SPBENCH: A MULTI-TASK EVALUATION BENCHMARK FOR EXPLORATION SEISMIC PROCESSING

Anonymous authors
Paper under double-blind review

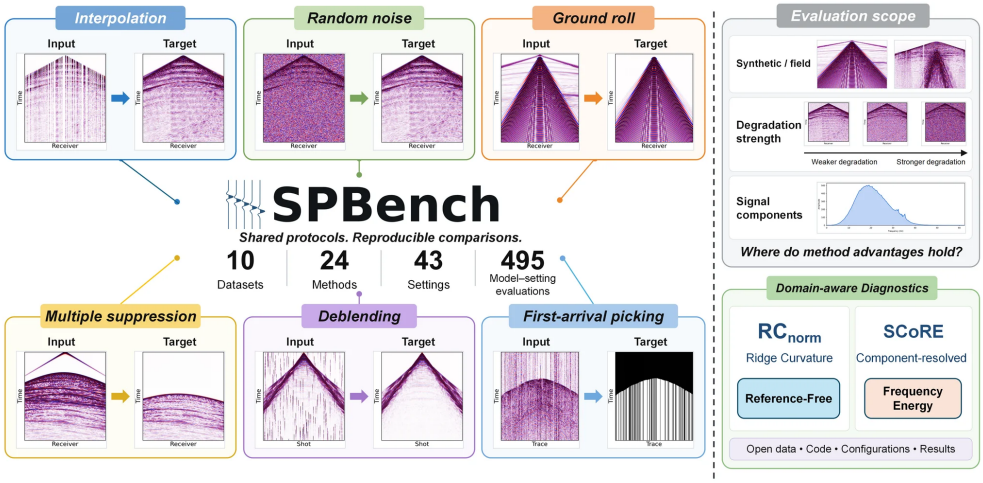


Figure 1: SPBENCH overview. SPBENCH covers six exploration seismic processing tasks under shared protocols, with open datasets, implementations, configurations, and results. Project page: `https://spbench-ano.github.io/SPBench/#/overview`.

ABSTRACT

Exploration seismic processing underpins subsurface imaging and resource exploration, where data quality shapes downstream interpretation and decision making. Learning-based methods have advanced rapidly, but their results remain difficult to compare across studies. Our literature survey of 368 papers reveals widespread reliance on private or difficult-to-reproduce datasets, and only 25 provide public code. This makes it difficult to attribute reported gains to model design rather than differences in experimental settings. To fill this gap, we introduce the Seismic Processing Benchmark (SPBENCH). SPBENCH covers six representative processing tasks, including random noise attenuation, trace interpolation, ground-roll noise suppression, multiple suppression, deblending, and first-arrival picking. We reproduce 24 supervised methods on 10 datasets under 43 standardized settings and release datasets, implementations, configurations, evaluation scripts, and results. Global scores hide frequency- and energy-dependent behavior, and per-trace pick errors ignore spatial continuity. We therefore introduce signal-component-resolved evaluation (SCoRE) for reconstruction and the reference-free ridge-curvature score ($\mathrm{RC}_{\mathrm{norm}}$) for first-arrival picking. Our analyses reveal three patterns. Synthetic rankings do not reliably predict field rankings, and the agreement varies by task when models train within each setting. As degradation strengthens, rankings reorder more under coherent ground roll than under random-like interference. The reference-free ridge score tracks MAE-based model rankings in the evaluated settings (mean Kendall correlation 0.881 across three field surveys), while SCoRE reveals component-dependent differences hidden by global scores. Together, these resources provide a fair and reproducible basis for comparing learning-based seismic-processing methods. The findings further characterize how method advantages vary across the evaluated conditions, and inform model selection.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

1 INTRODUCTION

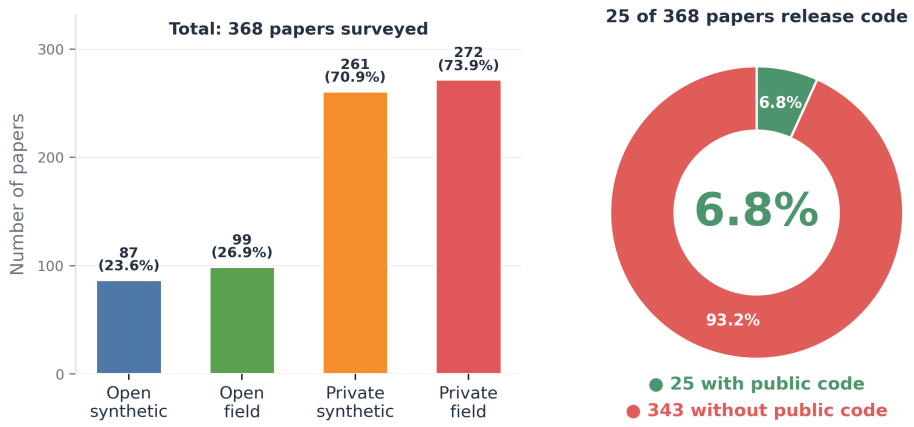
Oil and gas exploration relies on indirect geophysical measurements to infer subsurface structures and reservoir properties. Among these observations, exploration seismic data provide one of the most important sources of information for subsurface imaging, inversion, and geological interpretation (Yilmaz, 2001). These data are wavefields recorded at the surface (Appendix A). The quality of seismic data processing therefore directly affects later interpretation and exploration decisions. Machine learning has recently become a central tool in this pipeline, with neural networks increasingly used for seismic processing, interpretation, inversion, and foundation-model workflows (Bergen et al., 2019; Yu & Ma, 2021; Wu et al., 2019; Harsuko & Alkhalifah, 2022; Sheng et al., 2025; Liu et al., 2025). This growth has created a comparability problem. The literature contains many sophisticated architectures, but the field lacks shared evaluation infrastructure. It is difficult to tell which methods generalize, and whether reported gains reflect model design or differences in data, implementation, and evaluation.

We quantify this fragmentation with a curated corpus of 368 supervised deep-learning papers published between 2018 and 2026 in journals that regularly publish machine-learning research for exploration seismic processing (scope and criteria in Appendix B). Figure 2 summarizes the corpus. Two barriers stand out. First, most studies rely on private or difficult-to-reconstruct datasets with often incompletely specified generation procedures (Figure 2(a)). Second, released code is rare. Only 25 of the 368 papers include a public code link (Figure 2(b)). These patterns make published setups hard to reproduce and genuine progress hard to isolate.

To address these barriers, we introduce the Seismic Processing Benchmark (SPBENCH), an open multi-task benchmark for learning-based exploration seismic processing (Figure 1). SPBENCH improves comparability at three levels. Data construction, method implementations, and evaluation protocols are standardized across six representative tasks, including random noise attenuation, trace interpolation, ground-roll noise suppression, multiple suppression, deblending, and first-arrival picking (Liu & Ma, 2024; Li et al., 2024a; Yang et al., 2024; Wang et al., 2023a;b; St-Charles et al., 2024). Four tasks pair synthetic settings with field settings, and three tasks provide several controlled degradation levels, so methods can be compared across data conditions and degradation strengths. For multiple suppression, deblending, and ground-roll noise suppression, we release new paired synthetic datasets. The benchmark provides 10 datasets under 43 standardized settings. We implement and evaluate 24 supervised methods and report 495 model-setting evaluations over three random seeds. All datasets, implementations, splits, configurations, evaluation scripts, and per-seed results are released.

How the results are scored is part of the comparison itself. Global metrics such as the signal-to-noise ratio (SNR) and the mean absolute error (MAE) compress performance into a single number and ignore the physical structure of the data. Within one survey area, traces with similar source-receiver offsets have similar first-arrival times, so predicted picks should form a spatially continuous travel-time ridge. Building on this prior, the reference-free ridge-curvature score ($\mathrm{RC}_{\mathrm{norm}}$) measures the geometric continuity of predicted ridges, and on three field surveys it agrees with pick-error rankings at a mean Kendall rank correlation of 0.881. A second diagnostic, signal-component-resolved evaluation (SCoRE), resolves reconstruction fidelity across frequency bands with frequency-binned fidelity and recovery evaluation (FB-FRE) and across local reference-energy levels with energy-binned fidelity evaluation (EB-FE).

SPBENCH evaluates not only which methods perform well, but also how their relative advantages change across data settings, degradation strengths, and evaluation criteria. We decompose this question into three measurements. How consistent are method rankings between corresponding synthetic and field settings? How stable are rankings as degradation severity increases? Do the proposed diagnostics track the standard scores? The answers show that advantages are real but conditional. Between corresponding synthetic and field settings, where every method is trained and tested within the setting, ranking agreement varies by task, from a shared leading method to a complete turnover of the top three. Rankings remain moderately stable as random-like interference strengthens, but reorder under coherent ground roll. The ridge score agrees with pick-error rankings at 0.881, and the SCoRE partitions agree with SNR only to a degree. These advantages should therefore be read together with the conditions under which they were measured.



(a) Dataset-source usage **(b) Code availability**

Figure 2: Fragmented reproducibility in seismic-processing deep learning, measured on the 368-paper survey corpus. (a) Multi-label dataset-source usage. A paper may use several source types. (b) Only 25 of 368 papers include a public code URL.

Table 1: Comparison of representative exploration-geophysics benchmarks and resources.

Benchmark / reference	Primary scope	Type	Baselines	Code	Data	Params	Domain metrics
FaultSeg3D (Wu et al., 2019)	Fault detection	Single	×	✓	✓	×	×
OpenFWI (Deng et al., 2022)	Acoustic FWI	Single	4 methods	✓	✓	✓	×
E-FWI (Feng et al., 2023)	Elastic FWI	Single	4 methods	✓	✓	×	×
HardPicks (St-Charles et al., 2024)	First-arrival picking	Single	3 methods	✓	✓	×	×
Swell-noise (Barros et al., 2024)	Denoising	Single	2 methods	×	✓	×	×
OpenSeisML (Bhar et al., 2026)	Inversion	Single	×	×	✓	×	×
CIG-Bench (Dou et al., 2026)	Multi-task interpretation	Multi	8 methods	✓	✓	✓	×
SPBENCH	**Multi-task seismic processing**	**Multi**	**24 methods**	✓	✓	**>1000**	**SCoRE $\mathrm{RC}_{\mathrm{norm}}$**

2 RELATED WORK

Scientific machine-learning benchmarks increasingly promote comparable research by pairing shared datasets with reproduced baselines, evaluation protocols, and task-relevant metrics (Takamoto et al., 2022; Ohana et al., 2024; Rasp et al., 2024; Hu et al., 2026). Exploration-geophysics benchmark resources have begun to appear, but they target tasks that are mostly adjacent to seismic processing. OpenFWI (Deng et al., 2022) and E-FWI (Feng et al., 2023) focus on acoustic and elastic full-waveform inversion, with public datasets and reproduced baselines. OpenSeisML (Bhar et al., 2026) contributes real seismic and well-log data for inversion, but is primarily a data resource rather than a full evaluation suite with released code and parameters. FaultSeg3D (Wu et al., 2019) and CIG-Bench (Dou et al., 2026) address seismic interpretation, ranging from fault detection to multi-task subsurface interpretation. These resources are complementary to SPBENCH rather than overlapping. The inversion resources estimate subsurface models directly from recorded wavefields, and the interpretation resources analyze migrated images and volumes. SPBENCH instead standardizes the processing tasks that prepare data for imaging, inversion, and interpretation. HardPicks (St-Charles et al., 2024) and the swell-noise benchmark (Barros et al., 2024) provide valuable single-task resources for first-arrival picking and denoising. Existing open resources still lack task-specific synthetic datasets and standardized evaluation for multiple suppression, seismic deblending, and ground-roll noise suppression. As Table 1 summarizes, SPBENCH unifies six processing tasks under one shared protocol, pairs its datasets with 24 reproduced supervised baselines and 495 released configurations, and adds the seismic-specific diagnostics SCoRE and $\mathrm{RC}_{\mathrm{norm}}$, which existing resources do not provide.

3 SPBENCH OVERVIEW

3.1 TASKS AND DATA

SPBENCH defines 43 settings across six representative seismic-processing tasks and 10 datasets. All six tasks share one mathematical structure and differ only in the degradation operator and the prediction target. Let x denote a clean shot record and $\mathcal{C}_t$ a task-specific degradation operator. The observed input is

$$\tilde{x} = \mathcal{C}_t(x), \qquad \mathcal{C}_t(x) \in \{\, x + n,\;\; M \odot x,\;\; x + \alpha g,\;\; x + m,\;\; \boldsymbol{B}x \,\}, \tag{1}$$

covering additive random noise n (random-noise attenuation), a missing-trace mask M with $\odot$ denoting element-wise multiplication (interpolation), scaled ground roll αg (ground-roll suppression), multiple energy m (multiple suppression), and the blending operator $\boldsymbol{B}$ (deblending). First-arrival picking uses the recorded data directly and is annotated rather than degraded. Each method learns a mapping $f_\theta : \tilde{x} \mapsto \hat{y}$ with parameters θ, and the prediction target is chosen per task or method design. Most methods predict the clean signal directly. For ground-roll and multiple suppression, methods predict the degradation component, and the clean signal is recovered by subtraction as $\hat{x} = \tilde{x} - \hat{y}$. All methods are trained and evaluated under shared train/validation/test splits. Appendix C describes the datasets and forward modeling, Appendix E details the experiment design, and Appendix A explains how the benchmark data are recorded and simulated.

Seismic interpolation. Seismic interpolation restores absent traces so that a sparsely sampled acquisition approximates a spatially continuous wavefield. The benchmark uses the SEG C3 synthetic dataset and the Mobil AVO Viking Graben Line 12 field dataset, both openly distributed through the SEG open-data catalog[1]. Here SEG denotes the Society of Exploration Geophysicists and AVO denotes amplitude versus offset. We design three families of masks, random missing, uniform missing, and consecutive missing, each at several missing levels. This task contributes 16 benchmark settings.

Random noise attenuation. Random noise attenuation suppresses incoherent fluctuations while retaining coherent reflections, diffractions, and weak geological events. The additive noise n is seeded Gaussian or Poisson noise injected at a prescribed input SNR. The benchmark builds on the same two SEG open datasets as the interpolation task, with three SNR levels per noise family. This task contributes 12 benchmark settings.

Ground-roll noise suppression. Ground-roll noise suppression removes high-amplitude, low-velocity ground roll that masks reflection energy in land seismic records. The degradation g is elastically modeled ground roll scaled by an amplitude multiplier α. Methods predict $\hat{g}$, and the clean signal is recovered as $\hat{x} = \tilde{x} - \hat{g}$. We release the SPBENCH ground-roll datasets, comprising a synthetic set built on SEG C3 data and a companion field set. The synthetic set covers five multiplier levels. This task contributes 6 settings.

Multiple suppression. Multiple suppression separates primary reflections that directly describe subsurface interfaces from repeated-bounce energy that can produce misleading structural events. The degradation component m is the non-primary energy whose exact composition is specified in Appendix C. Methods predict $\hat{m}$, and the primaries are recovered as $\hat{x} = \tilde{x} - \hat{m}$. We release the SPBENCH multiples dataset, whose forward-modeling construction is detailed in Appendix C. This task contributes a single benchmark setting.

Deblending. Deblending disentangles overlapping wavefields recorded during simultaneous-source acquisition into the responses of individual shots. The blending operator $\boldsymbol{B}$ sums source responses shifted by random firing times. We release the SPBENCH deblending datasets, comprising a synthetic set of randomly blended common-receiver gathers with retained unblended references and a companion semi-synthetic field set built on Mobil AVO records. The synthetic settings cover three blending levels. This task contributes 4 settings.

[1] SEG open-data catalog: `https://wiki.seg.org/wiki/Open_data`

First-arrival picking. First-arrival picking identifies the earliest physically meaningful seismic onset on each trace, providing travel-time observations for statics correction and subsurface velocity estimation. The task maps a record x to per-trace arrival times t^*, realized by predicting a binary step mask per trace and reading t^* at its zero-to-one transition. The benchmark uses the Brunswick, Halfmile, and Lalor surveys from HardPicks (St-Charles et al., 2024), evaluated both per survey and on their combined evaluation set. This task contributes 4 benchmark settings.

3.2 Domain-Aware Evaluation Metrics

The benchmark reports standard reconstruction and picking metrics, including SNR, peak signal-to-noise ratio (PSNR), MAE, root mean squared error (RMSE), and tolerance-based hit rates (H@k). These metrics summarize how much error remains globally, but not how reconstruction fidelity varies across signal components or whether predictions preserve the spatial continuity of seismic events across traces. SPBench therefore adds two kinds of domain-aware diagnostics alongside the standard scores. They complement rather than replace the aggregate metrics.

Signal-component-resolved evaluation (SCoRE). SCoRE measures fidelity within partitions of the reference signal instead of only globally. Its two instances partition by different signal properties and report the same quantity per partition $\mathcal{S}$, the partition SNR

$$\mathrm{SNR}_{\mathcal{S}} = 10 \log_{10} \frac{\|r_{\mathcal{S}}\|_2^2}{\|p_{\mathcal{S}} - r_{\mathcal{S}}\|_2^2 + \varepsilon}, \tag{2}$$

where $r_{\mathcal{S}}$ and $p_{\mathcal{S}}$ are the reference and prediction restricted to the partition and ε is a small positive constant for numerical stability. FB-FRE partitions by frequency. It estimates the effective frequency band from the reference power spectrum, divides it into four adaptive subbands (low, mid, high, very high), and restricts the signals through frequency masks, with M_b the mask of subband b and RFFT and IRFFT the real discrete Fourier transform along time and its inverse,

$$r_{\mathcal{S}} = r_b = \mathrm{IRFFT}(M_b \cdot \mathrm{RFFT}(r)), \quad p_{\mathcal{S}} = p_b = \mathrm{IRFFT}(M_b \cdot \mathrm{RFFT}(p)). \tag{3}$$

A large gap between low- and high-frequency scores indicates that high-frequency detail is being removed together with noise. EB-FE partitions by local reference energy. Global scores can be dominated by strong reflections, so per-regime behavior is invisible in them. EB-FE builds a smoothed reference-energy map and groups samples into energy-percentile bins (very weak, weak, medium, strong), and each bin directly forms $\mathcal{S}$. Low-energy bins serve as a proxy for weak-signal regimes rather than explicitly identifying weak geological events. Both instances require a clean reference, and per-partition rankings agree with the global SNR ranking only to a degree (Appendix F). Full construction details are given in Appendix F.

First-arrival ridge continuity. Standard picking metrics score each trace independently. They do not use a basic physical prior. Within one survey area, traces with similar source-receiver offsets have similar first-arrival times, so the predicted picks should form a spatially continuous travel-time ridge. The ridge curvature score ($\mathrm{RC}_{\mathrm{norm}}$) measures how much the predicted ridge wiggles along the offset direction. It divides the offset range into B equal-width bins, represents each bin by its median predicted arrival, and averages the absolute slope change along the resulting ridge,

$$\mathrm{RC}_{\mathrm{norm}} = \frac{X_{\mathrm{range}}^2}{T_{\mathrm{range}}} \cdot \frac{1}{B-2} \sum_{j=3}^{B} |c_j|, \qquad c_j = \frac{d_j - d_{j-1}}{x_j - x_{j-1}}, \qquad d_j = \frac{t_j - t_{j-1}}{x_j - x_{j-1}}, \tag{4}$$

where $\{(x_j, t_j)\}_{j=1}^{B}$ are the ridge points in offset-time coordinates, X_{range} is the offset range of the valid picks and T_{range} the time range of the ridge, and the leading factor removes first-order differences in spatial and temporal scale. An affine ridge scores zero, and locally changing slopes increase the value. A low value therefore indicates a smooth aggregated ridge, not accurate picks on every trace. The median binning suppresses outlier traces by design, which adds robustness and also marks the diagnostic boundary. The metric is reference-free. It uses the geometry of the prediction itself, so it complements label-based accuracy metrics. Rearranging a fixed set of per-trace errors along the offset axis leaves MAE, RMSE, the mean bias error (MBE), and every H@k unchanged by construction, while $\mathrm{RC}_{\mathrm{norm}}$ changes, which isolates the geometric information it captures. Full construction details are given in Appendix F.

