Introduction inaugural ML4Seismic Partners Meeting

Felix J. Herrmann\textsuperscript{1,2,3}

November 22, 2021

Released to public domain under Creative Commons license type BY (https://creativecommons.org/licenses/by/4.0)
Copyright (c) 2021, Felix J. Herrmann (Georgia Tech)
Wellcome
Center for Machine Learning for Seismic Industry Partners Program is a

- public-private partnership between Georgia Institute of Technology, Oil & Gas industry & Cloud providers supported by Microsoft, Chevron, Equinor & Occidental
- collaboration between
  - Dr. Ghassan AlRegib (ECE) – Omni Lab for Intelligent Visual Engineering & Science (OLIVES)
  - Dr. Felix J. Herrmann (EAS/CSE/ECE) – Seismic Laboratory for Imaging & Modeling (SLIM)
- launched early 2020

During inaugural meeting

- report research progress & demo our open source software
- platform for feedback & discussion on future research directions
Who is here

Excited to welcome 124 registered participants

- 24 from ML4Seismic industry partners
  (Chevron, Equinor, Microsoft, Occidental)

- 100 from industry & academia
  (AWS, BGP, BHP, BP, CGG, Emerson, Extreme Scale Solutions, ExxonMobil, Haliburton, IHS Markit, ION, Osokey, Petrobras, PGS, RepsolSinopec, Saudi Aramco, Shell, Schlumberger, TGS, TotalEnergies, Woodside)

- 25 companies & several international universities
Background
Collaboration

**OLIVES:**
- machine learning, image processing & computer vision
- robust learning & learning w/ limited labels
- open source datasets & software for object recognition, facies classification, etc.

**SLIM:**
- wave-based imaging/monitoring, UQ & machine learning
- variational inference w/ normalizing flows & ML at scale
- open source datasets & software for FWI/RTM, UQ, CCS monitoring, etc.

Responsible for driving innovations & reducing costs…
Georgia Institute is a powerhouse in

- engineering
- machine learning

Combination OLIVES & SLIM

- complementary areas of expertise
- innovate w/i the intersection of seismic imaging/monitoring & machine learning
- train students in these areas

Transdisciplinary approach essential to navigate the energy transition…
Mission

Train next-generation of students to be well versed in

‣ machine learning & data science
‣ scientific computing
‣ specific domain knowledge for seismic imaging & quantitative interpretation

Seek balance between

‣ gaining fundamental understanding
‣ driving / accelerating innovations w/ tangible solutions to real problems
Dr. Ghassan AlRegib is recognized as a world-leader in

- algorithm development for digital image processing
- interactive seismic interpretation
- and is the recipient of
  - 2008 ECE Outstanding Junior Faculty Member Award
  - 2017 Steven A. Denning Global Engagement Award
  - best Paper Award, IEEE ICIP 2019 (out of 960 papers)
Recognition

Dr. Felix J. Herrmann is recognized as a world-leader in

- algorithm development for wave-based inversion
- (time-lapse) seismic data acquisition & wavefield reconstruction
- is the recipient of
  - 2019 SEG Distinguished lecturer “Sometimes it pays to be cheap – Compressive time-lapse seismic data acquisition”
  - 2020 Fessenden award
The Leading Edge®

The Leading Edge®

The Leading Edge®

The Leading Edge®

imageing
w/ multiples

FWI
w/ constraints

acquisition
w/ Compressive Sensing
ML4Seismic partnership
Aims to

- drive innovations in artificial-intelligence assisted seismic imaging, interpretation, analysis & monitoring in the cloud
- low-environmental impact time-lapse acquisition & monitoring
- data-constrained image segmentation & classification
- data-driven & physics-constrained machine learning & uncertainty quantification
- 2 programs in one

Philanthropic research directed gift program carries *no* overhead (opposed to 64.6% overhead for sponsored research* at Gatech)

*https://osp.gatech.edu/rates
Benefits

Charter Partner at $180k or more (Annual):

- possibilities for recruitment, internships, engagement w/ students & faculty
- up to 5 complementary registrations Fall & Spring meetings
- opportunities to engage w/ students and faculty on research projects
- access to reproduce our results via creative commons & open source licenses
- 2 complementary registrations professional training
- 1 x a year information session to discuss new research directions
- invitation of unto 5 company reps to attend ML4Seismic hackatons
- access to Beta testing ML4Seismic’s open source software
- recognition as Diamond sponsor
Executive Partner at $90-179k (Annual):

