





25 - 27 November 2022 Common Ground Bukit Bintang, Kuala Lumpur

Evaluating the Accuracy of ML-based Near Offset Data Reconstruction with Increasing Gap Size

Arvin Boutik Karpiah, Asiah Hanida Burhanuddin, Law May, N Farees Putra N Hassan, Syadhisy Dhanapal











Problem Statement



- 1. Near offset reconstruction is important in producing high-quality seismic data. It impacts many subsequent processes in the seismic processing workflow such as:
 - Surface-related multiple estimation (SRME)
 - Reverse Time Migration (RTM)
 - Amplitude Variation with Offset (AVO)
- 2. Challenges with conventional interpolation methods:
 - Operational issues during acquisition (large gap size)
 - Computational cost (large survey area)
- 3. Huge interest the geoscience community in applying ML-based solutions for seismic trace interpolation with good results in both synthetic and field dataset

Hypothesis: ML-based interpolation would struggle to fill in traces as gap size increases

Data Investigation

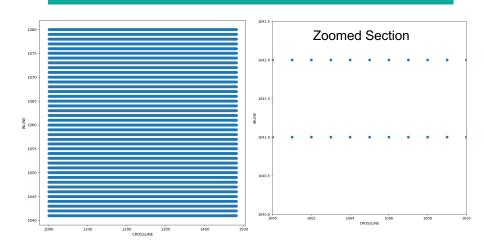


1. 3D Penobscot, Offshore Nova Scotia, Canada, working with 3D CMP Gathers

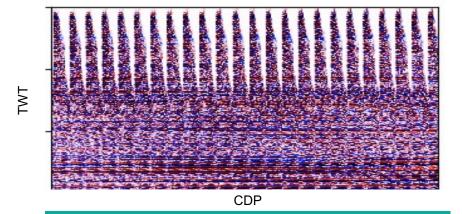
Trace Headers of 3D CMP Gathers (DataFrame)

	byte_loc	count	mean	std	min	25%	50%	75%	max
	Dyte_ioc	Count	IIIcaii	3tu		2570	3070	7570	IIIGA
TRACE_SEQUENCE_LINE		1000.0	500.500	288.819436	1.0	250.75	500.5	750.25	1000.0
TRACE_SEQUENCE_FILE	5	1000.0	1121.000	0.000000	1121.0	1121.00	1121.0	1121.00	1121.0
FieldRecord	9	1000.0	1008.653	4.740576	1000.0	1005.00	1009.0	1013.00	1017.0
TraceNumber	13	1000.0	1121.000	0.000000	1121.0	1121.00	1121.0	1121.00	1121.0
EnergySourcePoint	17	1000.0	1008.653	4.740576	1000.0	1005.00	1009.0	1013.00	1017.0
CDP	21	1000.0	1121.000	0.000000	1121.0	1121.00	1121.0	1121.00	1121.0
CDP_TRACE	25	1000.0	1008.653	4.740576	1000.0	1005.00	1009.0	1013.00	1017.0
TraceldentificationCode	29	1000.0	1.000	0.000000	1.0	1.00	1.0	1.00	1.0
NSummedTraces	31	1000.0	121.498	70.776389	2.0	60.00	124.0	184.00	240.0
NStackedTraces	33	1000.0	2.000	0.000000	2.0	2.00	2.0	2.00	2.0
DataUse	35	1000.0	0.000	0.000000	0.0	0.00	0.0	0.00	0.0
offset	37	1000.0	1658.200	886.404802	175.0	875.00	1625.0	2425.00	3175.0