3.3 REPRODUCED BASELINES

SPBENCH reproduces 24 supervised deep-learning methods under shared protocols within each task. The baseline set combines four common architectures, U-Net (Chai et al., 2020), ResUNet (Zhang et al., 2018), DnCNN (Yu & Ma, 2021), and attention-gated U-Net (Oktay et al., 2018), with 20 task-specific methods. The common architectures provide reference points for comparison across tasks, while the task-specific methods represent designs developed for individual processing problems. Together, they span representative model families, including convolutional and residual networks (Liao et al., 2023; Wang et al., 2022b), U-Net variants (Zhang et al., 2024; Wang et al., 2025), attention-based networks and Transformers (Yu & Wu, 2022; Jiang et al., 2023), generative adversarial networks (Tao et al., 2022; Yuan et al., 2020), diffusion models (Li et al., 2024b), and model-driven or physics-constrained networks (Wu et al., 2024; Pham & Li, 2022). This combination supports comparisons of how general-purpose and specialized designs perform across the evaluated settings. Appendix D provides the complete method list, references, descriptions, and per-task assignments.

4 EXPERIMENTAL SETUP AND RESULTS

4.1 EVALUATION PROTOCOL AND MAIN RESULTS

All methods are evaluated under a shared protocol. One benchmark setting pairs a dataset with a task and, where applicable, a controlled degradation configuration. Every method in a setting uses the same train/validation/test split, the same input-target construction, and the same evaluation scripts, and validation data are used only for model selection and early stopping. We call one method evaluated on one setting an evaluation and one complete train-and-test execution with a single seed a run. Each method–setting configuration is trained independently with three random seeds. Reported results are means over the three seeds. Hyperparameters follow the source papers, and for widely used architectures such as U-Net, multiple parameter groups appear as width variants in the result tables. Reported numbers are therefore not directly comparable to values in the original papers. Setting construction, corruption protocols, and split details are given in Appendix E. The 495 configurations therefore release more than 1,000 trained parameter sets. Each reported result traces to one released configuration and to the per-seed results of its three runs, all queryable on the project page.

Tables 2 and 3 list the top eight methods per representative setting from the synchronized public leaderboard, with standard deviations over the three seeds in small type. First-arrival picking is reported separately because its MAE and tolerance-based hit rates are not commensurate with reconstruction metrics. Complete results appear in Appendix H. The remainder of this section varies one evaluation condition at a time and asks whether the conclusions hold. Section 4.2 varies the data condition, Section 4.3 the degradation strength and interference type, and Section 4.4 the metric itself.

4.2 DO SYNTHETIC RANKINGS PREDICT FIELD RANKINGS?

Does performance on the synthetic setting predict performance on the corresponding field setting? Figure 3 shows every method evaluated on both settings of each pair, sorted by synthetic performance. All methods are trained and tested within each setting (Appendix E), so this comparison concerns ranking consistency rather than cross-domain deployment. We quantify ranking agreement by the Kendall rank correlation τ, which is the concordant minus the discordant fraction of all method pairs (Appendix F) and varies widely across tasks. On random noise the rankings partially survive ($\tau = 0.73$), yet 70% of methods still lose more than 1 dB in absolute SNR. On interpolation the correlation drops to $\tau = 0.49$ and every method loses performance, with a median drop of 2.7 dB. On ground-roll suppression the correlation drops to $\tau = 0.38$, so the synthetic ranking is only a weak guide to the field ranking. On deblending the direction inverts and every method gains on the field set (median +5.6 dB), yet the ranking agreement is the weakest of the four pairs ($\tau = 0.24$). The leading method is shared for random noise and deblending (SCRN), while the top three interpolation methods turn over completely. Absolute gaps partly reflect reference construction, so we read them as setting-difficulty differences rather than deployment degradation. Synthetic performance therefore does not, by itself, reliably predict field performance. **Takeaway.** Synthetic rankings carry only partial information about field rankings, and the degree varies by task, from a shared leading

Table 2: Representative reconstruction results on five benchmark settings (top eight methods by output SNR per setting). All scores are means over three seeds with standard deviations in small type. Rows are ordered by increasing SNR; shading is computed per block and metric. Setting labels: G0 is Gaussian noise at 0 dB input SNR, R50 is random missing of 50% of traces, L5 is ground-roll multiplier level 5 of five levels (L1–L9), and mod is the moderate of three blending levels.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
Random noise (SEG C3, G0)								
Attention UNet	7.85M	12.80±0.10	12.00±0.15	15.73±0.07	11.98±0.08	2.18±0.10	1.26±0.09	14.82±0.10
FBResNet	0.45M	12.97±0.00	11.99±0.04	15.45±0.01	12.12±0.02	2.91±0.01	1.47±0.02	15.00±0.01
QUNet	21.88M	13.08±0.08	12.59±0.16	15.81±0.04	12.18±0.05	2.61±0.05	1.34±0.06	15.19±0.09
UNet-L	31.04M	13.13±0.04	12.78±0.06	15.91±0.06	12.23±0.04	2.60±0.06	1.44±0.03	15.25±0.05
UNet++	9.05M	13.47±0.04	12.78±0.06	15.90±0.04	12.62±0.04	3.26±0.08	1.74±0.09	15.63±0.03
CBD-RDN	1.49M	13.58±0.45	13.23±0.59	15.92±0.37	12.57±0.42	3.53±0.53	1.67±0.24	16.02±0.53
DnCNN	0.56M	14.10±0.06	13.75±0.13	16.40±0.08	13.16±0.05	3.93±0.03	1.91±0.05	16.67±0.11
SCRN	0.43M	**14.38±0.01**	**14.01±0.00**	**16.60±0.02**	**13.40±0.01**	**4.29±0.01**	**2.19±0.00**	**16.92±0.01**
Interpolation (SEG C3, R50)								
DnCNN	0.14M	16.11±0.02	20.39±0.09	20.61±0.05	12.74±0.02	10.15±0.02	11.58±0.00	18.88±0.06
ResUNet-L	32.44M	16.23±0.06	**24.92±0.41**	20.83±0.18	12.79±0.03	10.50±0.02	11.42±0.02	19.61±0.16
CFunet	7.35M	16.23±0.09	22.69±0.11	20.72±0.09	12.88±0.10	10.56±0.13	11.48±0.08	19.42±0.13
UNet-L	31.04M	16.26±0.03	23.82±0.29	20.95±0.03	12.81±0.05	10.38±0.05	11.42±0.05	19.63±0.03
Attention UNet-L	31.39M	16.31±0.13	23.71±0.09	20.94±0.18	12.88±0.12	10.48±0.10	11.48±0.13	19.64±0.13
WRDL	35.83M	16.54±0.46	24.84±1.42	**21.51±0.43**	13.20±0.30	10.49±0.43	**12.22±0.08**	**20.10±0.08**
CA-Unet	7.79M	16.55±0.11	22.84±0.40	20.96±0.14	**13.27±0.10**	**10.73±0.15**	11.84±0.14	19.76±0.03
DnCNN-L	0.56M	**16.56±0.01**	22.39±0.16	21.25±0.10	13.13±0.04	10.59±0.07	11.88±0.04	19.62±0.04
Ground roll (SPBench, L5)								
SANet	0.23M	14.26±0.14	5.25±0.34	17.62±0.13	5.49±0.07	3.73±0.08	3.23±0.14	15.61±0.15
DnCNN	0.14M	15.57±0.22	9.13±0.39	16.65±0.21	10.90±0.38	9.40±0.50	9.83±0.33	15.84±0.23
ResUNet	8.11M	18.11±2.12	4.96±3.67	23.94±1.49	**20.25±1.45**	**19.23±0.52**	5.59±2.49	20.16±1.95
ResUNet-L	32.44M	19.52±0.29	11.98±0.15	20.64±0.38	16.39±0.00	14.99±0.13	10.57±0.27	20.23±0.31
UNet	7.76M	20.22±0.32	11.46±0.17	21.76±0.27	16.78±1.04	15.37±1.25	11.25±0.30	20.92±0.31
Attention UNet-L	31.39M	20.23±0.33	11.55±0.12	21.92±0.35	15.73±1.29	14.21±1.39	11.09±0.27	20.98±0.36
UNet-L	31.04M	20.39±0.13	11.44±0.31	22.12±0.24	16.33±0.21	14.49±0.22	10.81±0.30	21.25±0.17
Attention UNet	7.85M	**21.96±1.06**	**12.06±2.17**	**24.22±0.92**	18.43±0.45	17.55±0.36	**11.97±1.79**	**22.96±0.92**
Multiple (SPBench)								
DNNDAT	37.07M	18.51±0.46	19.18±1.12	19.21±0.68	18.71±0.40	15.36±1.19	-13.83±0.77	25.03±0.24
UNet	7.76M	20.52±0.58	22.65±0.44	21.69±0.63	20.05±0.64	15.17±0.35	-11.49±0.72	26.35±0.32
ResUNet	8.11M	20.77±0.71	22.64±0.47	21.92±0.67	20.82±0.86	15.99±1.06	-10.95±0.86	25.93±0.59
Attention UNet	7.85M	20.85±0.88	22.08±0.67	22.08±0.92	20.94±0.94	15.80±0.90	-11.13±1.06	27.03±0.16
ResUNet-L	32.44M	21.61±0.54	23.91±0.48	22.49±0.50	21.73±0.49	17.81±0.74	-10.08±0.61	26.98±0.48
Attention UNet-L	31.39M	22.33±0.46	23.77±0.54	23.23±0.54	22.71±0.86	**18.53±0.83**	-8.94±0.89	27.44±0.37
SAGAN	32.34M	22.44±1.46	20.27±1.53	**23.75±1.23**	22.29±2.10	16.42±1.71	-9.08±2.13	26.89±0.61
UNet-L	31.04M	**22.75±0.64**	**24.18±0.47**	23.68±0.57	**22.93±0.77**	18.18±0.63	**-8.37±0.90**	**27.51±0.25**
Deblending (SPBench, mod)								
Attention UNet	7.85M	14.56±0.14	11.38±0.17	15.10±0.15	15.20±0.17	9.55±0.31	6.85±0.07	15.30±0.18
UNet	7.76M	14.70±0.24	11.37±0.06	15.07±0.54	15.48±0.21	10.10±0.72	6.50±0.33	15.53±0.28
UNet++	9.05M	16.13±0.10	13.24±0.17	15.95±0.07	16.88±0.27	**13.26±0.09**	8.38±0.19	16.71±0.13
UNet-L	31.04M	16.68±0.64	13.61±0.61	16.58±0.71	17.49±0.55	13.13±0.16	8.45±0.59	17.38±0.66
FBResNet	0.45M	17.01±0.00	11.41±0.00	18.05±0.00	17.43±0.00	10.44±0.00	9.27±0.00	17.88±0.00
cDDPM	33.3M	17.24±0.23	**15.25±0.23**	19.50±0.13	18.77±0.16	11.60±0.32	8.80±0.24	19.51±0.15
DnCNN	0.56M	17.40±0.05	11.44±0.07	18.48±0.06	17.94±0.06	10.60±0.10	9.97±0.07	18.19±0.05
SCRN	0.43M	**19.36±0.04**	14.43±0.28	**20.16±0.10**	**19.40±0.08**	13.12±0.17	**11.93±0.09**	**19.94±0.03**

Table 3: Representative first-arrival picking results on the Lalor survey, listing the top eight methods by MAE. All scores are means over three seeds with standard deviations in small type. Hit rates (H@k) measure the fraction of picks within k samples, and $\mathrm{RC}_{\mathrm{norm}}$ measures the continuity of the predicted first-arrival ridge. Cell shading is computed independently for each metric.

Setting	Method	Params	MAE	$\mathrm{RC}_{\mathrm{norm}}$	H@1	H@5	H@9
First arrival	STUNet	71.55M	**2.27±0.31**	**18.78±0.94**	0.236±0.022	0.869±0.021	**0.962±0.009**
	DSU-Net	1.99M	5.69±0.72	21.85±4.43	0.545±0.011	**0.885±0.006**	0.949±0.005
	Attention UNet	7.85M	13.18±0.83	31.88±0.44	0.616±0.022	0.860±0.009	0.920±0.004
	Attention UNet-L	31.39M	15.63±1.85	41.21±3.26	**0.672±0.042**	0.858±0.011	0.912±0.010
	HUNet	7.76M	17.05±8.10	38.08±26.88	0.507±0.120	0.840±0.050	0.904±0.041
	UNet	7.76M	17.93±4.32	42.52±11.56	0.544±0.031	0.836±0.012	0.897±0.015
	ResUNet-L	32.44M	18.35±1.51	47.30±3.71	0.633±0.042	0.842±0.011	0.898±0.009
	UNet-L	31.04M	19.37±12.79	50.52±35.13	0.583±0.103	0.836±0.058	0.893±0.055

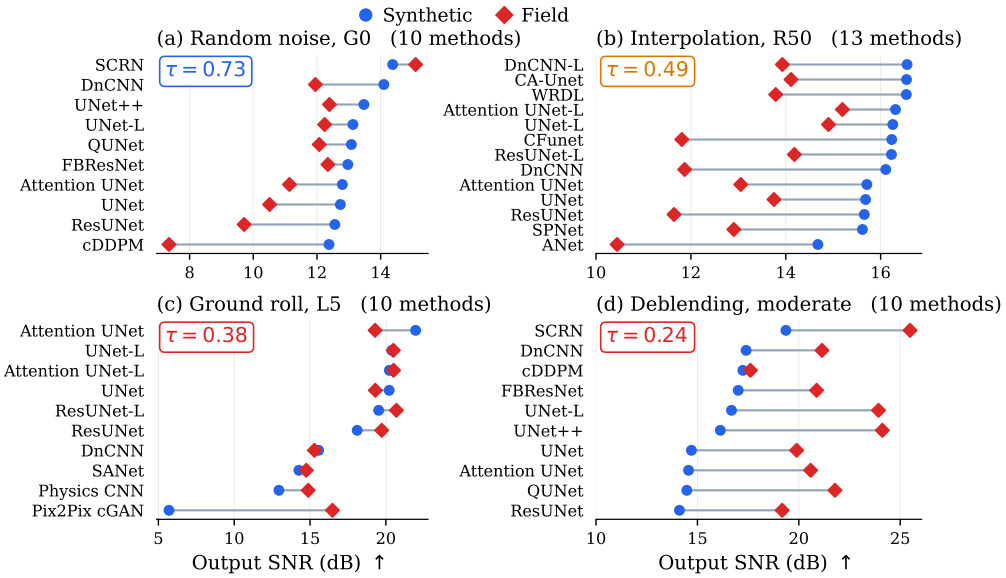


Figure 3: Synthetic-to-field performance under matched protocols. Each point reports the mean output SNR over three random seeds, and horizontal segments connect the same method across paired synthetic and field settings. Methods are ordered by their synthetic-setting performance, with the best-performing method placed at the top. Kendall's τ_b measures the agreement between the two method rankings. Models are trained and evaluated separately within each setting; the comparison therefore measures ranking consistency rather than cross-domain generalization. Only methods evaluated in both settings are included.

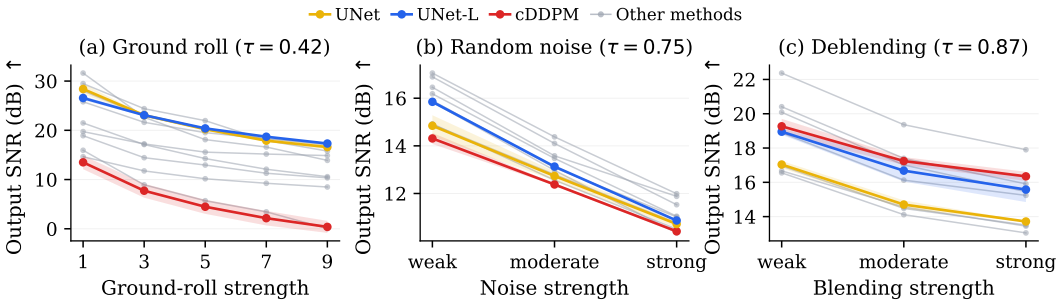


Figure 4: Output SNR across degradation strengths for the three multi-level tasks. Curves show means over three random seeds; shaded bands give one standard deviation for the highlighted UNet (yellow), its width-increased variant UNet-L (dark blue), and cDDPM (red); all other methods are in light gray. Panel titles report the Kendall τ between endpoint levels. (a) Rankings partially reorder as ground-roll strength grows: cDDPM collapses to 0.4 dB, and UNet-L matches UNet at L3 and leads from L5 onward. (b,c) UNet-L leads UNet at every noise and blending level.

method to a complete turnover of the top three. Model claims should therefore be validated on matched synthetic and field settings. Classic architectures remain strong baselines: DnCNN ranks in the top three of all three synthetic deblending settings and U-Net variants lead interpolation in 9 of its 16 settings.

4.3 How Robust Are Models to Noise Strength?

Figure 4 tracks output SNR as degradation strengthens in the three tasks that provide multiple levels. On random noise (Figure 4(b)) and on deblending (Figure 4(c)), absolute scores degrade almost in parallel and rankings remain moderately stable. The Kendall correlation between the weakest and strongest levels is 0.75 for Gaussian noise and 0.87 for deblending, and SCRN leads at every noise level. Ground roll behaves differently: rankings partially reorder as strength grows, with $\tau = 0.42$ between levels 1 and 9. ResUNet falls from first at 31.6 dB to seventh at 13.9 dB, and cDDPM

Table 4: Kendall rank correlation of the reference-free ridge score with MAE-based and hit-rate-based model rankings (rows 1–2) and between its two aggregation rules (row 3), per field picking survey over the ten evaluated methods.

Ranking pair	Brunswick	Halfmile	Lalor	Mean
$\mathrm{RC}_{\mathrm{norm}}$ vs pick error (MAE)	0.778	0.911	0.956	0.881
$\mathrm{RC}_{\mathrm{norm}}$ vs 1−H@5	0.689	0.899	0.822	0.803
Shot macro vs trace weighted	0.956	0.956	1.000	0.970

collapses from 13.5 dB to 0.4 dB. The width-increased variant UNet-L (Appendix D) behaves oppositely, starting 1.8 dB below UNet at L1, matching it at L3, and ending 0.7 dB ahead at L9. Under random noise and blending, UNet-L leads UNet at every level. The random-noise and ground-roll settings share the same clean reference shots (SEG C3 shots 1–9), so this contrast is observed between interference types on the same background. One plausible account is morphological. Ground roll forms dispersive, low-frequency, high-amplitude cones that overlap reflection events, unlike the random-like interference of the other two tasks. Random noise is incoherent by construction, and blended energy appears as incoherent vertical bursts in the common-receiver gathers used here, a form that classical random-noise suppression flows are designed to attenuate. This is an association, not a tested mechanism. **Takeaway.** Ranking stability under growing interference is a separate selection dimension. It is moderate under random-like corruption but weak under coherent ground roll, so it must be assessed per interference type.

