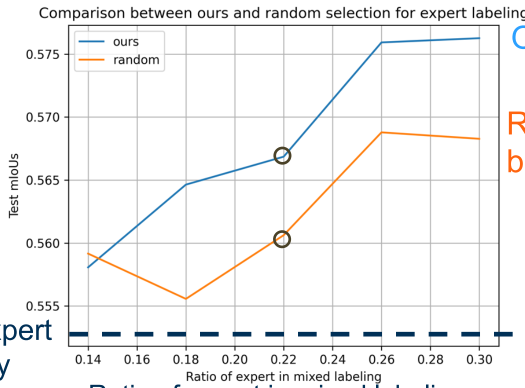



# Results

## Our Method Matches or Outperforms the Baseline

### Interpretation Distribution-aware and Random-based Expert Labeling

Test
mIoU




Ours

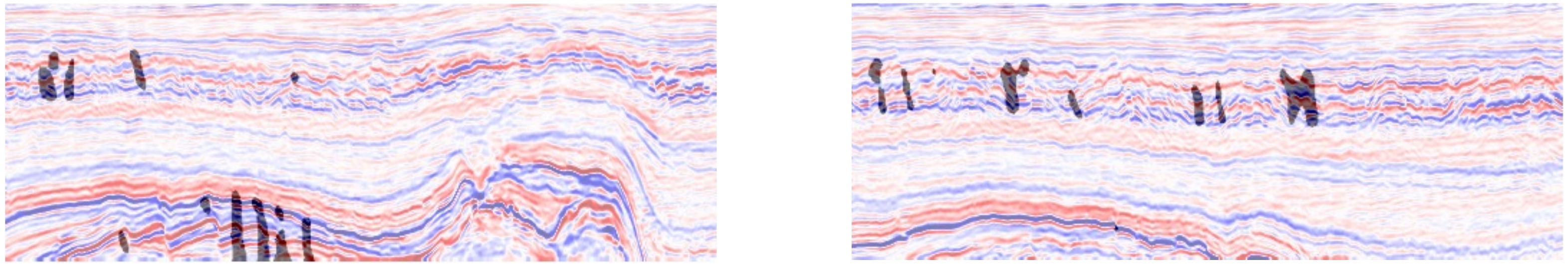
Random-
based

Non-expert
only

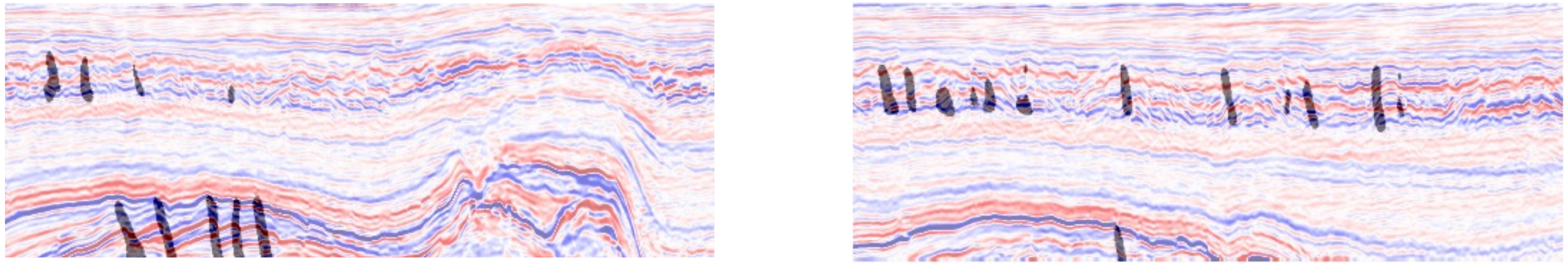
Ratio of expert in mixed labeling

22% expert
labeling

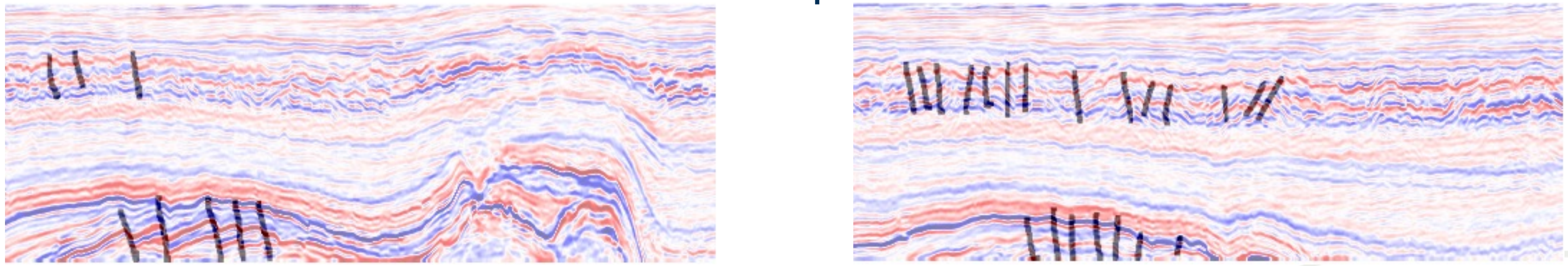
Random-based expert labeling



Interpretation distribution-aware expert labeling



Manual interpretation



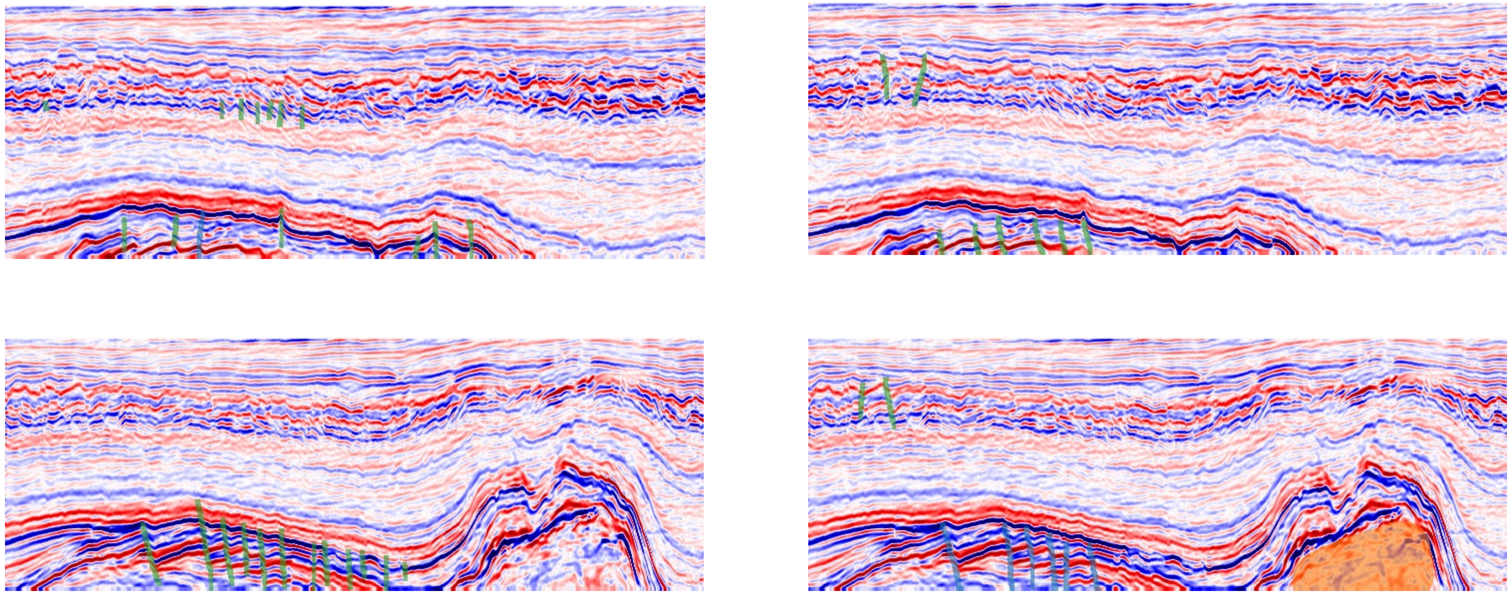
**Interpretation distribution-aware** expert labeling **matches** or **improves** over random-based labeling **baseline**.

# Qualitative Results – Ambiguous Training Sample Selection by our Framework

Sample selection

Variation of labels by non-experts

Labels by the expert



**Interpretation distribution-aware** expert labeling **select** samples with **high interpretational uncertainty**.

# Conclusion

- Training with **non-expert** versus **expert** interpretations shows significant **generalization gap**.
- We introduce an **interpretation distribution-aware** sample **selection** approach to characterize interpretational uncertainty.
- We demonstrate that the **generalization gap** can be **mitigated** by our proposed sample selection with **reduced expert labeling efforts**.

# Thanks for Listening
## Questions?

ML4SEISMIC

OLIVES @GeorgiaTech

**Publications** **Code**