- possibilities for recruitment, internships, engagement w/ students & faculty
- up to 3 complementary registrations Fall & Spring meetings
- opportunities to engage w/ students and faculty on research projects
- access to reproduce our results via creative commons & open source licenses
- 1 complementary registrations professional training
- 1 × a year information session to discuss new research directions
- invitation of unto 5 company reps to attend ML4Seismic hackatons
- access to Beta testing ML4Seismic’s open-source software
- recognition as Platinum sponsor
Two-tier program

Vetted by Georgia Tech to protect its Tax exempt status

- "Under no conditions will any particular RESEARCH RESULTS be linked, or given the perception of being linked to a specific donation by any individual PARTICIPANT. ” *

- no explicit deliverables

Most importantly you have direct access to our teams!

Option to give input to the research program...

* from OPERATING GUIDELINES Center for Machine Learning for Seismic Industry Partners Program (ML4Seismic)
Gifts from ML4Seismic partners

Your gifts are for 100% used to

- support our students (stipend, tuition, etc.)
- fund our post doctoral fellows
- pay for travel
- gain access to Cloud compute
- budget permitting free-up director’s time (summer salary & teaching release)
Open source model
advantages

New model:

- driving rapid innovations as in Machine Learning & AI
- Chevron, Total, & Equinor are releasing some of their codes as open source
- undergirds more & more public-private partnerships
- initiatives towards open-source endowed professorships

Northern Lights CCS Program (Equinor, Microsoft, Shell, Total):

- all codes released as open source
- part of the Open Group OSDU
Research findings are made available through

- publications in peer reviewed journals
- during presentations at conferences
- during presentations at ML4Seismic events (Creative Commons License)
- made reproducible open source software on GitHub and Zenodo (MIT License)

No formal IP

- agreement much simpler
- geared to rapidly drive innovations
- widely used model by US universities

Non-negotiable OPERATING GUIDELINES are available on request.
Repositories – slimgroup
open MIT licensed code on GitHub

General purpose open-source software (OSS) packages

- [JUDI.jl](#) – the Julia Devito Inversion framework
- [SetIntersectionProjection.jl](#) – projections onto intersections & sums of sets
- [InvertibleNetworks.jl](#) – building blocks for invertible neural networks
- [JOLI.jl](#) – serial and distributed linear operators in Julia

Specialized examples

- [FastApproximateInference.jl](#) – variational inference for inverse problems
- [XConv](#) – Julia/Python code for memory efficient CNNs
- [TimeProbeSeismic.jl](#) – low memory WE based inversion
Repositories – OLIVES on GitHub

Open-source software (OSS) packages

- **ISI** – Salt Dome Interpretation Tool
- **S3I** – seismic imaging, survey, and imaging
- **dippykit** – a Python Library for Digital Image Processing

Open source benchmark datasets

- **LANDMASS** – North-Sea Dataset of Migrated Aggregated Seismic Structures
- **F3Facies** – A Machine Learning Benchmark for Facies Classification
- **CoMMonS** – Challenging Microscopic Material Surface Dataset
Value proposition

Energy transition is having a major impact on Oil & Gas industry

- reduced research & development capacity
- increased demand for technology (e.g. CCS)

Public-private partnerships w/ universities

- lowers costs & increase rate of innovation
- access to next-generation of talent
- success hinges on access to training datasets & compute

Will allow us to tackle important problems & to become part of the solution...
This meeting
Formal plenary sessions
talks recorded / Q&A not recorded

Monday:

- **Deployment of Active Machine Learning**
  (Chair: Ali Siahkoohi)

- **Seismic Imaging and Processing in the Cloud**
  (Chair: Ahmad Mustafa)

Tuesday:

- **Explainability Paradigms in AI**
  (Chair: Mohit Prabhusankar)

- **Seismic Acquisition, Inversion, and Monitoring**
  (Chair: Rafael Orozco)
Informal parallel sessions
not recorded

Mon/Tuesday:

- **Breakout 1. Active Machine Learning, Explainability, and Uncertainty**
  (Chair: Ghassan AlRegib)

- **Breakout 2. Scalable Software in the Cloud**
  (Chair: Mathias Louboutin)

- **Breakout 3. Seismic Imaging & CCS Monitoring w/ Uncertainty Quantification**
  (Chair: Felix J. Herrmann)

Please stay in Teams meeting so you can be assigned to breakout room!
Hands-on tutorials

https://ml4shub.eastus.cloudapp.azure.com/
Program – Monday 22 November
### Plenary sessions

