Structured Output Learning for Automatic Geophysical Feature Detection

Chiyuan Zhang, Charlie Frogner, Tomaso Poggio, Mauricio Araya, Detlef Hohl

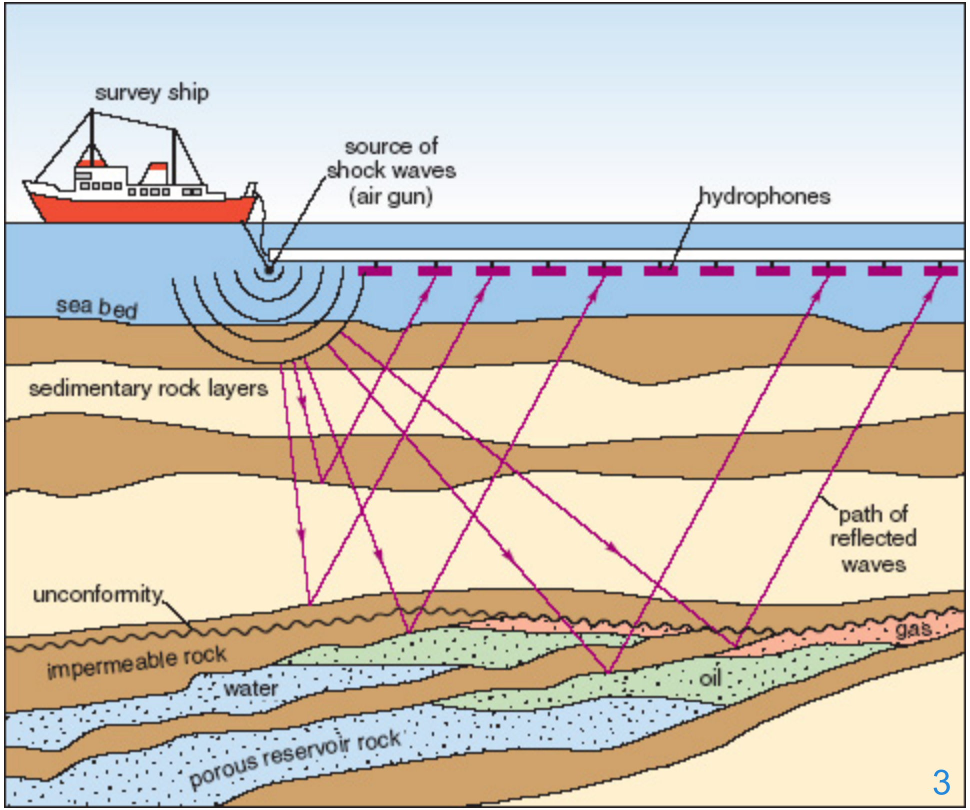
Outline

- Motivation
- Methods
- Results
- Conclusion & Outlook

Motivation: Seismic Survey

Seismic surveys are very important for discovering underground structures before deciding where to drill wells.

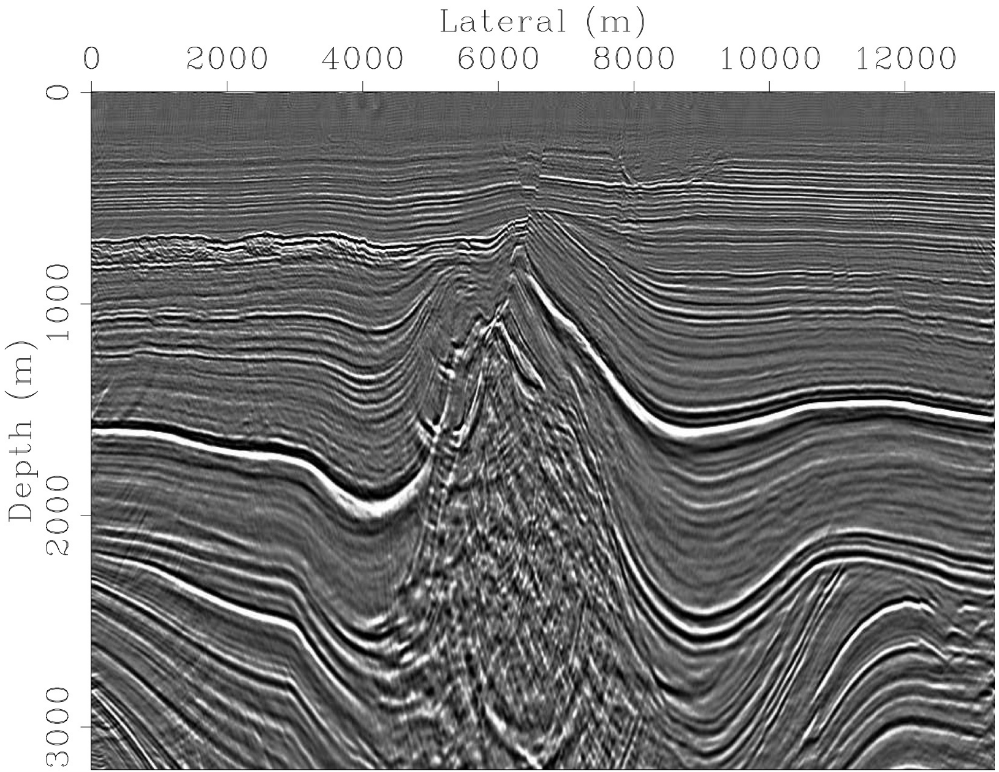
- Shock waves are generated (usually at many different places)
- The reflective waves from underground layers are recorded in an array of sensors
  - The time-series signals are called (raw) *seismic traces*