4.4 Can We Rank Picking Models Without Reference Labels?

MAE and hit rates both require reference picks, and on field data many traces lack usable labels entirely (Figure 11, Appendix C). $\mathrm{RC}_{\mathrm{norm}}$ avoids this dependency because it reads the geometry of the prediction itself. It is useful for model selection only if its model rankings agree with labeled evaluation, which we test directly. On each of the three field surveys, we rank the ten evaluated methods by mean $\mathrm{RC}_{\mathrm{norm}}$ and by MAE, and compute the Kendall rank correlation between the two rankings (Table 4). The agreement is strong on every survey, with a mean of 0.881, and holds under a hit-rate-based view as well (1−H@5, mean 0.803). It varies by survey, and Brunswick is weakest at 0.778. We therefore read the score per survey, not as a universal ranking guarantee. The two aggregation choices, shot-level macro averaging and trace-count weighting, rank the models nearly identically (mean 0.970), so the conclusions depend on neither the accuracy metric nor the aggregation rule.

$\mathrm{RC}_{\mathrm{norm}}$ has a clear boundary. It is blind to a global time shift, so a smooth but systematically offset prediction can still score well. This blindness is visible in the tolerance sweep (Table 8), where the agreement falls monotonically from 0.862 at H@9 to 0.244 at H@1. Where reference picks exist, MAE and hit rates remain the primary accuracy measures, and $\mathrm{RC}_{\mathrm{norm}}$ complements them. **Takeaway.** Across the ten evaluated methods on three field surveys, $\mathrm{RC}_{\mathrm{norm}}$ rankings show strong agreement with MAE rankings, with a mean Kendall correlation of 0.881. These results support its use as a reference-free auxiliary criterion for model comparison in the evaluated settings. $\mathrm{RC}_{\mathrm{norm}}$ measures ridge geometry and does not establish absolute picking accuracy.

5 Conclusion

In this work, to address the limited comparability of learning-based seismic processing studies, we introduce SPBench, an open multi-task benchmark. It incorporates six tasks, 10 datasets, 43 settings, 24 methods, 495 model-setting evaluations, and two diagnostic protocols. Our analyses show that conclusions are conditional on the data, the degradation, and the metric. Synthetic rankings carry only partial information about field rankings. Ranking stability under growing interference is a separate selection dimension, moderate under random-like corruption but weaker under coherent ground roll. For first-arrival picking, the reference-free ridge score ranks models consistently with pick-error evaluation without any labels. SPBench currently centers on supervised deep learning over two-dimensional gathers, and classical flows, foundation models, and three-dimensional acquisition remain to be covered. Future versions will add field datasets, baselines, and hidden-test evaluation. We hope SPBench will provide a more comprehensive and fair basis for comparing these methods and spur robust learning-based seismic processing on field data.

AI USE STATEMENT

In this work, generative AI tools assisted with developing the benchmark code framework, implementing and reproducing baseline methods, writing data-processing code, and preparing tables from computed experimental results. We also used these tools for language editing, LaTeX formatting, visualization prototyping, feedback on evaluation methodology and experimental design, and assistance with interpreting and presenting results. The authors assessed this feedback and made the final research decisions. Benchmark measurements were obtained by executing the experimental pipeline. The authors reviewed AI-assisted text and figures, executed and checked AI-assisted code, checked reproduced implementations against the source methods, and verified reported numerical results against experimental records. The authors take responsibility for the accuracy, originality, and integrity of the manuscript and accompanying artifacts.

ETHICS STATEMENT

This work studies benchmark infrastructure for exploration seismic data processing. Potential risks include benchmark overfitting, misuse of proprietary data, and overstating generalization from synthetic or restricted datasets. We mitigate these risks by documenting data provenance, separating provisional estimates from audited statistics, and emphasizing transparent evaluation protocols.

REPRODUCIBILITY STATEMENT

The benchmark is designed around fixed splits, public evaluation scripts, baseline implementations, and versioned benchmark artifacts. All code, parameter configurations, benchmark data, model artifacts, metric implementations, and leaderboard resources are available at `https://spbench-ano.github.io/SPBench/#/overview`. Each released configuration records the data split, degradation realization, normalization, input construction, training schedule, optimizer, stopping rule, and random seeds for one run, so every leaderboard number can be audited against its exact parameter setting. All resources are provided anonymously at the project page.

REFERENCES

Pablo M. Barros, Roosevelt de L. Sardinha, Giovanny A. M. Arboleda, Lessandro de S. S. Valente, Isabelle R. V. de Melo, Albino Aveleda, Andre Bulcao, Sergio L. Netto, and Alexandre G. Evsukoff. A real benchmark swell noise dataset for performing seismic data denoising via deep learning. *arXiv preprint arXiv:2410.08231*, 2024.

Karianne J. Bergen, Paul A. Johnson, Maarten V. de Hoop, and Gregory C. Beroza. Machine learning for data-driven discovery in solid Earth geoscience. *Science*, 363(6433), 2019. doi: 10.1126/science.aau0323.

Ipsita Bhar, Huseyin Tuna Erdinc, Thales Souza, Charles Jones, and Felix J. Herrmann. OpenSeisML: Open large-scale real seismic and well-log dataset for generative AI. *arXiv preprint arXiv:2605.20539*, 2026.

Xintao Chai, Genyang Tang, Shangxu Wang, Ronghua Peng, Wei Chen, and Jingnan Li. Deep learning for regularly missing data reconstruction. *IEEE Transactions on Geoscience and Remote Sensing*, 58(6):4406–4423, 2020. doi: 10.1109/TGRS.2020.2963928.

Chengyuan Deng, Shihang Feng, Hanchen Wang, Xitong Zhang, Peng Jin, Yinan Feng, Qili Zeng, Yinpeng Chen, and Youzuo Lin. OpenFWI: Large-scale multi-structural benchmark datasets for full waveform inversion. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL `https://proceedings.neurips.cc/paper_files/paper/2022/hash/27d3ef263c7cb8d542c4f9815a49b69b-Abstract-Datasets_and_Benchmarks.html`. Datasets and Benchmarks Track.

Yimin Dou, Xinming Wu, Hui Gao, Mingliang Liu, Tao Zhao, Zhi Zhong, Haibin Di, Min Jun Park, Robert G. Clapp, Zhixiang Guo, Long Han, and Sergey Fomel. CIG-Bench: A comprehensive survey and benchmark for AI-driven subsurface imaging understanding. *arXiv preprint arXiv:2606.09094v1*, 2026.

Shihang Feng, Hanchen Wang, Chengyuan Deng, Yinan Feng, Yanhua Liu, Min Zhu, Peng Jin, Yinpeng Chen, and Youzuo Lin. E-FWI: Multiparameter benchmark datasets for elastic full waveform inversion of geophysical properties. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://papers.nips.cc/paper_files/paper/2023/hash/4aa8d18aad014fb3d0076e0afd2e3b2e-Abstract-Datasets_and_Benchmarks.html. Datasets and Benchmarks Track.

Lei Gao, Housen Shen, and Fan Min. Swin Transformer for simultaneous denoising and interpolation of seismic data. *Computers & Geosciences*, 183:105510, 2024.

Randy Harsuko and Tariq A. Alkhalifah. StorSeismic: A new paradigm in deep learning for seismic processing. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022. doi: 10.1109/TGRS.2022.3216660.

Peiyan Hu, Haodong Feng, Hongyuan Liu, Tongtong Yan, Wenhao Deng, Tianrun Gao, Rong Zheng, Haoren Zheng, Chenglei Yu, Chuanrui Wang, Kaiwen Li, Zhi-Ming Ma, Dezhi Zhou, Xingcai Lu, Dixia Fan, and Tailin Wu. RealPDEBench: A benchmark for complex physical systems with real-world data. In *International Conference on Learning Representations*, 2026. URL https://arxiv.org/abs/2601.01829.

Peifan Jiang, Fei Deng, Xuben Wang, Pengfei Shuai, Wen Luo, and Yun Tang. Seismic first break picking through Swin Transformer feature extraction. *IEEE Geoscience and Remote Sensing Letters*, 20:1–5, 2023. doi: 10.1109/LGRS.2023.3248233.

Ji Li, Daniel Trad, and Dawei Liu. Robust seismic data denoising via self-supervised deep learning. *Geophysics*, 89(5):V437–V451, 2024a. doi: 10.1190/geo2023-0762.1.

Xinze Li, Bangyu Wu, Xu Zhu, and Hui Yang. Consecutively missing seismic data interpolation based on coordinate attention UNet. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022. doi: 10.1109/LGRS.2021.3128511.

Yuanyuan Li, Hao Zhang, Jianping Huang, and Zhenchun Li. Conditional denoising diffusion probabilistic model for ground-roll attenuation. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–12, 2024b. doi: 10.1109/TGRS.2024.3489716.

Zhangquan Liao, Yong Li, Yingtian Liu, Yifan Yang, and Yiming Zhang. Residual learning with feedback for strong random noise attenuation in seismic data. *Geophysics*, 88(1):V45–V57, 2023. doi: 10.1190/geo2021-0621.1.

Naihao Liu, Lukun Wu, Jiale Wang, Hao Wu, Jinghuai Gao, and Dehua Wang. Seismic data reconstruction via wavelet-based residual deep learning. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022. doi: 10.1109/TGRS.2022.3152984.

Qi Liu and Jianwei Ma. Generative interpolation via a diffusion probabilistic model. *Geophysics*, 89(1):V65–V85, 2024. doi: 10.1190/geo2023-0182.1.

Qi Liu, Zenggui Chen, and Jianwei Ma. Foundation models for exploration geophysics. *Geophysics*, 90(6):H29–H52, 2025. doi: 10.1190/geo2024-0916.1.

Gislain B. Madiba and George A. McMechan. Processing, inversion, and interpretation of a 2D seismic data set from the North Viking Graben, North Sea. *Geophysics*, 68(3):837–848, 2003. doi: 10.1190/1.1581036.

Ruben Ohana, Michael McCabe, Lucas Meyer, Rudy Morel, Fruzsina J. Agocs, Miguel Beneitez, Marsha Berger, Blakesley Burkhart, Stuart B. Dalziel, Drummond B. Fielding, Daniel Fortunato, Jared A. Goldberg, Keiya Hirashima, Yan-Fei Jiang, Rich R. Kerswell, Suryanarayana Maddu, Jonah Miller, Payel Mukhopadhyay, Stefan S. Nixon, Jeff Shen, Romain Watteaux, Bruno Régaldo-Saint Blancard, François Rozet, Liam H. Parker, Miles Cranmer, and Shirley Ho. The Well: A large-scale collection of diverse physics simulations for machine learning. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-1430. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/4f9a5acd91ac76569f2fe291b1f4772b-Abstract-Datasets_and_Benchmarks_Track.html. Datasets and Benchmarks Track.

Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention U-Net: Learning where to look for the pancreas. In *Medical Imaging with Deep Learning*, 2018.

Hanjoon Park, Jun-Woo Lee, Jongha Hwang, and Dong-Joo Min. Coarse-refine network with upsampling techniques and fourier loss for the reconstruction of missing seismic data. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022. doi: 10.1109/TGRS.2022.3190292.

Nam Pham and Weichang Li. Physics-constrained deep learning for ground roll attenuation. *Geophysics*, 87(1):V15–V27, 2022. doi: 10.1190/geo2020-0691.1.

Yitao Pu, Bo Zhang, Chenglin Wei, Yingyu Xu, and Hongfei Liu. Improving the accuracy of predicted first-arrival times using deep learning. *Geophysics*, 89(2):V169–V178, 2024. doi: 10.1190/geo2023-0154.1.

Stephan Rasp, Stephan Hoyer, Alexander Merose, Ian Langmore, Peter Battaglia, Tyler Russell, Alvaro Sanchez-Gonzalez, Vivian Yang, Rob Carver, Shreya Agrawal, Matthew Chantry, Zied Ben Bouallegue, Peter Dueben, Carla Bromberg, Jared Sisk, Luke Barrington, Aaron Bell, and Fei Sha. WeatherBench 2: A benchmark for the next generation of data-driven global weather models. *Journal of Advances in Modeling Earth Systems*, 16(6):e2023MS004019, 2024. doi: 10.1029/2023MS004019.

Hanlin Sheng, Xinming Wu, Xu Si, Jintao Li, Sibo Zhang, and Xudong Duan. Seismic foundation model: A next generation deep-learning model in geophysics. *Geophysics*, 90(2):IM59–IM79, 2025. doi: 10.1190/geo2024-0262.1.

Pierre-Luc St-Charles, Bruno Rousseau, Joumana Ghosn, Gilles Bellefleur, and Ernst Schetselaar. A deep learning benchmark for first break detection from hardrock seismic reflection data. *Geophysics*, 89(1):WA279–WA294, 2024. doi: 10.1190/geo2022-0741.1.

Siyu Sun, Hao Li, Guofa Li, and Zhewu Wang. Ground roll attenuation in seismic data based on enhanced deep learning framework with adaptive frequency modulation loss. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–16, 2024. doi: 10.1109/TGRS.2024.3427405.

Makoto Takamoto, Timothy Praditia, Raphael Leiteritz, Dan MacKinlay, Francesco Alesiani, Dirk Pflüger, and Mathias Niepert. PDEBench: An extensive benchmark for scientific machine learning. In *Advances in Neural Information Processing Systems*, volume 35, pp. 1596–1611, 2022.

Liurong Tao, Haoran Ren, Yueming Ye, and Jinsheng Jiang. Seismic surface-related multiples suppression based on SAGAN. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022.

Hongtao Wang, Rongyu Feng, Liangyi Wu, Mutian Liu, Yinuo Cui, Chunxia Zhang, and Zhenbo Guo. DSU-Net: Dynamic snake U-Net for 2-D seismic first break picking. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–13, 2024. doi: 10.1109/TGRS.2024.3457521.

Kunxi Wang, Tianyue Hu, Shangxu Wang, and Jianxin Wei. Seismic multiple suppression based on a deep neural network method for marine data. *Geophysics*, 87(4):V341–V365, 2022a. doi: 10.1190/geo2021-0206.1.

Kunxi Wang, Tianyue Hu, and Shangxu Wang. Iterative deblending using unsupervised learning with double-deep neural networks. *Geophysics*, 88(3):V187–V205, 2023a. doi: 10.1190/geo2022-0299.1.

Kunxi Wang, Tianyue Hu, Bangliu Zhao, and Shangxu Wang. Surface-related multiple attenuation based on a self-supervised deep neural network with local wavefield characteristics. *Geophysics*, 88(5):V387–V402, 2023b. doi: 10.1190/geo2022-0599.1.

Rongqian Wang, Ruixuan Zhang, Chenglong Bao, Lingyun Qiu, and Dinghui Yang. Adapting the residual dense network for seismic data denoising and upscaling. *Geophysics*, 87(4):V321–V340, 2022b. doi: 10.1190/geo2021-0294.1.

Xiaojing Wang, Yuhan Sui, and Jianwei Ma. Quadratic UNet for seismic random noise attenuation. *Geophysics*, 90(2):V53–V65, 2025. doi: 10.1190/geo2024-0019.1.

Mengyi Wu, Lihua Fu, Wenqian Fang, and Jiajia Cao. Sparse Prior-Net: A sparse prior-based deep network for seismic data interpolation. *Geophysics*, 89(1):V37–V47, 2024. doi: 10.1190/geo2022-0262.1.

Xinming Wu, Luming Liang, Yangkang Shi, and Sergey Fomel. FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. *Geophysics*, 84(3):IM35–IM45, 2019. doi: 10.1190/geo2018-0646.1.

Liuqing Yang, Sergey Fomel, Shoudong Wang, Xiaohong Chen, and Yangkang Chen. Deep learning with soft attention mechanism for small-scale ground roll attenuation. *Geophysics*, 89(1):WA179–WA193, 2024. doi: 10.1190/geo2023-0150.1.

Öz Yilmaz. *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data*. Society of Exploration Geophysicists, 2001. doi: 10.1190/1.9781560801580.

Jiaxu Yu and Bangyu Wu. Attention and hybrid loss guided deep learning for consecutively missing seismic data reconstruction. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–8, 2022. doi: 10.1109/TGRS.2021.3068279.

Siwei Yu and Jianwei Ma. Deep learning for geophysics: Current and future trends. *Reviews of Geophysics*, 59(3):e2021RG000742, 2021. doi: 10.1029/2021RG000742.

Yijun Yuan, Xu Si, and Yue Zheng. Ground-roll attenuation using generative adversarial networks. *Geophysics*, 85(4):WA255–WA267, 2020. doi: 10.1190/geo2019-0414.1.

Yongjie Zhang, Bingluo Gu, Zhiguang Sun, Xinyue Yan, and Shanshan Zhang. Seismic data denoising with two-step prediction strategy based on neural network. *Computers & Geosciences*, 187:105595, 2024.

Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018.

APPENDIX

A EXPLORATION SEISMIC DATA AS RECORDED AND SIMULATED WAVEFIELDS

This appendix is written for readers from the machine-learning community. It explains what exploration seismic data are and how they relate to the wave equation. The key point is that seismic data are samples of a wavefield. Some of the data in this benchmark are measured in the field, and some are computed by solving a partial differential equation. The two are connected by the same physics.

Exploration seismology probes the subsurface with controlled sources. An active source, such as an air-gun array at sea or a vibrator truck on land, injects mechanical energy into the Earth. The energy travels downward as elastic waves, and part of it returns after reflection, refraction, and diffraction at subsurface interfaces. Receivers at the surface record the returning wavefield over time. A shot gather is one of the primary data representations used in this benchmark, collecting the recordings for one source position. Each trace is one receiver's time series, so a shot gather is a spatiotemporal sample of the true wavefield of the Earth at the receiver locations. Other arrangements of the same wavefield, such as common-receiver gathers and three-dimensional receiver-grid volumes, are used where a task requires them. The field datasets in this benchmark, such as the Mobil AVO Viking Graben line and the field first-arrival data, are actual measurements obtained in this way. They are observations of the real Earth, not model outputs.