#### Monday

<table>
<thead>
<tr>
<th>Time</th>
<th>Speaker</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00—09:25 AM</td>
<td>Felix J. Herrmann</td>
<td>Introduction to the Meeting</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Deployment of Active Machine Learning (Chair: Ali Siahkoohi)</strong></td>
</tr>
<tr>
<td>09:25—09:50 AM</td>
<td>Ahmad Mustafa</td>
<td>Human in the Loop: Seismic Interpretation through Active Learning</td>
</tr>
<tr>
<td>09:50—10:15 AM</td>
<td>Ryan Benkert</td>
<td>The Value of Learning Dynamics in Seismic Interpretation</td>
</tr>
<tr>
<td>10:15—10:30 AM</td>
<td></td>
<td><strong>Coffee Break</strong></td>
</tr>
<tr>
<td>10:30—10:55 AM</td>
<td>Philipp Witte</td>
<td>Redwood - Towards clusterless supercomputing in the cloud</td>
</tr>
<tr>
<td>10:55—11:20 AM</td>
<td>Mathias Louboutin</td>
<td>ML4Seismic Open Source Software environment</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Seismic Imaging and Processing in the Cloud (Chair: Ahmad Mustafa)</strong></td>
</tr>
<tr>
<td>11:20—11:45 AM</td>
<td>Ali Siahkoohi</td>
<td>Uncertainty quantification in imaging and automatic horizon tracking – a Bayesian deep-prior based approach</td>
</tr>
<tr>
<td>11:45—12:10 PM</td>
<td>Ali Siahkoohi</td>
<td>Multifidelity conditional normalizing flows for physics-guided Bayesian inference</td>
</tr>
<tr>
<td>12:10—12:35 PM</td>
<td>Ryan Benkert</td>
<td>Underspecification in Seismic Interpretation</td>
</tr>
<tr>
<td>12:35—12:45 PM</td>
<td></td>
<td><strong>Discussion</strong></td>
</tr>
<tr>
<td>02:00—05:00 PM</td>
<td></td>
<td><strong>Informal sessions</strong></td>
</tr>
<tr>
<td>Time</td>
<td>Speaker</td>
<td>Topic</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>02:00—05:00 PM</td>
<td>Ghassan AlRegib</td>
<td>Breakout 1. Active Machine Learning and Explainability, and Uncertainty</td>
</tr>
<tr>
<td>02:00—03:00 PM</td>
<td></td>
<td>Human-in-the-Loop Systems</td>
</tr>
<tr>
<td>03:00—04:00 PM</td>
<td></td>
<td>Explainability and Uncertainty</td>
</tr>
<tr>
<td>04:00—05:00 PM</td>
<td></td>
<td>Trust in Seismic Interpretation Models</td>
</tr>
<tr>
<td>02:00—02:30 PM</td>
<td>Mathias Louboutin</td>
<td>Breakout 2. Scalable Software in the Cloud</td>
</tr>
<tr>
<td>02:30—03:15 PM</td>
<td>Mathias Louboutin</td>
<td>Walkthrough Julia installation</td>
</tr>
<tr>
<td>03:15—04:00 PM</td>
<td>Mathias Louboutin</td>
<td>Basics wave-based inversion (SegyIO, JUDI, constraints)</td>
</tr>
<tr>
<td>04:00—05:00 PM</td>
<td>Mathias Louboutin</td>
<td>Automatic differentiation (ChainRules, JOLI, JUDI)</td>
</tr>
<tr>
<td>04:00—05:00 PM</td>
<td>Mathias Louboutin</td>
<td>Serveless HPC in the Cloud (Devito and JUDI4Cloud)</td>
</tr>
<tr>
<td>02:00—02:30 PM</td>
<td>Ali Siahkoohi</td>
<td>Breakout 3. Seismic Imaging &amp; CCS Monitoring w/ Uncertainty Quantification</td>
</tr>
<tr>
<td>02:30—03:15 PM</td>
<td>Ali Siahkoohi</td>
<td>Background sampling from the posterior distributions via MCMC</td>
</tr>
<tr>
<td>03:15—04:00 PM</td>
<td>Rafael Orozco</td>
<td>Background (conditional) normalizing flows</td>
</tr>
<tr>
<td>04:00—04:30 PM</td>
<td>Felix J. Herrmann</td>
<td>Training of normalizing flows</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Future research directions</td>
</tr>
</tbody>
</table>