Motivation: Seismic Migration

Seismic migration uses an iterative procedure to recover the underground layerwise structure (seismic images).

- An initial prior velocity model from geologists is needed.
- Human intervention is needed during each iteration of refinement, to adjust the estimated velocity model to be more plausible/consistent with known geology, geophysics, etc.
- The whole procedure can take months to complete.



Post-Stack Depth Migration

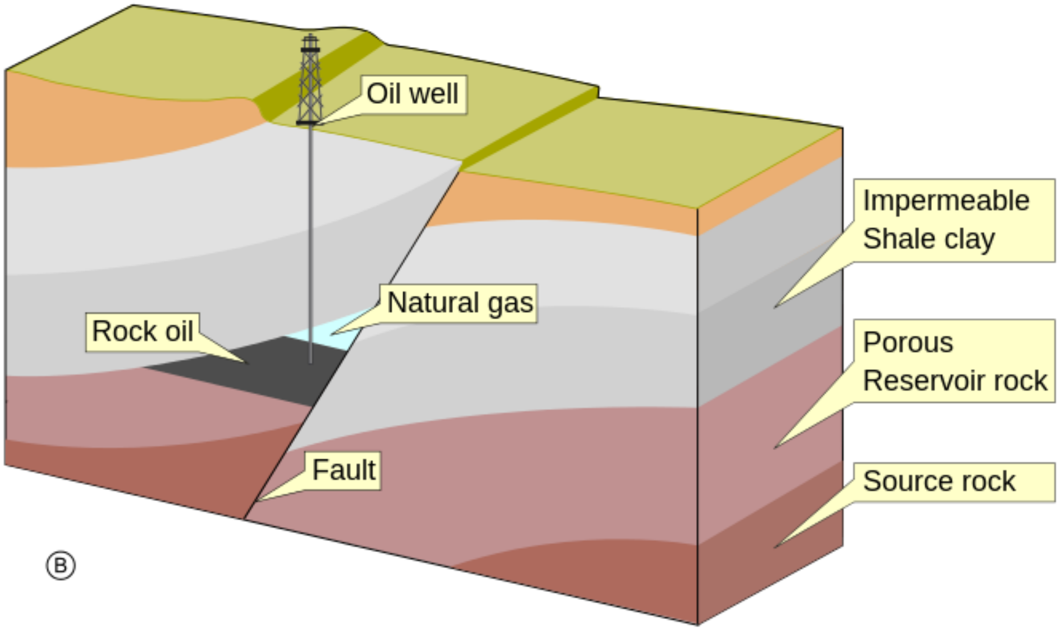
Automatic Geophysical Feature Detection

Can we bypass the costly migration step, and detect interesting geophysical features directly from the data?

Detecting Potential Traps of Oil/Gas

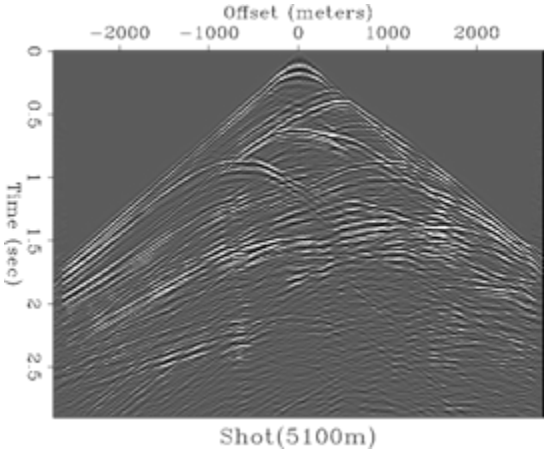
Common **structural traps** include anticlinal trap, **fault trap**, and salt dome trap.