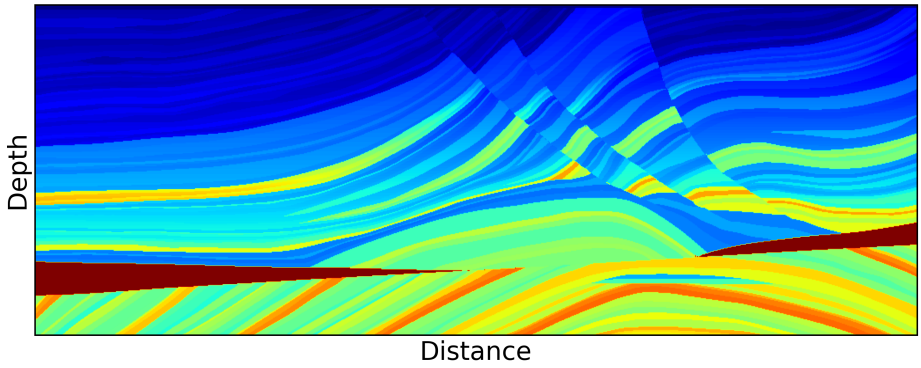
The same wavefield can also be computed. Wave propagation in the subsurface is governed by partial differential equations. In the simplest useful approximation, the pressure field $u(\boldsymbol{x}, t)$ in a medium with wave speed $v(\boldsymbol{x})$ satisfies the constant-density acoustic wave equation

$$\frac{1}{v(\boldsymbol{x})^2}\,\frac{\partial^2 u}{\partial t^2} \;=\; \nabla^2 u \;+\; s(\boldsymbol{x}, t), \tag{5}$$

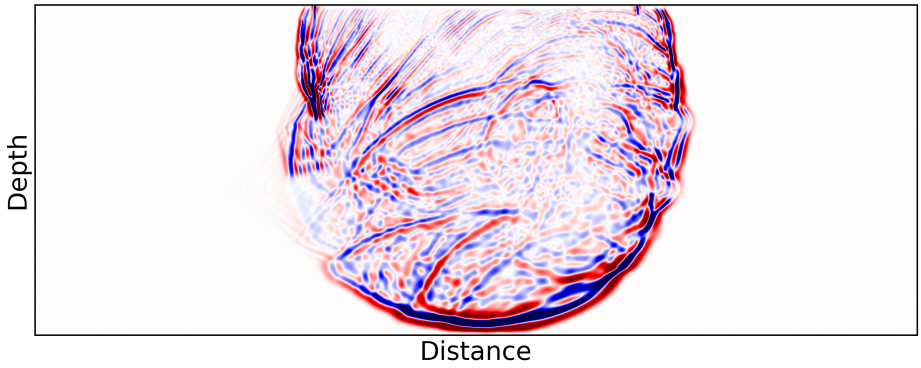
Given a velocity model $v(\boldsymbol{x})$ and a source term $s(\boldsymbol{x}, t)$, solving this equation numerically is called forward modeling. It produces a simulated wavefield, and sampling that wavefield at the receiver positions produces a synthetic shot gather. Writing $\boldsymbol{x}_s$ for the source position and $\boldsymbol{x}_r$ for the receiver position, the recorded data are the wavefield sampled at the receivers, $d(\boldsymbol{x}_r, t; \boldsymbol{x}_s) = u(\boldsymbol{x}_r, t; \boldsymbol{x}_s)$. Figure 5 illustrates this chain from a given model to a simulated wavefield to a recorded gather. More complete descriptions use the elastic wave equation, which also supports the dispersive surface-wave modes known as ground roll. The synthetic datasets in this benchmark are produced exactly in this way, by acoustic or elastic forward modeling on given velocity models.

This is the same data-generation paradigm as in AI-for-PDE benchmarks, where training and test data are numerical solutions of partial differential equations (Takamoto et al., 2022; Ohana et al., 2024). The synthetic component of SPBENCH is therefore directly comparable to AI-for-PDE data. The velocity model plays the role of the PDE coefficient field, the source term is the input, and the shot gather is a partial observation of the solution. Exploration seismology, however, also provides large amounts of real measured data governed by the same physics. SPBENCH contains both. Methods are trained separately within each setting and evaluated on PDE-simulated data and on real measured data under matched protocols, so the benchmark can also measure how rankings on simulation data relate to rankings on observed data (Section 4.2).

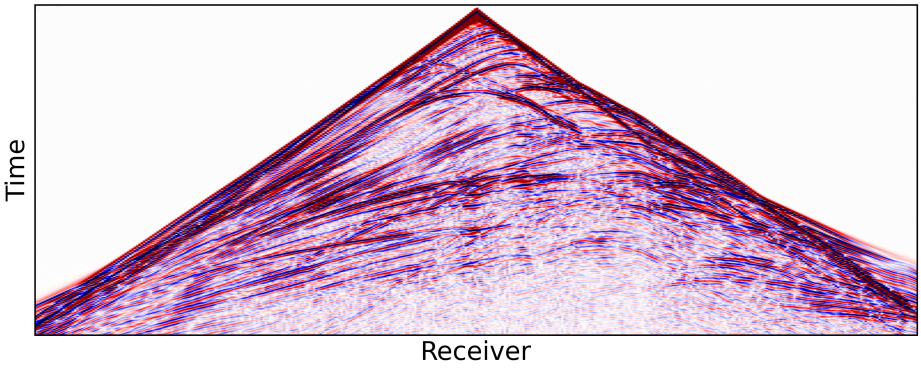
Each benchmark task is an operation on this wavefield. Random noise attenuation separates the deterministic wavefield from incoherent measurement noise. Trace interpolation restores a spatially undersampled wavefield. Ground-roll suppression removes the dispersive surface modes of the elastic equation that mask the weaker reflected body waves. Multiple suppression removes energy that has reflected more than once between interfaces. Deblending separates overlapping wavefields produced by simultaneous sources. First-arrival picking reads the traveltimes of the direct and refracted waves. Processing quality determines how much of the physics recorded in the data remains usable for imaging and inversion.



(a) Subsurface velocity model



(b) Wavefield snapshot



(c) Recorded shot gather

Figure 5: The chain from a subsurface model to recorded data. (a) A velocity model (Marmousi) specifies $v(\boldsymbol{x})$, the coefficient of the wave equation. (b) A numerical solution of the wave equation gives the propagating wavefield. (c) Surface receivers sample the wavefield over time, forming the shot gather that processing methods operate on.

B Literature Survey Scope and Selection Criteria

The reproducibility analysis in the Introduction and Figure 2 is based on a curated literature corpus. This appendix documents its scope, selection criteria, and annotation procedure.

Scope. We searched and screened papers published between January 2018 and May 2026 in nine English-language journals that regularly publish machine-learning research for exploration seismic data processing: *Geophysics*, *IEEE Transactions on Geoscience and Remote Sensing* (TGRS), *Petroleum Science*, *Computers & Geosciences*, *Journal of Geophysics and Engineering* (JGE), *IEEE Geoscience and Remote Sensing Letters* (GRSL), *Geophysical Prospecting*, *Journal of Seismic Exploration* (JSE), and *Journal of Applied Geophysics* (JAG), together with Chinese-language journals in the same field. Table 5 summarizes the resulting corpus composition.

Table 5: Composition of the 368-paper survey corpus by journal, with the share of each venue in the total.

Journal	Papers	Share
IEEE Transactions on Geoscience and Remote Sensing (TGRS)	130	35.3%
IEEE Geoscience and Remote Sensing Letters (GRSL)	63	17.1%
Geophysics	61	16.6%
Geophysical Prospecting	30	8.2%
Journal of Geophysics and Engineering (JGE)	20	5.4%
Journal of Seismic Exploration (JSE)	12	3.3%
Petroleum Science	8	2.2%
Computers & Geosciences	8	2.2%
Journal of Applied Geophysics (JAG)	4	1.1%
Chinese-language journals	32	8.7%
Total	368	100%

Selection criteria. We include papers that propose or evaluate supervised deep-learning methods for exploration seismic processing, covering random noise attenuation, trace interpolation, ground-roll noise suppression, multiple suppression, deblending, first-arrival picking, and closely related processing tasks. Supervised methods are defined as models trained on paired inputs and targets, such as degraded and reference data pairs or annotated arrival times. Papers based purely on unsupervised or self-supervised learning are outside the scope of this survey and are not counted. Each candidate paper was screened by title and abstract and then verified in full text.

Annotation. For every included paper we record the task labels, the dataset sources as multi-label categories (open synthetic, open field, private synthetic, private field), and whether a public code URL is provided. This procedure yields the 368-paper corpus reported in the Introduction, where Figure 2 summarizes the resulting dataset-source usage and code availability.

C Dataset Details

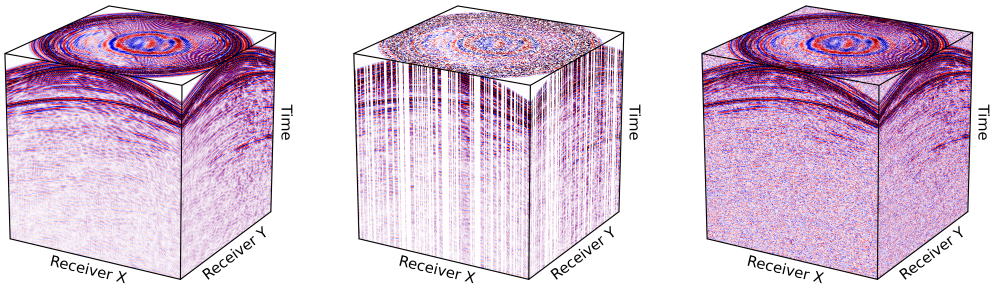
SEG C3 datasets. SEG C3 is an openly distributed synthetic dataset from the SEG open-data catalog.[2] It is a 45-shot subset of the SEG/EAGE salt model, originally extracted by Sandia National Laboratory for tests of the SALVO finite-difference imaging program and later reformatted by ARCO into uniform pre-stack SEG-Y files ordered by sail line. It is one of the most commonly used public datasets in seismic data processing research, particularly in interpolation and noise attenuation studies. Each shot is recorded on a 201×201 receiver grid with 625 time samples per trace at an 8 ms sampling interval. In SPBench it provides the base data for the interpolation and random-noise attenuation tasks. Example benchmark inputs built on SEG C3 are shown in Figure 6.

Mobil AVO datasets. The Mobil AVO Viking Graben Line 12 dataset is an openly distributed field dataset from the SEG open-data catalog (Madiba & McMechan, 2003). It is an east-west marine 2D line from the North Viking Graben in the North Sea, containing 1001 shot records, and was originally released for the 1994 SEG workshop on the comparison of seismic inversion methods. It has since become one of the most commonly used public field datasets in seismic processing, inversion, and AVO studies. In SPBench it provides the field-data base for the interpolation and random-noise attenuation tasks, and for a semi-synthetic field deblending setting. Figure 7 shows the field data and the derived benchmark inputs.

SPBench ground-roll datasets. The SPBench ground-roll datasets are new datasets for ground-roll noise suppression, comprising a synthetic set and a field set. The synthetic set is built on the SEG C3 dataset and contains nine shots, each recorded on a 201×201 receiver grid with 625 time samples per trace. Clean reflection data are modeled with the acoustic wave equation

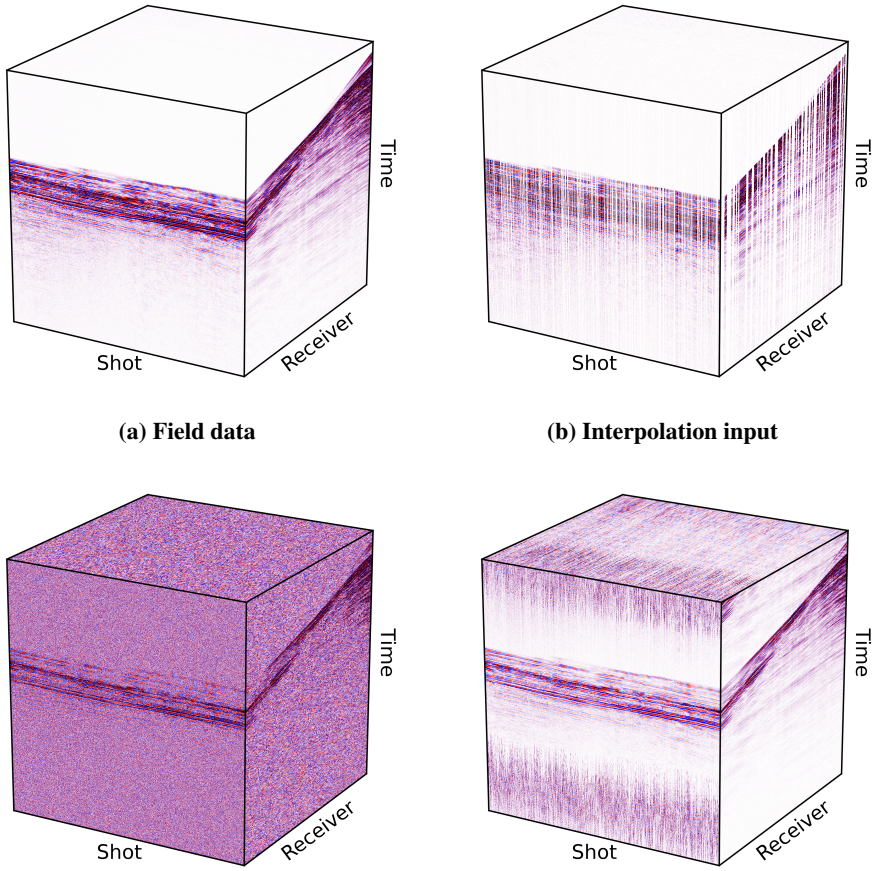
$$\frac{1}{c^2(\boldsymbol{x})}\frac{\partial^2 p}{\partial t^2} = \nabla^2 p + s,$$

[2]SEG C3 45-shot page: `https://wiki.seg.org/wiki/SEG_C3_45_shot`



(a) Reference **(b) Interpolation input** **(c) Random-noise input**

Figure 6: Example SEG C3 benchmark data. The clean reference data (a) are corrupted with missing traces to construct interpolation inputs (b) and with additive random noise to construct denoising inputs (c).



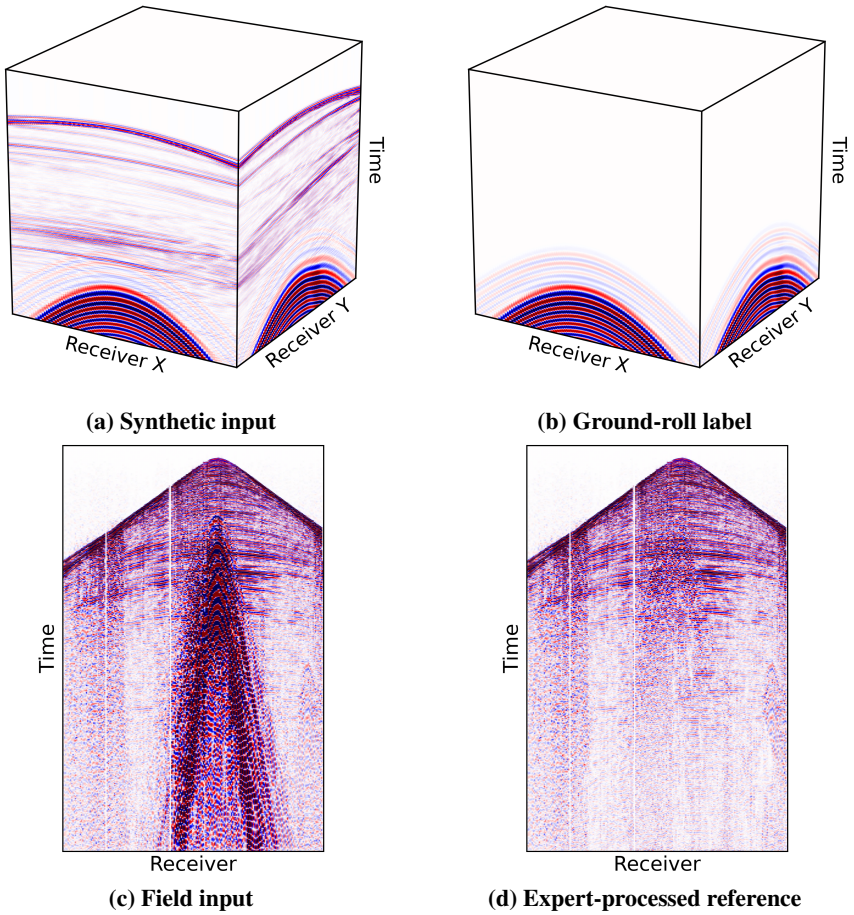
(a) Field data **(b) Interpolation input**

(c) Random-noise input **(d) Deblending input**

Figure 7: Example Mobil AVO benchmark data. The field records (a) serve as the base data for interpolation inputs with missing traces (b), random-noise inputs (c), and the semi-synthetic field deblending setting (d).

for the given survey geometry, where $c(\boldsymbol{x})$ is the wave speed, p the pressure field, and s the source term, and ground roll is modeled for the same geometry with the isotropic elastic wave equation

$$\rho \frac{\partial^2 \boldsymbol{u}}{\partial t^2} = (\lambda + \mu)\, \nabla(\nabla \cdot \boldsymbol{u}) + \mu \nabla^2 \boldsymbol{u},$$



(a) Synthetic input **(b) Ground-roll label**

(c) Field input **(d) Expert-processed reference**

Figure 8: Examples from the SPBENCH ground-roll datasets. Top row, synthetic set: the contaminated input (a) superposes modeled ground roll onto clean reflection data, and the ground-roll label (b) serves as the prediction target. Bottom row, field set: the field shot record (c) contains strong dispersive ground roll, and (d) is the reference produced by expert processing.

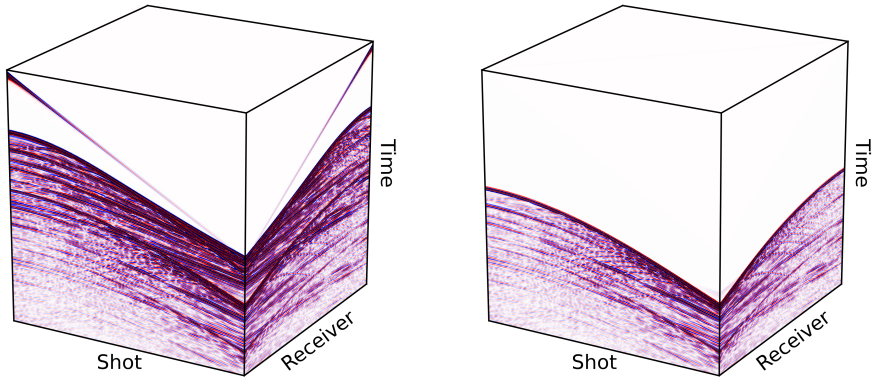
under a free-surface boundary condition, with density ρ, Lamé parameters λ and μ, and displacement $\boldsymbol{u}$. The elastic records are sorted by offset, and reflections are muted by exploiting their higher apparent velocity relative to the ground roll, yielding relatively pure ground-roll records. The traces are then restored to their original ordering, scaled to different amplitudes, and added to the clean acoustic data. The companion field set contains real acquisition records and extends the ground-roll task beyond synthetic conditions. Figure 8 shows example inputs and labels from the synthetic and field sets.