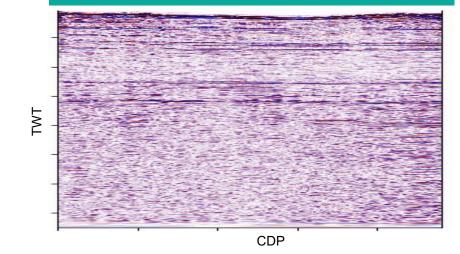
Bird's Eye View of CDP Location



3D CMP Gathers (CMP Domain) IL1401

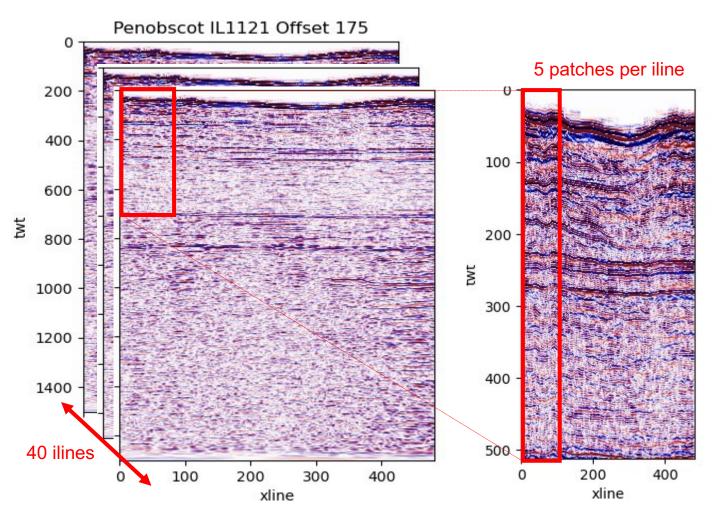


Offset 1 from IL1401

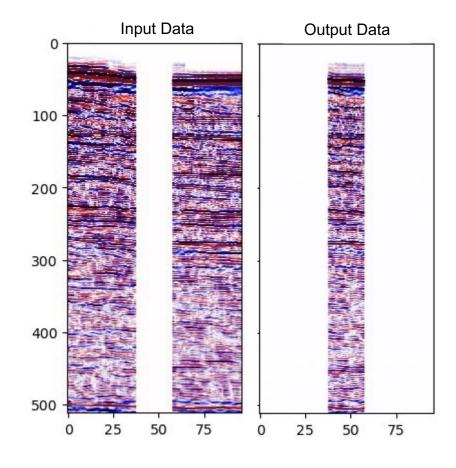


Data Preparation



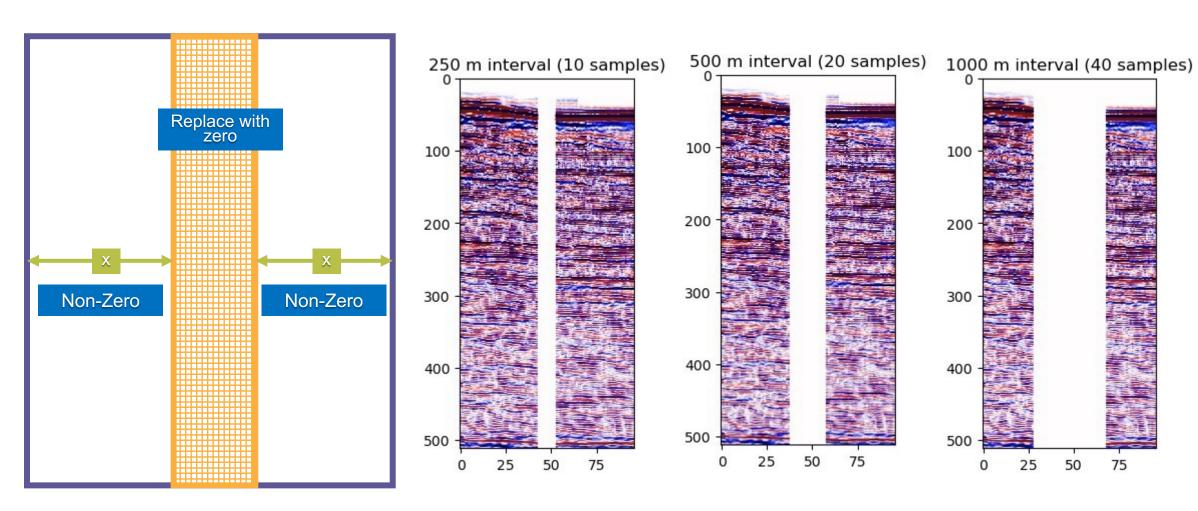


Input Data (40, 5, 512, 96) Output Data (40, 5, 512, 96)