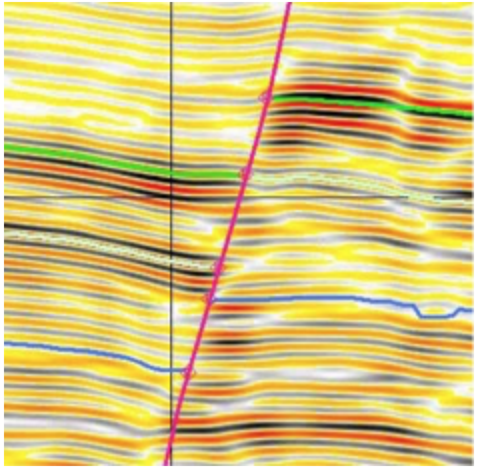
These traps block the upward migration of hydrocarbons and can lead to the formation of a **petroleum reservoir**.



https://en.wikipedia.org/wiki/Structural_trap

Current Goal: Fault Detection



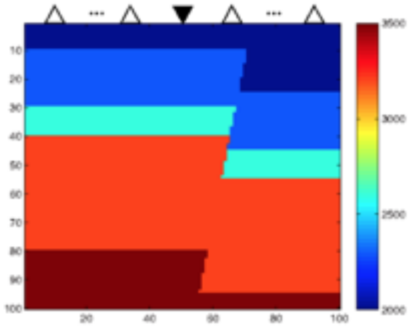


From raw **seismic traces**, discover (classification) and locate (structured prediction) faults in the underground structure, **without** running migration.

Machine Learning based Fault Detection

- Cast fault-detection as a machine learning problem
- Training data
    - Human labeled faults, acquired using **migrated** seismic images, along with corresponding raw seismic traces.
    - Synthetic data
        - Generate random velocity models.
        - Simulate seismic data for these models, using a finite difference approximation to the acoustic wave equations.

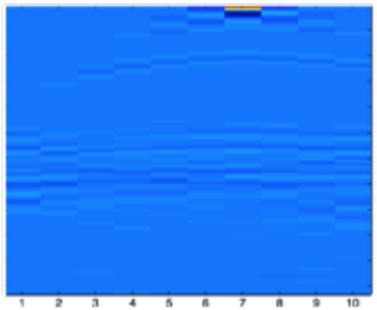
Workflow Overview



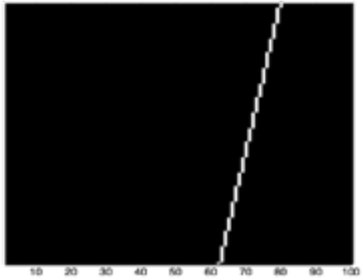
velocity model (latent, known only during data generation)

wave-equation simulation

seismic traces



Fault location (ground-truth)

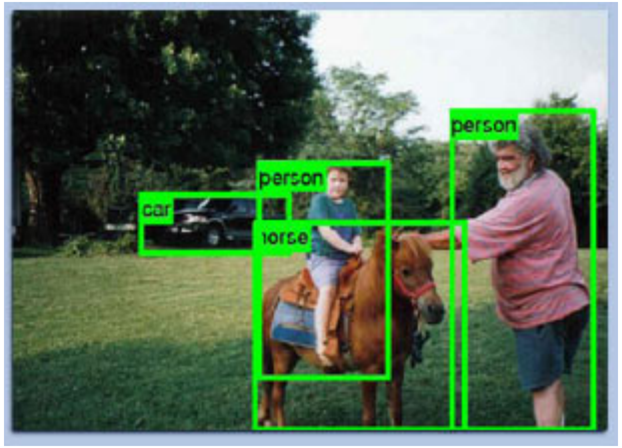


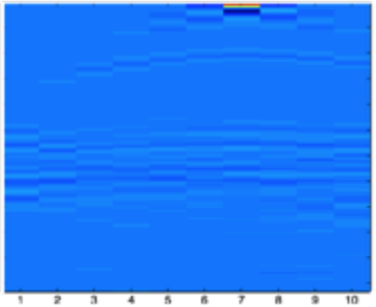
Learn a model to predict location of a fault from seismic traces.