SPBENCH multiples dataset. The SPBENCH multiples dataset is a new synthetic dataset for multiple suppression, forward-modeled on the SEAM velocity model.[3] Two wavefields are simulated for each shot. The total data are modeled with a free surface, so they contain primaries, source and receiver ghosts, surface-related multiples, and the direct wave. The primary references are modeled without a free surface using the mirror source/receiver method: four acoustic simulations with mirrored source and receiver depths are superposed to synthesize the primary reflection plus source and receiver ghosts,

$$p = p_1 + p_4 - p_2 - p_3,$$

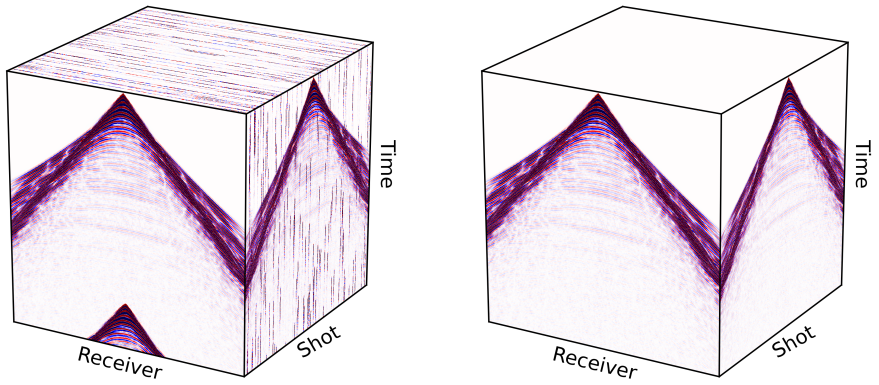
where p_1 is the primary reflection, p_2 and p_3 are the receiver-side and source-side ghosts, and p_4 is the source-plus-receiver ghost. The same combinations are used to compute the direct wave for the primary model. The true surface-related multiples are then obtained by subtracting the primaries

[3]SEG Advanced Modeling (SEAM) program: `https://www.seg.org/SEAM`



(a) Contaminated input **(b) Multiple label**

Figure 9: Example from the SPBENCH multiples dataset: a multiple-contaminated input gather (a) and the aligned additive multiple label (b) used as the prediction target.



(a) Blended input **(b) Unblended reference**

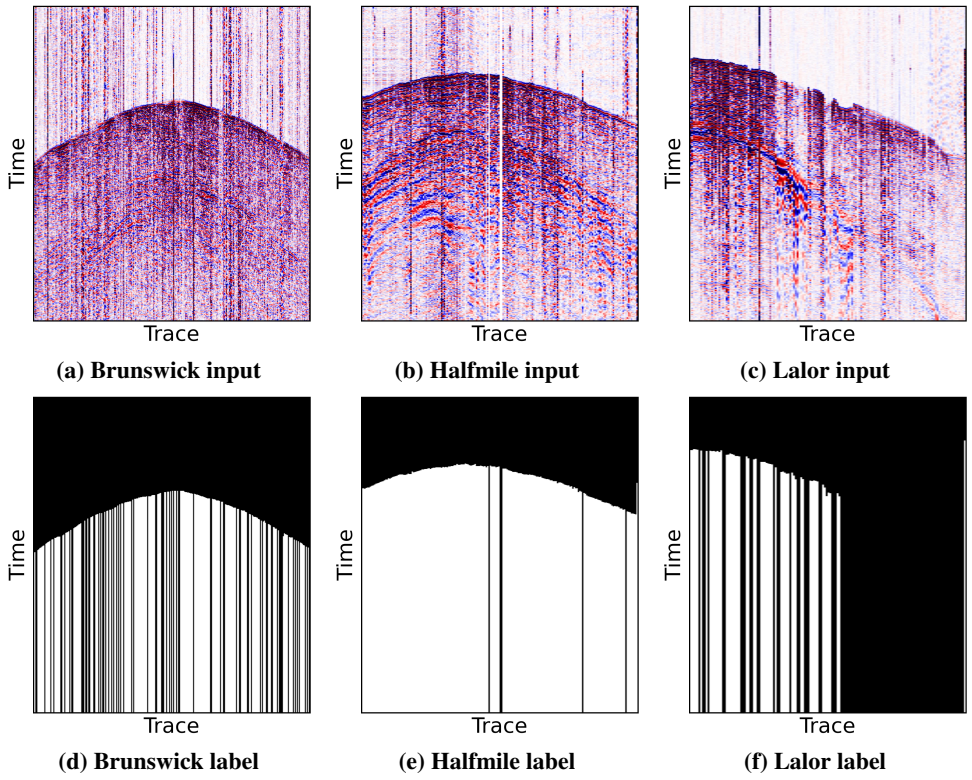
Figure 10: Example from the SPBENCH deblending datasets: a blended common-receiver gather (a) and the retained unblended reference (b).

from the total data after both have had their direct waves removed. The released set contains paired shot gathers with aligned additive multiple labels. Figure 9 shows an example input and its multiple label.

SPBENCH deblending datasets. The SPBENCH deblending datasets are new deblending datasets comprising a synthetic set and a field set. The synthetic set is constructed from single-shot seismic records generated by 3D acoustic forward modeling, containing 270 shots and 386 receivers with 2000 time samples per trace at a 2 ms sampling interval. Blended data are synthesized in the common-receiver domain by time-shifting and summing the single-source records at each receiver,

$$b(t;r) = \sum_{i=1}^{N_s} u(t - \tau_i; s_i; r),$$

where s_i and τ_i denote the position and firing time of the i-th source and N_s is the number of sources in the record. The original unblended records are retained as references, and all data are arranged as common-receiver gathers. The companion field set is a semi-synthetic field-data setting constructed by applying synthetic blending to real Mobil AVO Viking Graben Line 12 records, rather than a true simultaneous-source acquisition. Figure 10 shows an example blended input and its unblended reference.



(a) Brunswick input **(b) Halfmile input** **(c) Lalor input**

(d) Brunswick label **(e) Halfmile label** **(f) Lalor label**

Figure 11: Example first-arrival picking data. Top row: input shot gathers from the Brunswick, Halfmile, and Lalor surveys. Bottom row: the corresponding binary step-mask labels, which switch from zero to one at the annotated first arrival.

First-arrival picking datasets. The picking collection is built on HardPicks (St-Charles et al., 2024), a public benchmark of hardrock seismic reflection data collected from 3D surveys at mining sites in Canada and Finland, containing millions of traces with labeled first-break picks. SPBENCH uses its Brunswick, Halfmile, and Lalor surveys, which were acquired with dynamite sources, and evaluates them both per survey and as a combined set. Each input SEG-Y file is paired with a binary step mask $w(t) = H(t - t^*)$, where H is the Heaviside step function and t^* is the annotated first arrival, so the mask is zero before t^* and one from the arrival onward. Figure 11 shows example inputs and labels.

D REPRODUCED METHODS

This appendix lists the reproduced methods evaluated in each benchmark task. Four widely used architectures, U-Net, ResUNet, DnCNN, and attention-gated U-Net, are evaluated in every task as shared baselines. Beyond these baselines, the reproduced methods were selected from the task literature with clearly documented reproduction procedures. Table 6 summarizes them per task, with a short description derived from the source paper. Width-modified variants of the baselines (marked with an -L suffix on the leaderboard) follow the same architecture with increased base channels. Architecture notes, training schedules, and hyperparameters are recorded in the released configuration groups and the online artifact documentation.

Table 6: Reproduced methods per benchmark task beyond the four shared baselines.

Method	Description
Random noise attenuation and deblending	
CBD-RDN	Residual dense network adapted for seismic denoising and upscaling, combining local and global residual learning (Wang et al., 2022b).
FBResNet	Feedback residual network that iteratively refines strong-noise estimates (Liao et al., 2023).
cDDPM	Conditional denoising diffusion probabilistic model, evaluated on random-noise attenuation and adapted to blending-noise suppression (Li et al., 2024b).
QUNet	U-Net with quadratic neurons for richer feature representations, used for random-noise attenuation and transferred to deblending (Wang et al., 2025).
SCRN	Swin-Transformer convolutional residual network for joint denoising and interpolation, transferred to deblending (Gao et al., 2024).
U-Net++	Nested U-Net++ used within a two-step prediction strategy, transferred from denoising to deblending (Zhang et al., 2024).
Seismic interpolation	
SPNet	Sparse-prior network that integrates the POCS algorithm into a deep architecture for trace interpolation (Wu et al., 2024).
ANet	Attention-guided network with a hybrid loss for consecutively missing traces (Yu & Wu, 2022).
CA-Unet	Coordinate-attention U-Net for consecutive-missing interpolation (Li et al., 2022).
CFunet	Coarse-refine network with upsampling and a Fourier-domain loss (Park et al., 2022).
WRDL	Wavelet-based residual deep learning for seismic reconstruction (Liu et al., 2022).
Ground-roll noise suppression	
cDDPM	Conditional denoising diffusion probabilistic model for ground-roll attenuation (Li et al., 2024b).
Pix2Pix cGAN	Conditional GAN with a U-Net generator for ground-roll suppression (Yuan et al., 2020).
Physics CNN	Physics-constrained network combining learned and analytic separation filters (Pham & Li, 2022).
SANet	Soft-attention network for small-scale ground-roll attenuation (Yang et al., 2024).
Enhanced Atten-UNet	Attention U-Net with an adaptive frequency-modulation loss (Sun et al., 2024).
Multiple suppression	
DNNDAT	Deep neural network with data augmentation for marine multiple suppression (Wang et al., 2022a).
SAGAN	GAN with a U-Net generator, self-attention, and a Markovian discriminator (Tao et al., 2022).
First-arrival picking	
DSU-Net	U-Net with dynamic snake convolution for 2-D first-break picking (Wang et al., 2024).
HUNet	Uncertainty-aware picking network with prediction refinement (Pu et al., 2024).
STUNet	U-shaped network with Swin Transformer feature extraction (Jiang et al., 2023).

E Experiment Design

Seismic interpolation. Models take a masked shot gather as input and reconstruct the complete gather. Missing-trace experiments are organized into three families. Random missing removes traces at random positions, uniform missing removes traces at a fixed regular interval, and consecutive missing removes contiguous blocks of traces. Each family has fixed levels: 30%, 50%, and 70% for random missing, 50% and 75% for uniform missing, and gaps of 20, 30, and 40 traces for consecutive missing. All families run on both SEG C3 and Mobil AVO. The exact mask patterns and all parameter settings are available in the released code. Splits are made at the shot level to prevent leakage, and masks are generated only after the split is fixed. Observed traces are retained during evaluation. Most models train on patches of 256 time samples by 128 traces, or 256 by 120 on Mobil AVO, whose spatial extent is at most 120 traces. Data are normalized globally by the absolute maximum of the training set, and a few models use patch-wise normalization instead. Evaluation is performed on the full gather. Each run is executed with three random seeds. This task contributes 16 settings.

Random noise attenuation. Models take a noise-contaminated gather as input and predict the clean gather. Gaussian or Poisson noise is added to the clean data until the input SNR drops to -5, 0, or 5 dB, using experiment-specific random seeds, on SEG C3 and Mobil AVO. The exact noise realizations and all parameter settings are available in the released code. Shot-level train/validation/test partitions follow a 7:1:1 split to prevent leakage. Patching, normalization, and evaluation follow the interpolation protocol: most models train on patches of 256 time samples by 128 traces, or 256 by 120 on Mobil AVO, data are normalized globally by the absolute maximum of the training set, and

evaluation is performed on the full gather. Each run is executed with three random seeds. This task contributes 12 settings.

Ground-roll noise suppression. Models predict the additive ground-roll component. The estimate is subtracted from the contaminated input to recover the clean signal, and all metrics are computed on the recovered result. The synthetic settings scale the modeled ground roll to five amplitude multipliers, denoted as levels 1, 3, 5, 7, and 9. The SPBENCH field ground-roll set contributes a single setting with the same evaluation suite. Its labels are produced by expert processing rather than being ideal clean references, so results on this setting should be interpreted with care. Each run is executed with three random seeds. This task contributes 6 settings.

Multiple suppression. Models take a multiple-contaminated gather as input and estimate the additive multiple component, whose labels are obtained directly from physical modeling. Primary preservation can therefore be evaluated separately from noise removal by subtracting the estimate from the input. Shot gathers contain 638 traces and 2000 time samples, and models train on patches of 256 time samples by 512 traces. Shot-level train/validation/test partitions follow a 7:1:1 split, and data are normalized globally by the absolute maximum of the training set. Each run is executed with three random seeds. This task contributes a single setting.

Deblending. Models take a blended common-receiver gather as input and recover the unblended reference. Writing the blending process as $\boldsymbol{b} = \boldsymbol{B}\boldsymbol{U}$, the task is to recover the unblended single-source responses $\boldsymbol{U}$ from $\boldsymbol{b}$ given the blending operator $\boldsymbol{B}$. The synthetic settings cover three blending levels defined by the blending factor $bf \in \{0.2, 0.4, 0.6\}$, corresponding to weak, moderate, and strong blending. For each level, the firing schedule is generated recursively as $\tau_1 = 0$ and $\tau_{i+1} = \tau_i + \lfloor \epsilon_i \, n_t / bf \cdot d_t \rfloor$, with the firing times τ_i in milliseconds, n_t the number of time samples per trace, d_t the sampling interval, and ϵ_i a uniform random variate on $(0, 1)$, so a larger bf produces shorter average inter-shot intervals and stronger overlap. The Mobil AVO setting uses $bf = 5$ with a fixed 500-sample delay plus a random jitter of up to $n_t/5$ samples, mimicking streamer-style acquisition timing. Each configuration records the source combination, dither schedule, random seed, and split identifier so that every method is evaluated on identical interference patterns. Each run is executed with three random seeds. This task contributes 4 settings.

First-arrival picking. Models take a receiver-line patch as input and predict a binary step mask, from which the first-arrival pick is read at the mask transition. FFID-level splitting prevents gather leakage, and receiver-line segmentation prevents patches from crossing distant acquisition lines. All models use the same 128×512 patch geometry, and hit rates are reported at tolerances of 1, 3, 5, 7, and 9 samples. Each run is executed with three random seeds. This task contributes 4 settings.

Training schedules, optimizers, and stopping rules for every run are recorded per configuration group in the released artifacts.

F EVALUATION METRICS

Rank correlation. To compare the rankings produced by different metrics or data domains, we use the Kendall rank correlation

$$\tau = \frac{n_c - n_d}{\binom{n}{2}},$$

where n_c and n_d count concordant and discordant pairs among n ranked methods. τ ranges from -1 (fully reversed rankings) through 0 (no association) to $+1$ (identical rankings). This is the no-ties form. Where the compared scores contain ties, as in the per-survey model rankings of Section 4.4, we use the tie-corrected form (Kendall τ_b) computed by standard libraries, which reduces to the no-ties form on continuous scores. Sections 4.2, 4.3, and 4.4 use it to quantify ranking agreement between settings and between metrics.

All metrics are computed between a prediction p and a reference r, both treated as vectors in $\mathbb{R}^N$ with samples p_i and r_i. Each metric is first evaluated per test record, then averaged over records, and finally over the three random seeds of each run, so every reported value is a mean over seeds. Unless noted otherwise, larger values indicate better recovery.

Signal-to-noise ratios. The output signal-to-noise ratio, the standard global fidelity measure in seismic processing,

$$\mathrm{SNR} = 10 \log_{10} \frac{\|r\|_2^2}{\|p - r\|_2^2} \ (\mathrm{dB}),$$

compares the energy of the reference with the energy of the residual error. Higher is better, and every $+6$ dB roughly halves the residual amplitude. The peak signal-to-noise ratio

$$\mathrm{PSNR} = 10 \log_{10} \frac{N \, \mathrm{MAX}^2}{\|p - r\|_2^2} \ (\mathrm{dB})$$

normalizes the residual energy by the peak absolute amplitude $\mathrm{MAX} = \max_i |r_i|$ of the reference, so it is more comparable across datasets with different amplitude scales; higher is better. Since $\mathrm{MSE} = \|p - r\|_2^2 / N$, this equals the common form $10 \log_{10}(\mathrm{MAX}^2/\mathrm{MSE})$. Within one setting, MAX is fixed by the reference data, so PSNR and SNR differ only by a constant and produce identical rankings. The result tables therefore report SNR, and the PSNR values are available in the released results.

Pointwise error metrics. The mean squared error $\mathrm{MSE} = \frac{1}{N}\|p-r\|_2^2 = \frac{1}{N}\sum_i (p_i - r_i)^2$ measures the average squared deviation between prediction and reference, and the root mean squared error $\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$ expresses it on the signal-amplitude scale; smaller is better for both. Both weight errors by squared amplitude, so they are dominated by strong events. The mean absolute error $\mathrm{MAE} = \frac{1}{N}\sum_i |p_i - r_i|$ measures the average absolute deviation, weights all errors linearly, and is less sensitive to large outliers than MSE; smaller is better.

Structural similarity. The structural similarity index (SSIM) compares prediction and reference through luminance, contrast, and structure terms,

$$\mathrm{SSIM}(p, r) = \frac{(2\mu_p\mu_r + c_1)(2\sigma_{pr} + c_2)}{(\mu_p^2 + \mu_r^2 + c_1)(\sigma_p^2 + \sigma_r^2 + c_2)},$$

where μ_p and μ_r denote local means, σ_p^2 and σ_r^2 local variances, and σ_{pr} the local covariance, and c_1, c_2 are small stabilizing constants. SSIM is upper-bounded by 1, with larger values indicating greater structural similarity. It can be negative on sign-alternating seismic amplitudes, and it captures perceptual and structural fidelity rather than pointwise error, so it can rank models differently from SNR.

Component-resolved fidelity construction (SCoRE). FB-FRE partitions the reference by frequency. Given a reference shot record r, prediction p, and sampling interval dt, we compute the real Fourier transform along the time axis and estimate the average reference power spectrum

$$P(f_k) = \frac{1}{N_{\mathrm{avg}}} \sum_{\text{non-time axes}} |\mathrm{RFFT}(r)(f_k)|^2, \quad f_k = \mathrm{rfftfreq}(N_t, d = dt).$$

Here N_t is the number of time samples per trace, and the average runs over the N_{avg} non-time slices of the record. The effective frequency band is defined by the frequencies whose power exceeds a relative threshold $\eta = 10^{-3}$ of the peak power:

$$f_{\min} = \min\{f_k : P(f_k) \geq \eta \max_f P(f)\}, \quad f_{\max} = \max\{f_k : P(f_k) \geq \eta \max_f P(f)\}.$$

This band is divided into four contiguous subbands, low, mid, high, and very high, using relative widths $(0.20, 0.30, 0.30, 0.20)$. The band content and the per-band scores follow Section 3.2, and the leaderboard columns labeled FB-FRE report the per-band partition SNR, $\mathrm{SNR}_{\mathcal{S}}$ of Section 3.2.

EB-FE partitions the reference by local energy. It builds a smoothed reference-energy map

$$E_i = \sqrt{\mathrm{gaussian_filter}(r_i^2, \sigma)},$$

removes strictly zero-reference samples, and sorts the remaining N_{valid} samples by E_i. The filter width σ is recorded in the released configuration. For an energy-percentile bin $\mathcal{B}(p_L, p_H)$, samples whose sorted ranks fall in

$$\left[N_{\mathrm{valid}} \frac{p_L}{100}, \ N_{\mathrm{valid}} \frac{p_H}{100}\right)$$

are evaluated separately. The default bins are very weak $(5, 20)$, weak $(20, 40)$, medium $(40, 70)$, and strong $(70, 100)$. The leaderboard columns labeled EB-FE report the per-bin partition SNR, likewise $\mathrm{SNR}_{\mathcal{S}}$. The main-text tables report the medium and strong bins, and all bins are available in the released metric summaries. The released metric summaries additionally report per-partition normalized errors, $\mathrm{NE}_b = \|p_b - r_b\|_2/(\|r_b\|_2 + \varepsilon)$ and the analogously defined $\mathrm{NE}_{\mathcal{B}}$, and the reference energy ratio $\rho_b = \|r_b\|_2^2/(\|r\|_2^2 + \varepsilon)$.