Data Masking

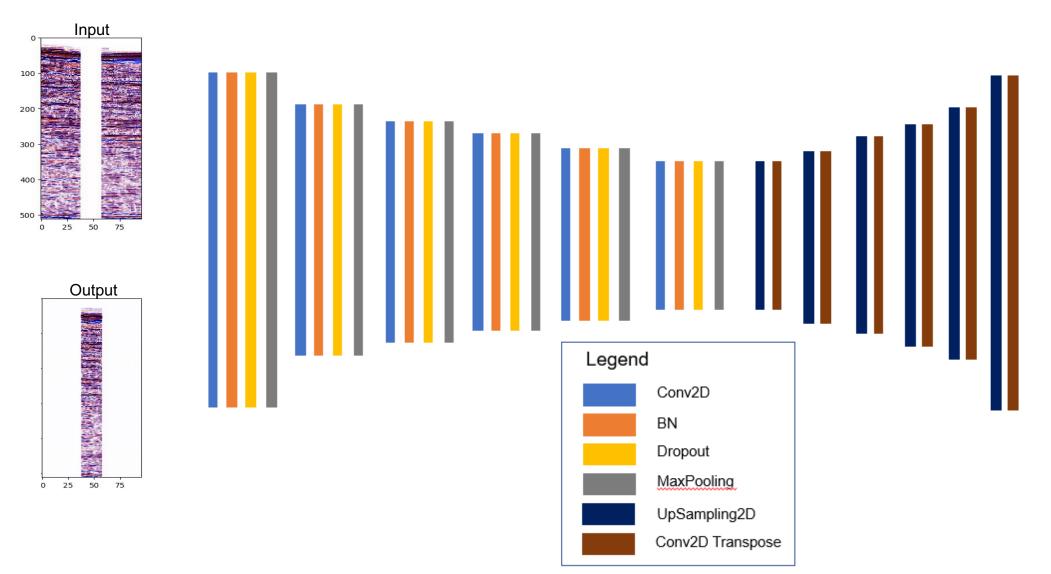




Train-test Split: 66.6% - 33.3%

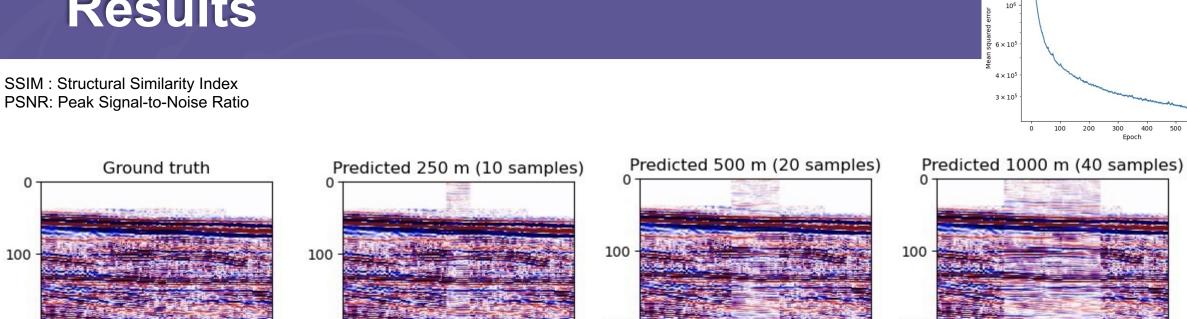
ML model - Autoencoder

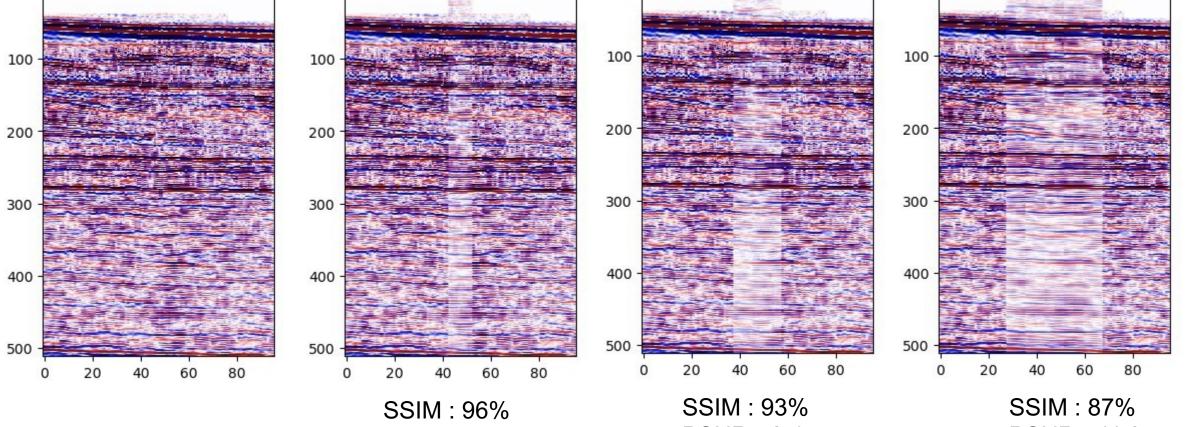




Results

PSNR: Peak Signal-to-Noise Ratio





PSNR: -8.4 PSNR: -11.3 PSNR: -4.6

Loss curve - 500 m (20 samples)

Conclusion & Way Forward



Hypothesis: ML-based interpolation would struggle to fill in traces as gap size increases

Conclusion:

- Autoencoder-based interpolation technique is able to fill in traces within the identified gap sizes.
- 2. However, there are some caveats:
 - Significant amplitude contrast at the boundary of prediction
 - Performance of autoencoder suffers as gap size increases

Way Forward:

- Spend more time to do hyperparameter tuning, test autoencoder with different loss function, gradient descent methods, etc. to improve amplitude recovery of infilled traces
- Extending 2D nature of the code in a 3D environment by utilising data in the crossline direction
- 3. Inter-offset contribution

Challenges & Lessons Learned



Challenges

- 1. Incorrect byte location during Penobscot gathers loading
- 2. Domain knowledge
- Converting longitude/latitude into the XY coordinate using the right projection datum for synthetic seismogram

Lesson Learnt

- Data is KING!
 - Always refer to official documents (repository website, acquisition/ processing report, etc.) to verify important acquisition and processing parameters
 - Spending more time on data investigation and train/test data preparation

Appreciation



• Thanks to committee, mentors, organizers and judges for the opportunity

• Work dedicated to Arvin's Patti (Grandmother)

Thank you!