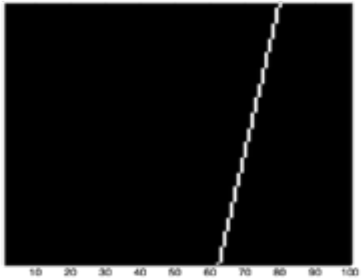
Difference from Detection in Computer Vision

Unknown correspondence between input and output domain

- CV: pixel ⇔ pixel
- Fault detection
  - Input: Time-by-Sensor (1000x10)
  - Output: Space-by-space (e.g. 100x100)
  - Correspondence depends on unknown velocity model

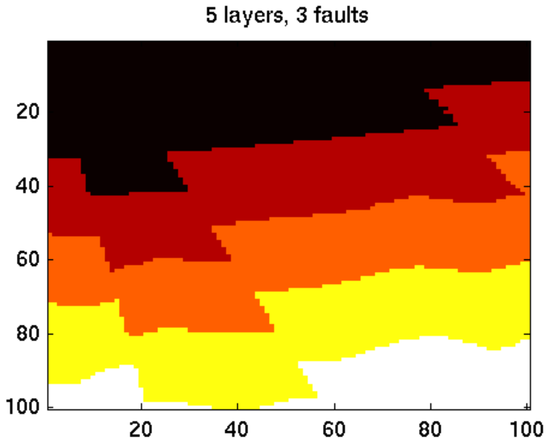




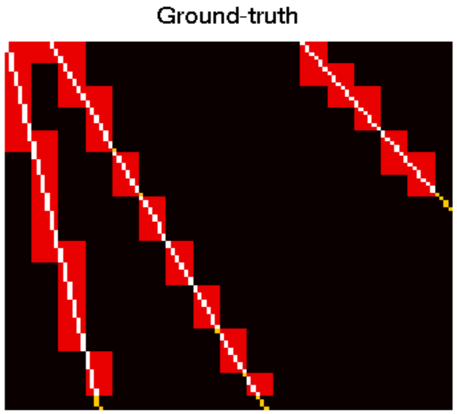


Problem Formulation

A grid of binary fault PRESENT/NOT regions



Velocity model (unknown even during training)



Label (fault) representation, 2D "pixel" map

Learning to predict a binary bit map - each pixel is "on" if a fault crosses the corresponding spatial region.

Similar to semantic segmentation in Computer Vision, but no easy pixel correspondence between input and output.

Wasserstein Distance

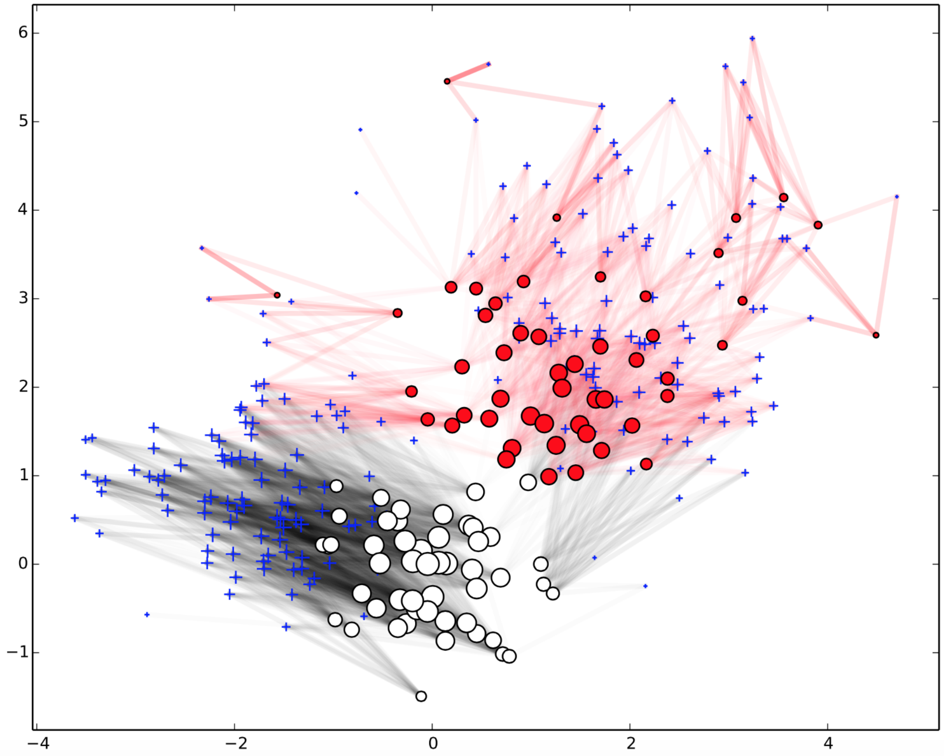


image source: http://remi.flamary.com/biblio/courty2014domain.pdf

Total cost of the optimal transport plan from the source (prediction) distribution to the target (ground truth) distribution. A.k.a. Earth Mover's Distance.