Partition-ranking agreement with SNR. For each task, we compute the Kendall rank correlation between the SNR ranking and each partition-SNR ranking and average it over the task's settings (Table 7). The agreement is moderate to strong but never perfect, so the partition rankings carry information that the aggregate score does not. The very-weak and weak energy bins are excluded from this comparison because their low reference energy makes the partition SNR unreliable.

Table 7: Mean Kendall rank correlation between the SNR ranking and each SCoRE partition ranking, averaged over the settings of each task.

Partition	Noise	Interp.	Ground roll	Multiple	Deblend
FB-FRE Low	0.752	0.699	0.848	0.500	0.644
FB-FRE Mid	0.854	0.820	0.838	0.944	0.900
FB-FRE High	0.818	0.800	0.696	0.944	0.944
FB-FRE VHigh	0.800	0.521	0.740	0.722	0.489
EB-FE Med	0.738	0.505	0.791	0.889	0.822
EB-FE Strong	0.716	0.900	0.922	0.722	0.878

Picking metrics. For first-arrival picking, let t_i^* be the reference pick on trace i and $\hat{t}_i$ the predicted pick, both in samples, over N_{tr} traces, and write $\Delta_i = \hat{t}_i - t_i^*$ for the signed pick error. The mean absolute error

$$\mathrm{MAE} = \frac{1}{N_{\mathrm{tr}}} \sum_{i=1}^{N_{\mathrm{tr}}} |\Delta_i|$$

measures the average pick displacement in samples, and multiplying by the sampling interval converts it to milliseconds; smaller is better. The root mean squared pick error follows as $\mathrm{RMSE}_{\mathrm{pick}} = \sqrt{\frac{1}{N_{\mathrm{tr}}} \sum_i (\hat{t}_i - t_i^*)^2}$, also in samples; smaller is better. The mean bias error $\mathrm{MBE} = \frac{1}{N_{\mathrm{tr}}} \sum_i \Delta_i$ preserves the sign and therefore reveals systematic early or late picking that absolute errors cancel out. The hit rate at tolerance k,

$$\mathrm{H@}k = \frac{\#\{i : |\Delta_i| < k\}}{N_{\mathrm{tr}}}, \qquad k \in \{1, 3, 5, 7, 9\},$$

is the fraction of picks with an error below k samples, where $\#\{\cdot\}$ counts the traces satisfying the condition; higher is better, and the curve $\mathrm{H@}k$ versus k shows how quickly picks concentrate around the reference. For detection quality, precision, recall, and F1 are computed pixel-wise between the predicted binary step mask and the reference mask. With TP the number of samples where both masks equal one, and FP and FN defined accordingly, $\mathrm{P} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP})$, $\mathrm{R} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$, and $\mathrm{F1} = 2\mathrm{PR}/(\mathrm{P} + \mathrm{R})$; higher is better. This mask-level F1 measures overlap of the predicted step mask and is not comparable to a pick-level F1 at a fixed error tolerance, which would coincide with the corresponding hit rate.

Ridge-continuity construction. The model outputs a first-arrival probability $p_i(s)$ for each trace i at time sample s. With the threshold $\gamma = 0.5$, the predicted pick in samples is $\hat{s}_i = \min\{s : p_i(s) \geq \gamma\}$, the same quantity denoted $\hat{t}_i$ in the pick-error metrics above, and traces without any sample above the threshold are excluded. Let $\mathcal{V}$ denote the set of valid picks and $N_v = |\mathcal{V}|$. Valid picks are sorted by absolute source-receiver offset x_i. A robust offset-time ridge is then estimated in three steps. First, the offset range is divided into $B = \max(3, \min(50, \lfloor N_v/5 \rfloor, N_v))$ equal bins, and each bin is represented by the median pick time of its members, so that isolated gross mispicks do not bend the ridge. Second, empty bins are filled by linear interpolation from adjacent valid bins. Third, the bin sequence is smoothed by a moving average whose window is $0.07B$, adjusted to an odd number not exceeding B, which suppresses high-frequency fluctuations caused by the finite sample count, and the smoothed ridge is recentered so that the median residual over the valid picks

is zero. Let $\{(\tilde{x}_j, \tilde{t}_j)\}_{j=1}^{B}$ denote the resulting ridge at the offset-bin centers. Its discrete slope and slope change are

$$d_j = \frac{\tilde{t}_j - \tilde{t}_{j-1}}{\tilde{x}_j - \tilde{x}_{j-1}}, \qquad j = 2, \ldots, B, \qquad c_j = \frac{d_j - d_{j-1}}{\tilde{x}_j - \tilde{x}_{j-1}}, \qquad j = 3, \ldots, B,$$

and the ridge curvature is the mean absolute slope change, $\mathrm{RC} = \frac{1}{B-2} \sum_{j=3}^{B} |c_j|$. This quantity is a discrete second derivative of travel time with respect to offset and carries units of time per offset squared. The normalized ridge curvature removes first-order differences in spatial and temporal scale through $\mathrm{RC}_{\mathrm{norm}} = \mathrm{RC}\, X_{\mathrm{range}}^2 / T_{\mathrm{range}}$, where $X_{\mathrm{range}} = \max_i x_i - \min_i x_i$ and $T_{\mathrm{range}} = \max_j \tilde{t}_j - \min_j \tilde{t}_j$, which makes $\mathrm{RC}_{\mathrm{norm}}$ dimensionless. Under rescalings $x' = \alpha x$ and $t' = \beta t$, the slopes scale as $d'_j = (\beta/\alpha)\, d_j$ and the slope changes as $c'_j = (\beta/\alpha^2)\, c_j$, so RC changes by $|\beta|/\alpha^2$ while the factor $X_{\mathrm{range}}^2 / T_{\mathrm{range}}$ changes by $\alpha^2/|\beta|$, and $\mathrm{RC}_{\mathrm{norm}}$ is invariant. A numerically flat ridge, with $T_{\mathrm{range}} \le \varepsilon_t$ and $\mathrm{RC} = 0$, is assigned $\mathrm{RC}_{\mathrm{norm}} = 0$, where ε_t is a numerical zero tolerance scaled to the sampling interval.

G Validation of the Reference-Free Ridge Score

Systematic-shift failure mode. One important counterexample is a global systematic shift. If every predicted arrival is shifted by a constant b, $\hat{s}'_i = \hat{s}_i + b$, the recentered ridge shifts by the same b, so the residual statistics and the ridge curvature remain unchanged. A smooth curve that consistently follows the wrong event can therefore still obtain good scores. This failure mode can only be identified through manual labels, waveform evidence, or other physical constraints.

Tolerance sweep of the ridge-score agreement. Table 8 extends the validation of Section 4.4 to all five released hit-rate tolerances. The agreement with the labeled rankings rises monotonically with the tolerance, from 0.244 at H@1 to 0.862 at H@9, approaching the MAE-based agreement of 0.881. Strict tolerances penalize every sample of deviation, including a constant offset shared by all traces, while the ridge score is invariant to such shifts. The gap between the strict-tolerance rows and the MAE row therefore quantifies the systematic-shift blindness described above. The main text reports H@5 as the middle tolerance of the released hit-rate set.

Table 8: Kendall rank correlation of the reference-free ridge score with labeled accuracy rankings and between its two aggregation rules, per field picking survey over the ten evaluated methods.

Ranking pair	Brunswick	Halfmile	Lalor	Mean
$\mathrm{RC}_{\mathrm{norm}}$ vs pick error (MAE)	0.778	0.911	0.956	0.881
$\mathrm{RC}_{\mathrm{norm}}$ vs 1−H@1	0.244	0.422	0.067	0.244
$\mathrm{RC}_{\mathrm{norm}}$ vs 1−H@3	0.556	0.644	0.333	0.511
$\mathrm{RC}_{\mathrm{norm}}$ vs 1−H@5	0.689	0.899	0.822	0.803
$\mathrm{RC}_{\mathrm{norm}}$ vs 1−H@7	0.719	0.911	0.911	0.847
$\mathrm{RC}_{\mathrm{norm}}$ vs 1−H@9	0.764	0.911	0.911	0.862
Shot macro vs trace weighted	0.956	0.956	1.000	0.970

Aggregation choices. The main-text scores use shot-level macro averaging, and weighting by the valid trace count per shot instead produces nearly identical model rankings (Table 8, last row).

H Complete Leaderboard Results

The following tables report 482 result records across all benchmark settings. The remaining 13 records of architecture variants are omitted from the tables and included in the released results. Reconstruction rows are ordered by increasing output SNR, and first-arrival rows are ordered by increasing MAE. All scores are means over three random seeds with standard deviations shown in small type. PSNR differs from SNR by a per-setting constant and yields identical rankings, so it is omitted from the tables and reported in the released results. A dash indicates that the record does not report that field. Cell shading is computed independently within each setting and metric. Per-seed results and complete machine-readable metric dictionaries are available on the project website.

H.1 RANDOM-NOISE ATTENUATION

Table 9: SEGC3 Random Noise Gaussian SNR -5 dB.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
FBResNet	0.45M	10.35±0.06	9.09±0.08	12.46±0.07	9.55±0.05	1.01±0.04	0.17±0.04	11.68±0.08
cDDPM	33.3M	10.41±0.07	9.92±0.15	12.78±0.06	9.92±0.06	0.89±0.08	-0.72±0.04	12.65±0.10
ResUNet	8.11M	10.50±0.12	10.07±0.30	12.95±0.22	9.62±0.08	0.66±0.06	-0.05±0.04	11.89±0.18
UNet	7.76M	10.72±0.16	10.08±0.33	13.19±0.20	9.82±0.14	0.81±0.10	0.12±0.04	12.12±0.21
Attention UNet	7.85M	10.76±0.04	10.08±0.07	13.22±0.08	9.89±0.03	0.82±0.01	0.15±0.04	12.16±0.05
UNet-L	31.04M	10.87±0.03	10.57±0.05	13.30±0.03	9.94±0.04	0.94±0.03	0.17±0.01	12.31±0.05
QUNet	21.88M	11.01±0.21	10.70±0.30	13.34±0.24	10.16±0.20	1.11±0.14	0.17±0.06	12.52±0.28
UNet++	9.05M	11.04±0.14	10.55±0.26	13.21±0.14	10.30±0.07	1.30±0.04	0.15±0.13	12.57±0.15
DnCNN	0.56M	11.53±0.03	11.09±0.03	13.62±0.06	10.73±0.03	1.79±0.03	0.30±0.08	13.28±0.03
CBD-RDN	1.49M	**11.88±0.05**	**11.80±0.09**	**13.89±0.05**	**10.87±0.04**	**2.32±0.06**	**0.57±0.01**	**13.67±0.06**
SCRN	0.43M	**11.99±0.00**	**11.68±0.05**	**13.96±0.01**	**11.14±0.01**	**2.30±0.02**	**0.65±0.01**	**13.79±0.00**

Table 10: SEGC3 Random Noise Gaussian SNR 0 dB.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
cDDPM	33.3M	12.38±0.04	11.58±0.09	15.27±0.09	11.95±0.05	2.44±0.07	0.55±0.03	15.57±0.11
ResUNet	8.11M	12.56±0.11	12.05±0.17	15.57±0.08	11.73±0.10	1.90±0.14	1.04±0.12	14.55±0.11
UNet	7.76M	12.73±0.20	11.99±0.32	15.69±0.10	11.90±0.19	2.09±0.24	1.17±0.21	14.73±0.22
Attention UNet	7.85M	12.80±0.10	12.00±0.15	15.73±0.07	11.98±0.08	2.18±0.10	1.26±0.09	14.82±0.10
FBResNet	0.45M	12.97±0.00	11.99±0.04	15.45±0.01	12.12±0.02	2.91±0.01	1.47±0.02	15.00±0.01
QUNet	21.88M	13.08±0.08	12.59±0.16	15.81±0.04	12.18±0.05	2.61±0.05	1.34±0.06	15.19±0.09
UNet-L	31.04M	13.13±0.04	12.78±0.06	15.91±0.06	12.23±0.04	2.60±0.06	1.44±0.03	15.25±0.05
UNet++	9.05M	13.47±0.04	12.78±0.06	15.90±0.04	12.62±0.04	3.26±0.08	1.74±0.09	15.63±0.03
CBD-RDN	1.49M	13.58±0.45	13.23±0.59	15.92±0.37	12.57±0.42	3.53±0.53	1.67±0.24	16.02±0.53
DnCNN	0.56M	**14.10±0.06**	**13.75±0.13**	**16.40±0.08**	**13.16±0.05**	**3.93±0.03**	**1.91±0.05**	**16.67±0.11**
SCRN	0.43M	**14.38±0.01**	**14.01±0.00**	**16.60±0.02**	**13.40±0.01**	**4.29±0.01**	**2.19±0.00**	**16.92±0.01**

Table 11: SEGC3 Random Noise Gaussian SNR 5 dB.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
cDDPM	33.3M	14.31±0.23	13.01±0.29	17.87±0.14	14.00±0.18	4.26±0.17	2.35±0.17	18.56±0.14
UNet	7.76M	14.84±0.44	13.78±0.59	18.19±0.35	13.96±0.42	3.96±0.45	3.04±0.43	17.30±0.48
Attention UNet	7.85M	14.89±0.23	13.61±0.29	18.21±0.18	14.08±0.19	4.08±0.27	3.09±0.20	17.38±0.24
ResUNet	8.11M	14.92±0.05	14.15±0.11	18.47±0.03	14.05±0.03	3.98±0.04	3.18±0.07	17.37±0.04
FBResNet	0.45M	15.84±0.02	14.76±0.04	18.65±0.04	14.87±0.00	5.43±0.01	3.72±0.00	18.56±0.02
UNet-L	31.04M	15.85±0.07	15.22±0.13	18.88±0.14	14.77±0.06	5.18±0.04	3.79±0.06	18.55±0.10
QUNet	21.88M	15.90±0.15	15.21±0.33	18.98±0.14	14.83±0.10	5.16±0.18	3.80±0.12	18.63±0.23
UNet++	9.05M	16.19±0.05	15.08±0.14	18.89±0.01	15.18±0.06	5.76±0.09	4.16±0.11	18.86±0.02
CBD-RDN	1.49M	16.46±0.47	16.00±0.58	19.05±0.40	15.32±0.43	6.26±0.56	3.99±0.37	19.65±0.53
DnCNN	0.56M	**16.90±0.02**	**16.48±0.04**	**19.53±0.03**	**15.73±0.02**	**6.43±0.02**	**4.32±0.03**	**20.15±0.01**
SCRN	0.43M	**17.06±0.01**	**16.64±0.02**	**19.64±0.02**	**15.87±0.02**	**6.59±0.03**	**4.57±0.03**	**20.24±0.01**

Table 12: SEGC3 Random Noise Poisson SNR -5 dB.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
FBResNet	0.45M	10.37±0.04	9.17±0.04	12.49±0.06	9.59±0.04	0.99±0.01	0.15±0.02	11.72±0.07
cDDPM	33.3M	10.43±0.06	9.92±0.11	12.75±0.04	10.00±0.05	0.93±0.06	-0.69±0.05	12.64±0.06
ResUNet	8.11M	10.53±0.17	10.10±0.36	12.96±0.27	9.66±0.12	0.65±0.08	-0.08±0.02	11.94±0.23
UNet	7.76M	10.72±0.17	10.06±0.39	13.19±0.17	9.83±0.16	0.78±0.15	0.06±0.10	12.14±0.22
Attention UNet	7.85M	10.84±0.07	10.19±0.14	13.30±0.10	9.98±0.08	0.85±0.03	0.15±0.05	12.28±0.11
UNet-L	31.04M	10.93±0.05	10.63±0.16	13.24±0.07	10.04±0.04	1.05±0.06	0.13±0.03	12.43±0.07
QUNet	21.88M	10.94±0.08	10.52±0.13	13.34±0.16	10.10±0.09	1.02±0.02	0.09±0.05	12.44±0.15
UNet++	9.05M	10.98±0.09	10.51±0.17	13.22±0.11	10.25±0.04	1.23±0.03	0.05±0.14	12.51±0.09
DnCNN	0.56M	**11.53±0.05**	**11.10±0.02**	**13.65±0.01**	**10.71±0.06**	**1.70±0.06**	**0.27±0.08**	**13.33±0.05**
SCRN	0.43M	**11.99±0.01**	**11.62±0.06**	**13.97±0.03**	**11.15±0.01**	**2.29±0.01**	**0.65±0.02**	**13.80±0.01**

Table 13: SEGC3 Random Noise Poisson SNR 0 dB.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
cDDPM	33.3M	12.37±0.02	11.49±0.02	15.24±0.00	11.97±0.03	2.46±0.02	0.54±0.02	15.53±0.01
ResUNet	8.11M	12.57±0.09	12.04±0.16	15.57±0.08	11.75±0.09	1.91±0.13	1.01±0.09	14.58±0.10
UNet	7.76M	12.73±0.19	11.99±0.31	15.66±0.09	11.91±0.19	2.08±0.24	1.14±0.23	14.74±0.21
Attention UNet	7.85M	12.80±0.12	12.01±0.16	15.74±0.08	11.98±0.11	2.17±0.13	1.24±0.09	14.83±0.14
FBResNet	0.45M	12.98±0.02	11.97±0.06	15.45±0.03	12.14±0.01	2.91±0.01	1.46±0.01	15.02±0.03
QUNet	21.88M	13.03±0.06	12.50±0.08	15.81±0.05	12.17±0.03	2.55±0.07	1.29±0.08	15.14±0.07
UNet-L	31.04M	13.13±0.07	12.80±0.08	15.90±0.02	12.23±0.06	2.62±0.17	1.43±0.04	15.25±0.08
UNet++	9.05M	13.49±0.08	12.75±0.16	15.85±0.02	12.65±0.08	3.36±0.09	1.80±0.13	15.63±0.06
DnCNN	0.56M	**14.11±0.03**	**13.71±0.04**	**16.40±0.06**	**13.19±0.02**	**3.93±0.04**	**1.87±0.02**	**16.68±0.05**
SCRN	0.43M	**14.38±0.01**	**14.01±0.04**	**16.61±0.01**	**13.41±0.01**	**4.26±0.02**	**2.19±0.01**	**16.94±0.01**

Table 14: SEGC3 Random Noise Poisson SNR 5 dB.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
cDDPM	33.3M	14.36±0.09	13.03±0.14	17.89±0.07	14.03±0.07	4.30±0.06	2.38±0.03	18.60±0.06
UNet	7.76M	14.89±0.37	13.81±0.49	18.26±0.26	14.00±0.37	4.01±0.39	3.11±0.34	17.35±0.41
Attention UNet	7.85M	14.92±0.22	13.60±0.29	18.24±0.17	14.11±0.18	4.13±0.25	3.10±0.19	17.41±0.23
ResUNet	8.11M	14.92±0.05	14.14±0.10	18.47±0.03	14.05±0.02	3.96±0.05	3.17±0.07	17.38±0.04
QUNet	21.88M	15.77±0.04	14.97±0.05	18.88±0.02	14.72±0.05	5.01±0.03	3.74±0.02	18.45±0.01
UNet-L	31.04M	15.82±0.05	15.13±0.06	18.87±0.11	14.79±0.06	5.13±0.02	3.79±0.04	18.50±0.06
FBResNet	0.45M	15.86±0.01	14.78±0.07	18.67±0.04	14.89±0.01	5.42±0.03	3.71±0.02	18.60±0.02
UNet++	9.05M	16.16±0.07	15.11±0.19	18.94±0.06	15.14±0.05	5.68±0.02	4.13±0.12	18.85±0.05
DnCNN	0.56M	**16.94±0.03**	**16.56±0.12**	**19.57±0.02**	**15.78±0.01**	**6.45±0.03**	**4.36±0.06**	**20.22±0.05**
SCRN	0.43M	**17.07±0.01**	**16.65±0.05**	**19.65±0.02**	**15.88±0.01**	**6.60±0.01**	**4.58±0.01**	**20.26±0.01**