Program – Tuesday 23 November
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:50—09:00 AM</td>
<td><strong>Sign in on virtual event</strong></td>
</tr>
<tr>
<td>09:00—09:25 AM</td>
<td><strong>Explainability Paradigms in AI (Chair: Mohit Prabhushankar)</strong></td>
</tr>
<tr>
<td>Rafael Orozco</td>
<td>Variational inference for artifact removal of adjoint solutions in photoacoustic inverse problems</td>
</tr>
<tr>
<td>09:25—09:50 AM</td>
<td>Mohit Prabhushankar</td>
</tr>
<tr>
<td>Contrastive explanations and robustness for recognition in data</td>
<td></td>
</tr>
<tr>
<td>09:50—10:15 AM</td>
<td>Ahmad Mustafa</td>
</tr>
<tr>
<td>Making Black-boxes Transparent through Explainable and Interpretable Machine Learning</td>
<td></td>
</tr>
<tr>
<td>10:15—10:30 AM</td>
<td><strong>Coffee Break</strong></td>
</tr>
<tr>
<td>10:30—10:55 AM</td>
<td><strong>Seismic Acquisition, Inversion, and Monitoring (Chair: Rafael Orozco)</strong></td>
</tr>
<tr>
<td>Yijun Zhang</td>
<td>Improved seismic survey design by maximizing the spectral gap with global optimization</td>
</tr>
<tr>
<td>10:55—11:20 AM</td>
<td>Gabrio Rizzuti</td>
</tr>
<tr>
<td>A dual formulation of wavefield reconstruction inversion for large-scale seismic inversion</td>
<td></td>
</tr>
<tr>
<td>11:20—11:45 AM</td>
<td>Ziyi (Francis) Yin</td>
</tr>
<tr>
<td>Improved seismic monitoring of CO\textsubscript{2} sequestration with the weighted joint recovery model</td>
<td></td>
</tr>
<tr>
<td>11:45—12:10 PM</td>
<td><strong>Machine Learning at Scale (Chair: Ziyi (Francis) Yin)</strong></td>
</tr>
<tr>
<td>Mathias Louboutin</td>
<td>Randomized linear algebra for inversion</td>
</tr>
<tr>
<td>12:10—12:35 PM</td>
<td>Thomas Grady</td>
</tr>
<tr>
<td>Distributed Fourier Neural Operators</td>
<td></td>
</tr>
<tr>
<td>12:35—12:45 PM</td>
<td><strong>Discussion</strong></td>
</tr>
<tr>
<td>Time</td>
<td>Presenter</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>02:00—03:00 PM</td>
<td>Ghassan AlRegib</td>
</tr>
<tr>
<td>03:00—04:00 PM</td>
<td></td>
</tr>
<tr>
<td>04:00—05:00 PM</td>
<td></td>
</tr>
<tr>
<td>02:00—02:30 PM</td>
<td>Mathias Louboutin</td>
</tr>
<tr>
<td>02:30—03:00 PM</td>
<td>Rafael Orozco</td>
</tr>
<tr>
<td>03:00—03:45 PM</td>
<td>Ziyi (Francis) Yin</td>
</tr>
<tr>
<td>03:45—04:15 PM</td>
<td>Yijun Zhang</td>
</tr>
<tr>
<td>04:15—04:45 PM</td>
<td>Mathias Louboutin</td>
</tr>
<tr>
<td>02:00—02:45 PM</td>
<td>Felix J. Herrmann</td>
</tr>
<tr>
<td>02:45—03:45 PM</td>
<td>Yijun Zhang</td>
</tr>
<tr>
<td>03:45—04:15 PM</td>
<td>Ziyi (Francis) Yin</td>
</tr>
<tr>
<td>04:15—04:45 PM</td>
<td>Felix J. Herrmann</td>
</tr>
</tbody>
</table>
Your input

To facilitate interaction

- Q&A after each 20 min talk
- extra 10 min Q&A after each morning session
- extensive Q&A & discussions during afternoon sessions

Questions can be asked at end formal talks & during+after parallel breakout sessions.

To ask a question

- post in chat and/or “raise hand”
- you will be unmuted
- Q&A and informal sessions will not be recorded
Outcomes
ML4Seismic meeting

At the end of the meeting

- good idea what we are up to
- who we are & what our students are capable of
- what value we create for your organization
- what is next
- ways to improve
- accelerate innovation & training of talent
Enjoy the meeting!