Transport cost computed with respect to an underlying *ground metric*. In contrast, standard divergence-based or L^p distance, or hamming distance ignore the ground metric.

Wasserstein Distance

Primal LP

$$W_p^p(h(\cdot|x), y(\cdot)) = \inf_{T \in \Pi(h(x), y)} \langle T, M \rangle$$

$$\Pi(h(x), y) = \{T \in \mathbb{R}_+^{K \times K} : T\mathbf{1} = h(x),\ T^\top \mathbf{1} = y\}$$

Dual LP

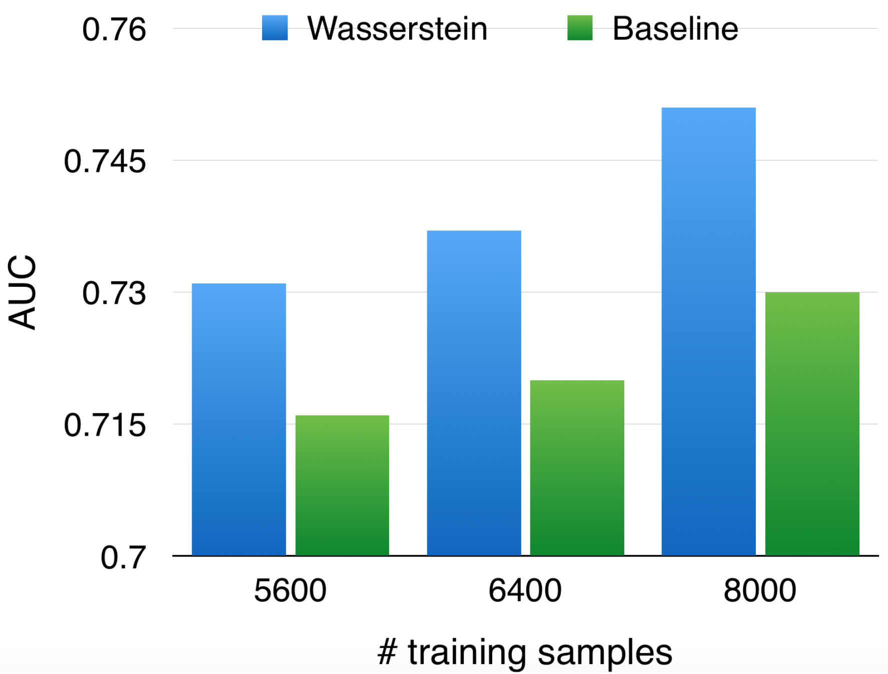
$${}^{d}W_p^p(h(x), y) = \sup_{\alpha, \beta \in C_M} \alpha^\top h(x) + \beta^\top y$$

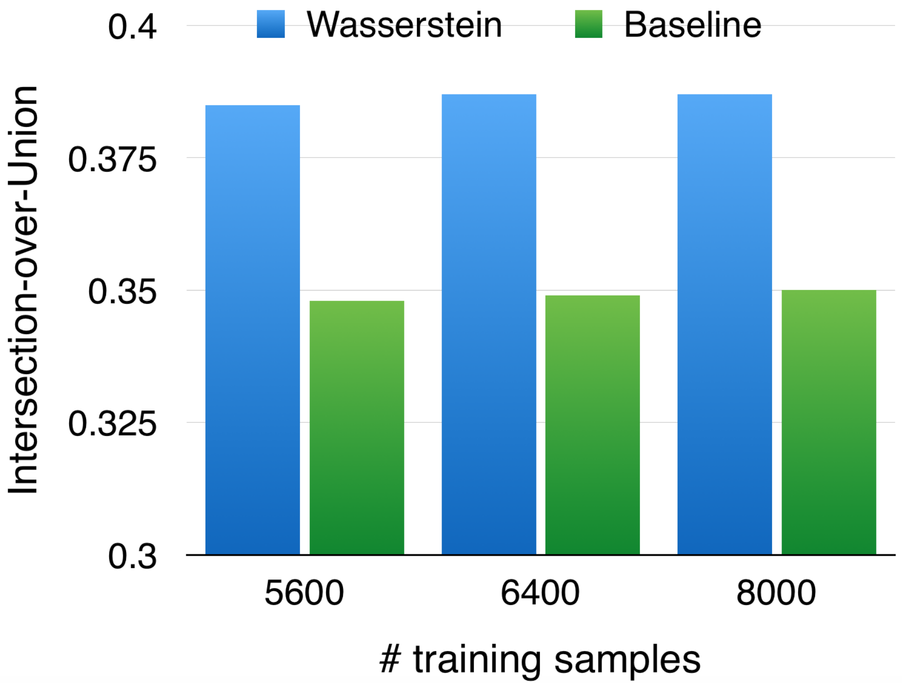
$$C_M = \{(\alpha, \beta) \in \mathbb{R}^{K \times K} : \alpha_\kappa + \beta_{\kappa'} \leq M_{\kappa, \kappa'}\}$$