Table 15: Mobile AVO Random Noise Gaussian SNR -5 dB.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
cDDPM	33.3M	7.16±0.16	9.75±0.12	9.55±0.13	6.83±0.18	-11.45±0.25	-15.45±0.16	10.26±0.10
ResUNet	8.11M	8.01±0.00	9.51±0.01	9.78±0.01	7.10±0.01	-9.12±0.01	-10.31±0.00	8.76±0.00
UNet	7.76M	8.94±0.00	9.12±0.01	10.22±0.00	8.62±0.01	-5.21±0.01	-7.65±0.00	9.40±0.00
DnCNN	0.56M	8.97±0.00	9.47±0.01	10.55±0.00	8.73±0.00	-7.44±0.01	-9.81±0.00	9.79±0.00
UNet++	9.05M	9.23±0.11	9.77±0.15	10.12±0.13	7.89±0.24	**-2.73±0.22**	-2.34±0.83	9.32±0.14
Attention UNet	7.85M	9.27±0.00	10.00±0.01	10.45±0.01	8.44±0.01	-5.77±0.00	-3.04±0.01	9.41±0.00
FBResNet	0.45M	9.59±0.00	9.51±0.01	10.78±0.01	**9.31±0.01**	-4.63±0.01	-7.50±0.00	10.10±0.00
QUNet	21.88M	9.70±0.48	9.65±0.36	10.77±0.43	9.13±0.64	-3.92±0.74	-2.34±1.00	9.82±0.50
UNet-L	31.04M	**10.34±0.00**	**10.97±0.01**	**11.27±0.01**	9.05±0.01	-3.29±0.01	**-2.03±0.01**	**10.44±0.00**
SCRN	0.43M	**12.73±0.00**	**12.93±0.01**	**13.05±0.01**	**12.21±0.02**	**0.52±0.01**	**-1.42±0.01**	**12.88±0.00**

Table 16: Mobile AVO Random Noise Gaussian SNR 0 dB.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
cDDPM	33.3M	7.35±0.13	10.54±0.11	10.18±0.10	6.90±0.13	-12.21±0.15	-15.96±0.11	11.52±0.12
ResUNet	8.11M	9.71±0.00	11.77±0.00	11.75±0.01	8.88±0.01	-8.47±0.01	-8.65±0.00	10.44±0.00
UNet	7.76M	10.51±0.00	10.52±0.01	12.05±0.01	10.61±0.01	-4.23±0.01	-5.39±0.01	10.89±0.00
Attention UNet	7.85M	11.14±0.00	12.44±0.00	12.71±0.00	10.10±0.01	-5.89±0.01	-2.85±0.01	11.32±0.00
DnCNN	0.56M	11.95±0.00	12.76±0.01	13.48±0.00	11.42±0.01	-4.65±0.01	-6.59±0.00	12.68±0.00
QUNet	21.88M	12.07±0.36	12.77±0.55	13.26±0.28	10.95±0.59	-3.05±0.77	-1.43±0.16	12.20±0.37
UNet-L	31.04M	12.24±0.00	12.86±0.00	13.50±0.01	10.78±0.01	-2.56±0.00	**-1.10±0.00**	12.36±0.00
FBResNet	0.45M	12.34±0.00	12.63±0.01	**13.55±0.00**	**11.78±0.01**	-2.68±0.01	-5.35±0.00	**12.92±0.00**
UNet++	9.05M	**12.39±0.30**	**13.20±0.39**	13.41±0.34	10.90±0.41	**-1.27±0.61**	-1.59±0.55	12.55±0.33
SCRN	0.43M	**15.10±0.00**	**15.60±0.00**	**15.49±0.00**	**14.32±0.01**	**1.13±0.00**	**-0.94±0.00**	**15.33±0.00**

Table 17: Mobile AVO Random Noise Gaussian SNR +5 dB.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
cDDPM	33.3M	7.68±0.18	11.19±0.17	10.92±0.16	7.29±0.18	-12.22±0.19	-15.99±0.14	12.58±0.17
ResUNet	8.11M	10.83±0.00	13.38±0.00	13.23±0.00	10.21±0.01	-8.46±0.01	-6.49±0.00	11.38±0.00
Attention UNet	7.85M	12.35±0.01	14.06±0.00	14.24±0.00	11.32±0.01	-5.41±0.00	-2.29±0.00	12.56±0.01
UNet	7.76M	12.72±0.00	13.35±0.01	14.36±0.00	12.46±0.00	-3.54±0.01	-3.41±0.00	13.10±0.00
QUNet	21.88M	14.15±0.76	15.65±0.62	15.68±0.53	12.99±0.64	-2.98±1.39	-2.42±2.30	14.55±0.49
DnCNN	0.56M	15.10±0.00	16.31±0.01	**16.73±0.00**	14.52±0.00	-2.12±0.01	-3.20±0.00	15.71±0.01
UNet-L	31.04M	15.11±0.00	16.28±0.01	16.40±0.00	13.72±0.01	-1.48±0.01	**-0.30±0.01**	15.29±0.00
FBResNet	0.45M	15.36±0.00	16.19±0.01	16.66±0.00	**14.63±0.01**	-0.98±0.00	-3.18±0.00	**16.00±0.00**
UNet++	9.05M	**15.72±0.19**	**16.59±0.34**	16.55±0.20	14.45±0.06	**0.64±0.42**	-0.45±0.26	15.95±0.17
SCRN	0.43M	**17.77±0.00**	**18.63±0.00**	**18.31±0.01**	**16.78±0.00**	**2.24±0.00**	**-0.36±0.00**	**18.16±0.00**

Table 18: Mobile AVO Random Noise Poisson SNR -5 dB.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
cDDPM	33.3M	6.87±0.24	9.37±0.26	9.20±0.23	6.46±0.29	-11.92±0.37	-15.76±0.23	10.04±0.15
ResUNet	8.11M	8.06±0.00	9.54±0.01	9.84±0.01	7.17±0.00	-9.08±0.01	-10.31±0.00	8.82±0.00
UNet	7.76M	8.99±0.00	9.20±0.01	10.25±0.01	8.66±0.00	-5.20±0.00	-7.64±0.01	9.45±0.00
DnCNN	0.56M	9.15±0.00	9.69±0.01	10.62±0.00	8.83±0.00	-7.03±0.00	-9.42±0.01	9.94±0.01
Attention UNet	7.85M	9.27±0.01	9.99±0.02	10.48±0.01	8.41±0.01	-5.87±0.01	-3.27±0.02	9.42±0.01
UNet++	9.05M	9.28±0.16	9.66±0.12	10.20±0.05	8.00±0.30	**-2.52±0.40**	**-1.94±0.55**	9.36±0.18
FBResNet	0.45M	9.60±0.00	9.52±0.01	10.78±0.00	**9.33±0.01**	-4.64±0.01	-7.50±0.02	10.12±0.01
QUNet	21.88M	9.68±0.58	9.56±0.55	10.77±0.53	9.10±0.63	-3.85±0.70	-2.29±0.94	9.79±0.60
UNet-L	31.04M	**10.11±0.00**	**10.74±0.02**	**11.12±0.01**	8.88±0.01	-3.67±0.01	-2.52±0.02	**10.22±0.01**
SCRN	0.43M	**12.72±0.01**	**12.95±0.02**	**13.05±0.01**	**12.17±0.00**	**0.38±0.01**	**-1.31±0.03**	**12.87±0.01**

Table 19: Mobile AVO Random Noise Poisson SNR 0 dB.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
cDDPM	33.3M	7.05±0.62	10.22±0.61	9.93±0.56	6.63±0.60	-12.51±0.66	-16.24±0.55	11.28±0.51
ResUNet	8.11M	9.65±0.00	11.67±0.01	11.71±0.01	8.86±0.00	-8.63±0.00	-8.75±0.00	10.39±0.00
UNet	7.76M	10.27±0.00	10.28±0.01	11.99±0.00	10.42±0.01	-4.44±0.01	-5.52±0.00	10.63±0.00
Attention UNet	7.85M	11.16±0.01	12.37±0.01	12.69±0.01	10.17±0.00	-5.61±0.01	-2.86±0.00	11.34±0.01
DnCNN	0.56M	11.98±0.00	12.74±0.01	13.50±0.00	11.52±0.00	-4.66±0.00	-6.35±0.00	12.67±0.00
QUNet	21.88M	12.01±0.41	12.67±0.66	13.20±0.28	10.89±0.60	-2.97±0.69	-1.37±0.21	12.14±0.42
UNet-L	31.04M	12.21±0.00	12.65±0.00	13.55±0.00	10.88±0.01	-2.89±0.00	**-1.21±0.00**	12.33±0.00
UNet++	9.05M	12.28±0.21	**13.00±0.25**	13.24±0.20	10.81±0.45	**-1.27±0.33**	-1.28±0.22	12.41±0.22
FBResNet	0.45M	**12.35±0.00**	12.60±0.01	**13.56±0.00**	**11.79±0.00**	-2.69±0.00	-5.31±0.00	**12.91±0.00**
SCRN	0.43M	**15.11±0.00**	**15.61±0.01**	**15.50±0.00**	**14.33±0.00**	**1.16±0.02**	**-1.16±0.01**	**15.36±0.00**

Table 20: Mobile AVO Random Noise Poisson SNR +5 dB.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
cDDPM	33.3M	7.78±0.18	11.28±0.18	11.01±0.17	7.39±0.18	-12.12±0.19	-15.89±0.13	12.67±0.20
ResUNet	8.11M	10.82±0.00	13.44±0.00	13.25±0.01	10.14±0.01	-8.52±0.00	-6.51±0.00	11.37±0.00
UNet	7.76M	12.67±0.00	13.31±0.01	14.35±0.00	12.44±0.01	-3.62±0.02	-3.41±0.00	13.05±0.00
Attention UNet	7.85M	12.68±0.00	14.28±0.00	14.68±0.01	11.56±0.00	-4.93±0.01	-2.07±0.00	12.90±0.00
QUNet	21.88M	14.10±0.76	15.67±0.72	15.66±0.52	12.92±0.60	-3.23±1.24	-2.42±2.18	14.49±0.51
DnCNN	0.56M	14.96±0.00	16.13±0.01	16.61±0.00	14.37±0.00	-2.31±0.01	-3.64±0.00	15.63±0.00
UNet-L	31.04M	15.15±0.00	16.13±0.00	16.49±0.01	13.87±0.00	-1.72±0.01	-0.34±0.01	15.34±0.00
FBResNet	0.45M	15.33±0.00	16.12±0.01	16.63±0.01	14.62±0.00	-0.98±0.00	-3.13±0.01	15.97±0.00
UNet++	9.05M	**15.94±0.15**	**16.74±0.20**	**16.78±0.20**	**14.78±0.01**	**0.68±0.47**	**-0.24±0.08**	**16.16±0.16**
SCRN	0.43M	**17.77±0.00**	**18.59±0.00**	**18.31±0.01**	**16.79±0.00**	**2.25±0.00**	**-0.30±0.01**	**18.14±0.00**

H.2 SEISMIC INTERPOLATION

Table 21: Mobile AVO Continuous Missing 20 Traces.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
ANet	7.05M	9.76±0.70	11.36±1.03	9.41±0.45	8.75±0.64	1.52±1.17	7.04±0.47	9.88±0.71
CFunet	7.35M	10.92±0.06	12.63±0.06	10.12±0.04	10.55±0.32	5.74±0.45	**8.19±0.08**	11.04±0.06
CA-Unet	7.79M	11.63±0.33	13.55±0.72	10.72±0.22	11.54±0.12	7.64±0.10	**8.72±0.25**	11.76±0.36
WRDL	35.83M	11.72±0.22	14.33±0.39	10.34±0.16	11.25±0.21	6.97±0.22	7.87±0.21	11.91±0.23
DnCNN	0.14M	14.05±0.14	12.20±0.17	15.26±0.08	15.34±0.05	**10.28±0.08**	6.28±0.26	14.08±0.13
SPNet	0.67M	14.45±0.13	12.77±0.11	15.63±0.17	**15.84±0.17**	**11.04±0.36**	8.08±0.18	14.56±0.13
DnCNN-L	0.56M	15.02±0.14	13.53±0.33	15.94±0.16	**15.87±0.13**	9.98±0.17	6.73±0.14	15.06±0.14
ResUNet	8.11M	15.05±0.46	14.88±0.75	15.10±0.48	13.99±0.25	4.80±0.66	5.72±0.64	14.50±0.47
Attention UNet	7.85M	15.47±0.39	15.56±0.08	15.73±0.11	13.70±0.88	5.64±0.32	6.63±0.03	15.12±0.09
UNet	7.76M	16.08±0.09	15.73±0.18	16.11±0.10	14.11±0.27	4.96±0.87	7.00±0.48	15.08±0.43
Attention UNet-L	31.39M	16.49±0.04	16.72±0.20	16.75±0.05	14.97±0.21	7.63±0.12	7.88±0.26	**16.19±0.01**
ResUNet-L	32.44M	**16.70±0.41**	**16.97±0.46**	**17.12±0.14**	15.20±0.04	7.99±0.66	7.44±0.32	**16.51±0.20**
UNet-L	31.04M	**16.85±0.14**	**17.02±0.22**	**17.16±0.25**	15.03±0.14	7.43±0.65	7.52±0.45	16.12±0.31

Table 22: Mobile AVO Continuous Missing 30 Traces.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
ANet	7.05M	7.49±0.90	8.32±1.04	7.46±0.32	7.61±0.87	1.15±1.33	4.99±0.73	7.69±0.74
CFunet	7.35M	8.18±0.19	8.76±0.14	7.92±0.21	9.12±0.32	4.93±0.28	**6.23±0.06**	8.25±0.20
CA-Unet	7.79M	9.40±0.41	10.90±0.90	8.47±0.18	9.75±0.31	6.32±0.24	**6.29±0.37**	9.54±0.45
WRDL	35.83M	9.71±0.28	11.58±0.31	8.49±0.22	9.79±0.36	5.79±0.13	5.38±0.15	9.93±0.30
DnCNN	0.14M	11.63±0.12	9.22±0.05	13.18±0.17	13.55±0.20	**8.68±0.06**	4.63±0.23	11.66±0.12
SPNet	0.67M	11.89±0.06	9.58±0.04	13.39±0.08	**13.90±0.06**	**9.34±0.19**	5.96±0.14	11.96±0.07
DnCNN-L	0.56M	12.17±0.21	9.95±0.07	13.54±0.28	13.87±0.30	8.42±0.14	4.33±0.42	12.19±0.17
ResUNet	8.11M	13.12±1.00	12.60±1.64	13.35±0.90	12.62±0.27	3.79±0.53	4.43±0.18	12.70±1.05
Attention UNet	7.85M	13.96±0.13	13.64±0.11	14.14±0.10	12.64±0.50	4.91±0.30	4.94±0.04	13.64±0.09
UNet	7.76M	14.27±0.08	14.02±0.36	14.38±0.11	12.80±0.15	4.39±0.84	5.06±0.40	13.67±0.21
ResUNet-L	32.44M	14.33±0.43	14.28±0.48	14.86±0.48	13.56±0.43	6.43±0.72	4.80±0.74	14.32±0.44
Attention UNet-L	31.39M	**14.86±0.48**	**14.46±0.59**	**15.23±0.46**	**13.99±0.28**	6.63±0.37	5.34±0.13	**14.72±0.47**
UNet-L	31.04M	**15.13±0.12**	**15.01±0.09**	**15.48±0.15**	13.83±0.11	6.80±0.18	5.38±0.32	**14.62±0.53**

Table 23: Mobile AVO Continuous Missing 40 Traces.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
ANet	7.05M	6.40±0.57	6.70±0.71	6.48±0.34	6.94±0.67	0.90±1.36	4.25±0.30	6.48±0.58
CFunet	7.35M	6.64±0.16	6.75±0.14	6.64±0.17	8.11±0.33	4.29±0.31	**4.88±0.08**	6.70±0.16
WRDL	35.83M	7.96±0.12	9.41±0.11	6.80±0.13	8.53±0.14	4.98±0.16	3.46±0.10	8.20±0.12
CA-Unet	7.79M	8.00±0.06	9.30±0.11	7.01±0.01	8.56±0.08	5.20±0.16	4.29±0.10	8.16±0.06
DnCNN	0.14M	10.18±0.09	7.46±0.02	11.88±0.13	12.38±0.15	**7.69±0.18**	3.49±0.41	10.21±0.10
SPNet	0.67M	10.28±0.09	7.63±0.12	11.93±0.08	**12.56±0.10**	**8.15±0.13**	**4.76±0.09**	10.33±0.11
DnCNN-L	0.56M	10.47±0.19	7.89±0.13	12.06±0.25	**12.55±0.24**	7.40±0.24	3.31±0.51	10.52±0.19
ResUNet	8.11M	11.77±0.52	11.32±0.71	12.02±0.51	11.24±0.24	3.00±0.40	2.74±0.28	11.45±0.56
Attention UNet	7.85M	12.50±0.36	12.11±0.35	12.84±0.28	11.38±0.52	4.03±0.64	3.13±0.58	12.39±0.24
ResUNet-L	32.44M	12.80±0.55	12.60±0.21	13.33±0.25	12.32±0.19	5.28±0.20	3.76±0.72	12.85±0.27
UNet-L	31.04M	12.89±0.36	**12.67±0.71**	**13.40±0.67**	12.34±0.66	5.61±0.66	3.53±1.02	**12.91±0.65**
Attention UNet-L	31.39M	**12.91±0.29**	12.52±0.31	12.84±0.45	12.18±0.28	5.43±0.16	4.03±0.36	**12.88±0.39**
UNet	7.76M	**13.20±0.24**	**13.00±0.38**	**13.45±0.36**	11.79±0.07	3.77±0.51	3.64±0.41	12.75±0.43