Learning with Wasserstein Loss

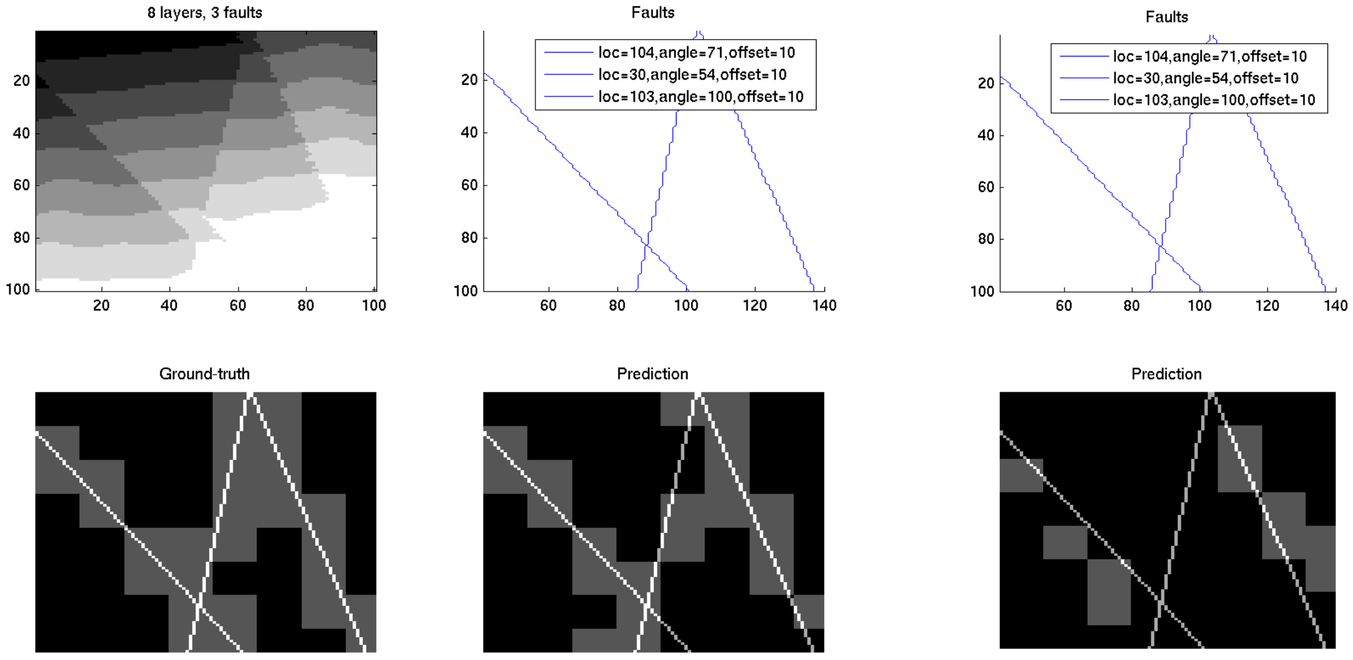
- Non-decomposable loss, penalize mis-predictions that are “far away” from groundtruth.
- Dual formulation: gradient given by the dual solution, back-propagate into model parameters via chain-rule.
- Fast computation: Sinkhorn iteration [MC13] or other matrix scaling algorithms [FZMAP15].

Empirical Performance





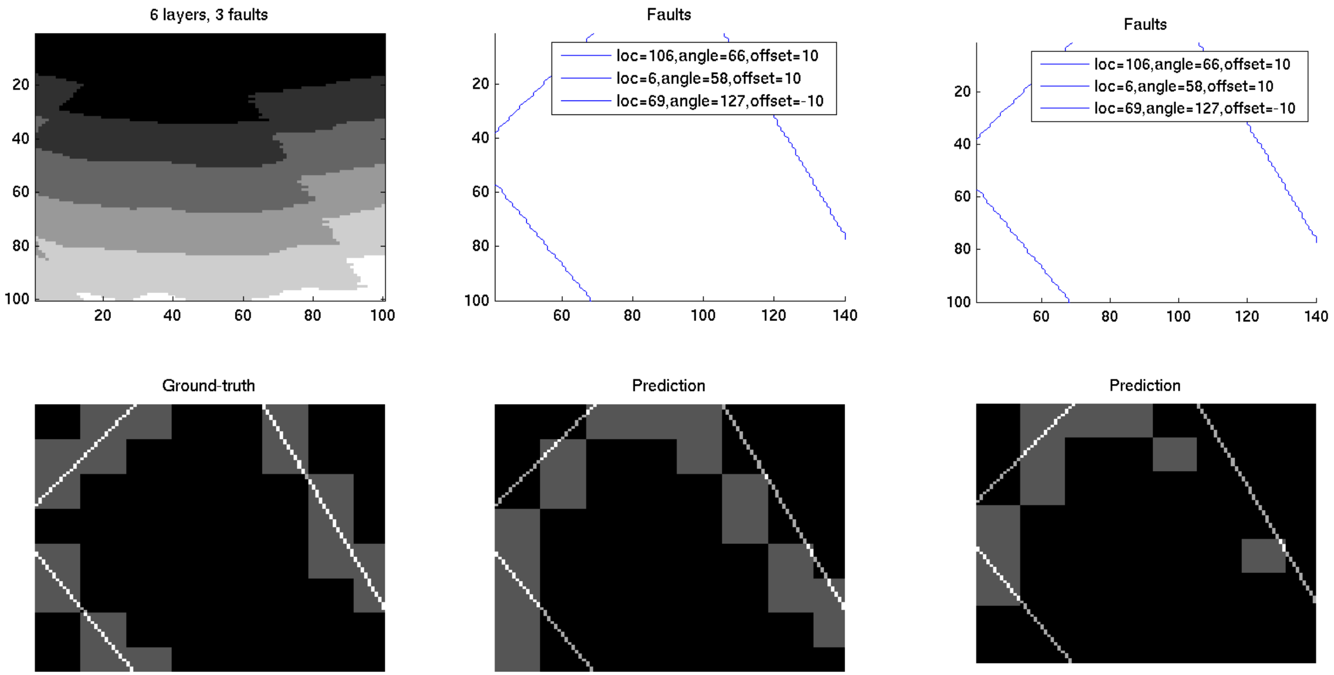
Visualization



Wasserstein

Baseline

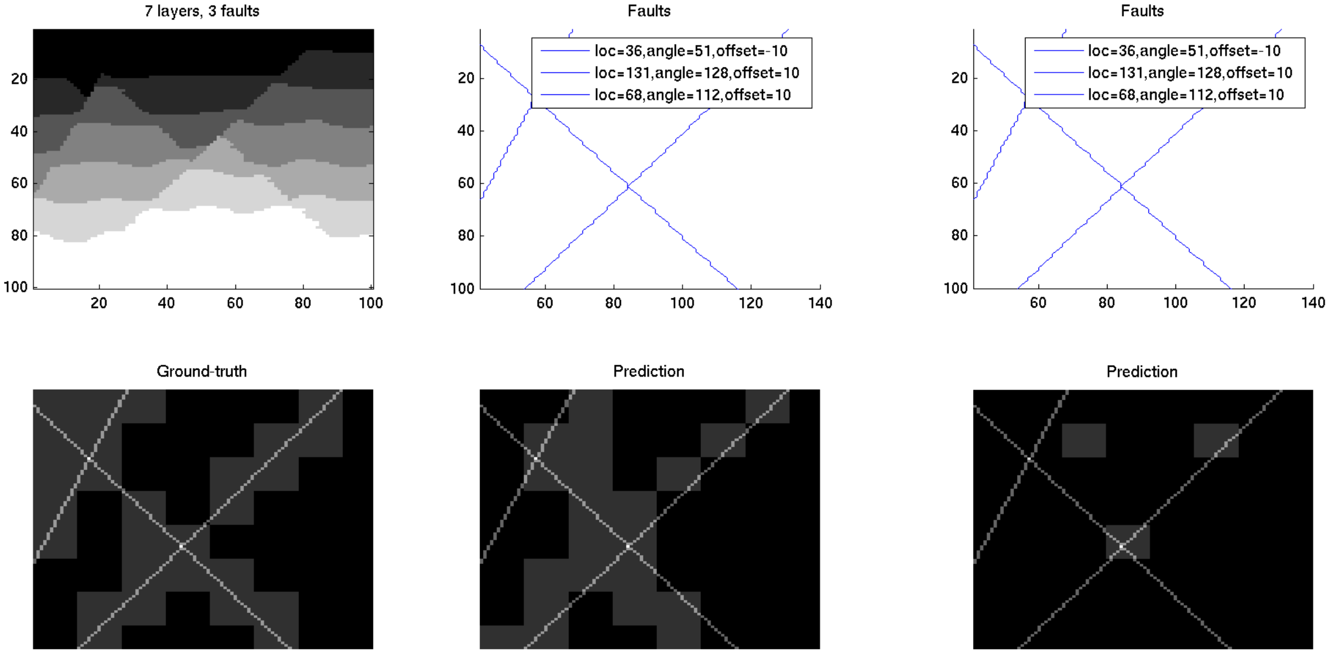
Visualization



Wasserstein

Baseline

Visualization



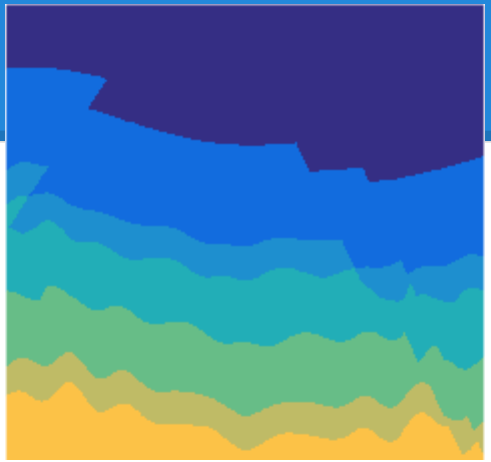
Wasserstein

Baseline

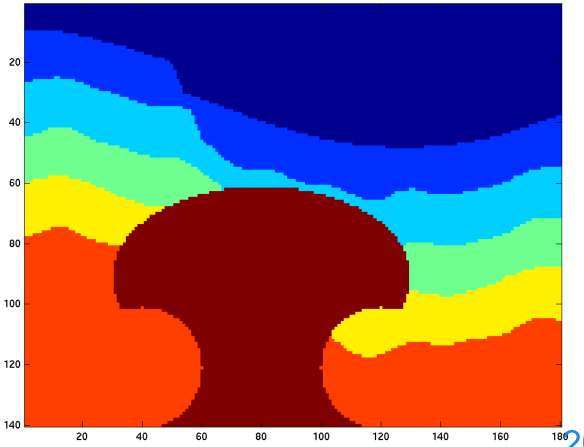
Conclusion

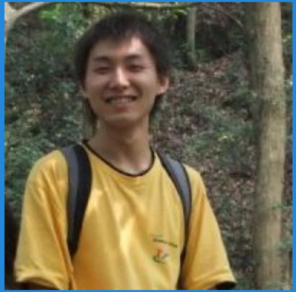
- Automatic geophysical feature detection, directly from seismic data, is a groundbreaking and cost-reducing approach.
- Can be formulated as a structured output prediction problem, but unlike many standard structured prediction problems, there’s no direct input-output mapping.
- Preliminary experiments show promising results.

Outlook



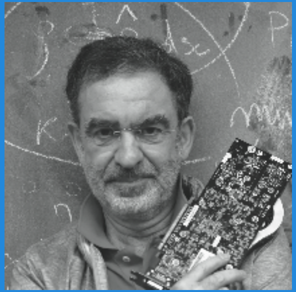
- More realistic velocity models
  - Partial, 3D models, salt domes, real data
- More advanced structured prediction algorithms
  - High-order priors: faults tend to be “linear” structures
- Prediction of other geophysical features





Chiyuan Zhang

Charlie Frogner



Tomaso Poggio

Mauricio Araya

Detlef Hohl

Massachusetts Institute of Technology

Thank you!