Table 24: Mobile AVO Random Missing 30%.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
ANet	7.05M	12.79±0.64	18.01±0.52	12.27±0.59	8.55±0.81	0.57±0.49	8.69±0.30	12.99±0.67
CFunet	7.35M	14.45±0.39	20.57±0.44	13.34±0.34	10.28±0.62	4.50±0.37	10.13±0.21	14.67±0.41
DnCNN	0.14M	15.36±0.13	17.99±0.09	16.01±0.18	11.56±0.15	4.02±0.16	8.63±0.00	15.15±0.15
ResUNet	8.11M	15.59±0.86	19.05±1.14	16.92±0.94	10.76±1.03	0.71±1.34	7.26±1.56	15.08±1.09
Attention UNet	7.85M	15.95±0.31	19.03±0.45	16.90±0.49	11.45±0.24	2.15±0.32	8.38±0.67	15.54±0.37
UNet	7.76M	16.00±0.99	19.21±1.15	17.04±1.18	11.34±0.87	2.47±0.72	8.43±0.46	15.59±1.05
SPNet	0.67M	17.17±0.10	20.68±0.08	17.76±0.06	13.76±0.16	**5.88±0.13**	10.41±0.09	17.23±0.10
ResUNet-L	32.44M	17.83±0.62	22.23±0.38	19.10±0.54	13.24±0.70	3.87±0.42	9.60±0.63	17.63±0.54
CA-Unet	7.79M	17.85±0.20	**24.22±0.77**	16.90±0.23	13.87±0.07	5.58±0.11	**11.37±0.05**	18.28±0.21
DnCNN-L	0.56M	17.92±0.28	21.13±0.13	18.31±0.24	14.66±0.40	**7.05±0.37**	10.23±0.09	17.90±0.28
WRDL	35.83M	17.94±0.16	**26.52±0.30**	17.03±0.07	13.42±0.19	5.11±0.23	**11.15±0.05**	18.43±0.19
Attention UNet-L	31.39M	**18.98±0.16**	22.85±0.42	**19.91±0.12**	**14.74±0.26**	5.14±0.28	10.41±0.53	**18.72±0.18**
UNet-L	31.04M	**19.17±0.05**	23.15±0.17	**20.29±0.06**	**14.79±0.08**	4.77±0.15	10.19±0.38	**18.96±0.09**

Table 25: Mobile AVO Random Missing 50%.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
ANet	7.05M	10.44±0.18	15.44±0.16	9.62±0.12	6.38±0.24	-0.36±0.19	6.12±0.13	10.66±0.18
ResUNet	8.11M	11.64±0.54	15.73±0.22	13.36±0.07	7.05±0.14	-1.49±0.36	4.77±0.68	11.62±0.11
CFunet	7.35M	11.81±0.19	17.66±0.16	10.58±0.13	7.66±0.46	2.91±0.24	7.32±0.09	12.04±0.21
DnCNN	0.14M	11.87±0.16	14.49±0.22	12.48±0.14	8.36±0.16	2.12±0.04	5.84±0.46	11.79±0.14
SPNet	0.67M	12.90±0.14	15.75±0.13	13.42±0.17	9.77±0.18	3.05±0.17	6.67±0.08	12.95±0.14
Attention UNet	7.85M	13.05±0.81	16.41±0.97	14.13±0.78	8.73±0.85	0.22±0.34	5.81±0.34	12.90±0.75
UNet	7.76M	13.75±0.31	17.30±0.54	14.80±0.29	9.26±0.35	0.62±0.48	6.37±0.06	13.51±0.36
WRDL	35.83M	13.79±0.10	**21.16±0.19**	12.45±0.03	9.69±0.18	2.55±0.29	**7.39±0.24**	14.22±0.15
DnCNN-L	0.56M	13.93±0.30	16.72±0.44	14.22±0.25	**11.04±0.25**	**3.92±0.17**	6.70±0.34	13.93±0.30
CA-Unet	7.79M	14.11±0.10	**19.96±0.05**	12.79±0.06	10.47±0.20	**3.41±0.17**	**7.93±0.05**	14.51±0.12
ResUNet-L	32.44M	14.18±0.45	18.49±0.44	15.22±0.59	10.02±0.45	1.19±0.75	6.09±0.54	14.12±0.59
UNet-L	31.04M	**14.90±0.62**	18.91±0.87	**15.59±0.93**	11.01±0.38	1.91±0.66	6.52±0.16	**14.76±0.74**
Attention UNet-L	31.39M	**15.19±0.13**	19.03±0.32	**15.99±0.21**	**11.27±0.29**	2.52±0.32	6.82±0.35	**15.09±0.26**

Table 26: Mobile AVO Random Missing 70%.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
DnCNN	0.14M	8.22±0.22	9.93±0.34	8.68±0.27	5.30±0.30	0.45±0.16	2.63±0.64	8.14±0.29
SPNet	0.67M	8.42±0.10	10.21±0.05	8.86±0.09	5.86±0.14	**0.85±0.13**	**3.58±0.05**	8.45±0.10
ResUNet	8.11M	9.08±0.65	12.74±0.76	10.14±0.60	4.80±0.73	-2.32±0.77	2.81±0.73	8.97±0.70
Attention UNet	7.85M	9.60±0.24	12.99±0.37	10.63±0.16	5.84±0.16	-1.37±0.03	2.68±0.59	9.70±0.17
DnCNN-L	0.56M	10.09±0.21	12.12±0.32	10.34±0.17	**7.61±0.18**	**1.47±0.10**	3.22±0.38	10.07±0.22
UNet	7.76M	10.31±0.19	13.66±0.37	11.12±0.16	6.40±0.18	-0.95±0.18	**3.49±0.12**	10.26±0.19
ResUNet-L	32.44M	10.58±0.37	**14.48±0.13**	11.35±0.28	6.93±0.33	-0.50±0.22	3.24±0.67	10.60±0.26
Attention UNet-L	31.39M	**10.78±0.10**	13.73±0.38	**11.41±0.04**	7.60±0.23	0.04±0.17	2.59±1.26	**10.69±0.06**
UNet-L	31.04M	**11.09±0.45**	**14.34±0.62**	**11.71±0.46**	**7.61±0.35**	-0.06±0.16	3.40±0.54	**11.08±0.40**

Table 27: Mobile AVO Uniform Missing 50%.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
CFunet	7.35M	13.37±0.28	20.35±0.38	12.27±0.24	8.75±0.60	3.24±0.26	9.14±0.18	13.58±0.30
ANet	7.05M	14.17±0.58	20.99±0.98	13.39±0.49	9.70±0.57	1.49±0.56	8.94±0.26	14.46±0.61
DnCNN	0.14M	16.35±0.27	20.29±0.62	17.42±0.27	12.05±0.27	2.72±0.25	8.37±0.34	16.34±0.30
ResUNet	8.11M	16.36±0.18	21.67±0.37	18.70±0.39	11.58±0.50	0.58±0.03	7.08±0.78	16.28±0.43
SPNet	0.67M	17.43±0.31	21.45±0.18	18.36±0.25	13.40±0.46	3.98±0.37	9.21±0.08	17.50±0.33
WRDL	35.83M	18.57±0.04	**29.98±0.26**	17.59±0.22	13.69±0.05	5.31±0.26	**9.88±0.29**	19.41±0.04
UNet	7.76M	18.75±0.22	23.46±0.84	20.49±0.24	14.03±0.25	2.33±0.28	8.83±0.20	18.54±0.24
Attention UNet	7.85M	18.86±0.54	23.89±0.81	20.52±0.54	13.86±0.95	2.97±0.21	8.71±0.61	18.78±0.45
DnCNN-L	0.56M	19.05±0.43	23.90±0.63	19.84±0.55	14.84±0.27	**5.86±0.14**	9.29±0.25	19.05±0.45
CA-Unet	7.79M	19.06±0.16	**28.78±0.50**	18.07±0.30	14.36±0.05	**6.29±0.18**	**10.44±0.05**	19.87±0.20
ResUNet-L	32.44M	19.20±0.83	25.99±1.46	20.88±0.65	14.47±0.78	2.96±0.82	9.26±0.45	19.14±0.85
Attention UNet-L	31.39M	**20.04±0.19**	26.29±0.51	**21.44±0.17**	**15.41±0.11**	5.06±0.58	9.72±0.44	**19.96±0.16**
UNet-L	31.04M	**20.26±0.27**	27.41±0.58	**21.69±0.22**	**15.59±0.22**	4.86±0.29	9.54±0.28	**20.22±0.39**

Table 28: Mobile AVO Uniform Missing 75%.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
ANet	7.05M	11.04±1.23	17.93±1.35	10.41±0.89	6.14±1.61	-0.72±1.22	6.05±0.25	11.32±1.32
CFunet	7.35M	11.60±0.26	17.74±0.28	10.46±0.13	7.21±0.46	2.13±0.23	6.56±0.07	11.88±0.28
DnCNN	0.14M	11.91±0.21	15.65±0.41	12.69±0.21	7.84±0.22	0.47±0.12	4.82±0.17	11.89±0.22
SPNet	0.67M	12.37±0.01	15.78±0.23	13.08±0.01	8.64±0.07	1.00±0.08	4.99±0.11	12.43±0.01
ResUNet	8.11M	12.58±0.12	18.24±0.21	14.24±0.12	7.95±0.18	-1.35±0.25	4.72±0.32	12.64±0.05
DnCNN-L	0.56M	13.75±0.24	18.17±0.46	14.31±0.24	9.91±0.20	**2.16±0.19**	5.10±0.23	13.78±0.26
UNet	7.76M	13.92±0.20	19.50±0.34	15.02±0.12	9.52±0.17	-0.12±0.46	5.24±0.10	13.92±0.20
Attention UNet	7.85M	13.93±0.21	19.53±0.35	14.94±0.14	9.62±0.13	0.13±0.47	5.12±0.23	13.93±0.18
ResUNet-L	32.44M	13.93±0.32	20.08±0.15	14.97±0.28	9.50±0.48	0.65±0.24	5.36±0.14	13.97±0.34
UNet-L	31.04M	14.19±0.40	20.68±0.53	**15.20±0.12**	10.32±0.13	1.32±0.18	5.31±0.34	14.20±0.43
Attention UNet-L	31.39M	14.42±0.43	20.71±0.35	**15.31±0.20**	**10.46±0.20**	1.72±0.57	5.00±0.62	14.54±0.32
WRDL	35.83M	**14.52±0.05**	**23.64±0.21**	13.17±0.09	10.26±0.09	1.74±0.17	**6.94±0.02**	**15.12±0.06**
CA-Unet	7.79M	**15.43±0.19**	**23.87±0.48**	13.87±0.23	**11.42±0.14**	**3.52±0.16**	**7.28±0.08**	**16.15±0.22**

Table 29: SEGC3 Random Missing 30%.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
ANet	7.05M	19.06±0.09	21.28±0.41	24.27±0.13	15.74±0.04	12.16±0.18	15.12±0.15	21.29±0.07
CFunet	7.35M	20.08±0.10	24.44±0.23	26.44±0.12	16.31±0.11	13.65±0.06	15.38±0.11	23.38±0.11
ResUNet	8.11M	20.39±0.36	26.15±0.16	26.99±0.48	16.60±0.35	13.52±0.35	15.67±0.31	23.68±0.46
Attention UNet	7.85M	20.44±0.13	23.71±0.09	26.71±0.13	16.61±0.13	13.25±0.23	15.67±0.15	23.25±0.13
UNet	7.76M	20.45±0.13	24.11±0.37	26.89±0.15	16.61±0.15	13.23±0.17	15.67±0.15	23.40±0.09
SPNet	0.67M	20.87±0.18	28.49±0.51	28.46±0.36	16.89±0.15	13.93±0.23	16.35±0.14	24.37±0.25
UNet-L	31.04M	21.28±0.07	28.28±0.32	29.24±0.11	17.18±0.08	14.37±0.10	16.32±0.07	25.00±0.09
CA-Unet	7.79M	21.30±0.06	26.45±0.34	28.54±0.20	17.49±0.05	14.52±0.07	16.62±0.06	24.70±0.06
ResUNet-L	32.44M	21.36±0.15	**30.50±0.07**	29.57±0.24	17.25±0.14	14.50±0.15	16.38±0.15	25.33±0.17
Attention UNet-L	31.39M	21.50±0.26	28.05±0.16	29.30±0.21	17.40±0.26	14.53±0.24	16.53±0.28	25.18±0.23
DnCNN	0.14M	21.50±0.05	26.14±0.41	29.37±0.13	17.46±0.03	14.58±0.04	16.80±0.03	24.87±0.06
DnCNN-L	0.56M	**21.97±0.06**	27.46±0.13	**29.97±0.54**	**17.90±0.01**	**15.01±0.07**	**17.18±0.03**	**25.56±0.13**
WRDL	35.83M	**22.28±0.22**	**29.78±0.74**	**31.19±0.69**	**18.38±0.23**	**14.99±0.23**	**17.81±0.23**	**25.71±0.33**

Table 30: SEGC3 Random Missing 50%.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
ANet	7.05M	14.68±0.67	18.14±1.10	18.72±0.79	11.65±0.59	8.25±0.85	10.43±0.54	17.14±0.88
SPNet	0.67M	15.62±0.07	21.39±0.19	20.11±0.10	12.31±0.05	9.36±0.06	11.25±0.07	18.54±0.08
ResUNet	8.11M	15.66±0.04	22.23±0.16	19.87±0.07	12.45±0.06	9.70±0.07	11.02±0.04	18.66±0.04
UNet	7.76M	15.68±0.07	20.67±0.24	19.92±0.09	12.46±0.06	9.45±0.06	11.04±0.08	18.54±0.05
Attention UNet	7.85M	15.71±0.17	20.48±0.26	19.90±0.29	12.50±0.14	9.55±0.14	11.09±0.15	18.50±0.27
DnCNN	0.14M	16.11±0.02	20.39±0.09	20.61±0.05	12.74±0.02	10.15±0.02	11.58±0.00	18.88±0.06
ResUNet-L	32.44M	16.23±0.06	**24.92±0.41**	20.83±0.18	12.79±0.03	10.50±0.02	11.42±0.02	19.61±0.16
CFunet	7.35M	16.23±0.09	22.69±0.11	20.72±0.09	12.88±0.10	10.56±0.13	11.48±0.08	19.42±0.13
UNet-L	31.04M	16.26±0.03	23.82±0.29	20.95±0.03	12.81±0.05	10.38±0.05	11.42±0.05	19.63±0.03
Attention UNet-L	31.39M	16.31±0.13	23.71±0.09	20.94±0.18	12.88±0.12	10.48±0.10	11.48±0.13	19.64±0.13
WRDL	35.83M	16.54±0.46	**24.84±1.42**	**21.51±0.43**	**13.20±0.30**	10.49±0.43	**12.22±0.08**	**20.10±0.08**
CA-Unet	7.79M	**16.55±0.11**	22.84±0.40	20.96±0.14	**13.27±0.10**	**10.73±0.15**	11.84±0.14	**19.76±0.03**
DnCNN-L	0.56M	**16.56±0.01**	22.39±0.16	**21.25±0.10**	13.13±0.04	**10.59±0.07**	**11.88±0.04**	19.62±0.04

Table 31: SEGC3 Random Missing 70%.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
SPNet	0.67M	10.27±0.05	14.23±0.22	13.05±0.03	7.72±0.05	4.72±0.10	6.32±0.05	12.34±0.05
DnCNN	0.14M	10.60±0.02	14.11±0.10	13.27±0.04	8.07±0.02	5.33±0.03	6.56±0.01	12.56±0.04
UNet	7.76M	10.93±0.02	16.00±0.25	13.38±0.06	8.55±0.02	5.68±0.04	6.68±0.04	13.18±0.05
Attention UNet	7.85M	10.96±0.10	15.84±0.26	13.42±0.15	8.58±0.08	5.67±0.09	6.69±0.07	13.23±0.14
ResUNet	8.11M	11.00±0.02	16.86±0.29	13.51±0.05	8.57±0.03	5.83±0.05	6.71±0.03	13.31±0.08
DnCNN-L	0.56M	11.11±0.04	15.76±0.04	**13.79±0.04**	8.56±0.04	5.81±0.07	6.83±0.03	13.35±0.03
UNet-L	31.04M	11.19±0.08	17.67±0.12	13.64±0.09	8.75±0.07	6.24±0.13	6.76±0.08	13.73±0.10
Attention UNet-L	31.39M	**11.32±0.06**	**17.82±0.11**	13.78±0.07	**8.86±0.07**	**6.40±0.07**	**6.86±0.05**	**13.87±0.09**
ResUNet-L	32.44M	**11.37±0.03**	**18.37±0.09**	**13.90±0.04**	**8.87±0.03**	**6.40±0.02**	**6.89±0.03**	**14.02±0.04**

Table 32: SEGC3 Uniform Missing 50%.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
CFunet	7.35M	16.52±0.23	24.29±0.45	27.95±0.15	12.30±0.28	9.58±0.16	11.44±0.26	20.17±0.25
DnCNN	0.14M	16.66±0.06	22.77±0.11	28.90±0.22	12.47±0.06	9.10±0.02	11.49±0.05	20.36±0.10
ANet	7.05M	16.78±0.20	22.12±0.92	26.19±0.49	12.91±0.22	9.38±0.43	11.92±0.26	20.14±0.12
DnCNN-L	0.56M	16.89±0.06	24.90±0.55	29.69±0.93	12.74±0.05	8.86±0.04	11.69±0.04	20.88±0.13
SPNet	0.67M	17.12±0.04	27.93±0.19	31.35±0.17	12.79±0.06	9.84±0.03	12.14±0.04	21.37±0.04
Attention UNet	7.85M	18.08±0.03	27.77±0.51	32.09±0.15	13.85±0.03	10.32±0.03	12.84±0.03	22.30±0.06
UNet	7.76M	18.08±0.06	27.77±0.49	32.20±0.26	13.84±0.06	10.37±0.08	12.85±0.05	22.31±0.07
ResUNet	8.11M	18.14±0.03	30.43±0.09	32.98±0.14	13.87±0.03	10.48±0.02	12.90±0.03	22.54±0.02
ResUNet-L	32.44M	18.18±0.03	**33.00±0.16**	**33.84±0.07**	13.84±0.03	10.67±0.06	12.91±0.04	**22.71±0.01**
Attention UNet-L	31.39M	18.20±0.06	31.10±0.20	33.54±0.08	13.88±0.06	10.69±0.07	12.92±0.05	22.65±0.04
UNet-L	31.04M	18.22±0.03	31.20±0.34	33.67±0.04	13.90±0.03	10.65±0.03	12.95±0.03	22.70±0.06
CA-Unet	7.79M	**18.26±0.02**	28.40±0.66	32.11±0.67	**14.07±0.05**	**10.79±0.08**	**13.08±0.06**	22.61±0.04
WRDL	35.83M	**18.56±0.01**	**31.75±0.35**	**33.83±0.12**	**14.32±0.04**	**11.40±0.04**	**13.35±0.03**	**23.02±0.04**

Table 33: SEGC3 Uniform Missing 75%.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
DnCNN	0.14M	12.29±0.05	19.03±0.12	17.41±0.06	8.96±0.05	5.55±0.06	7.39±0.05	15.37±0.09
SPNet	0.67M	12.47±0.08	21.55±0.10	17.66±0.08	9.15±0.07	5.71±0.15	7.58±0.05	15.66±0.17
DnCNN-L	0.56M	12.84±0.07	21.38±0.11	17.87±0.06	9.52±0.07	6.04±0.11	7.81±0.05	16.27±0.12
ResUNet-L	32.44M	13.26±0.05	**26.48±0.05**	17.94±0.05	9.95±0.07	7.20±0.08	8.12±0.06	17.14±0.04
UNet-L	31.04M	13.26±0.05	25.44±0.08	17.80±0.02	10.00±0.06	7.27±0.07	8.12±0.05	17.09±0.05
UNet	7.76M	13.30±0.02	22.96±0.31	17.94±0.04	10.08±0.02	6.97±0.04	8.21±0.02	16.96±0.02
Attention UNet-L	31.39M	13.31±0.06	25.36±0.27	17.82±0.07	10.06±0.07	7.35±0.04	8.17±0.08	17.15±0.04
ResUNet	8.11M	13.35±0.04	23.97±0.15	18.02±0.04	10.12±0.03	7.16±0.08	8.27±0.02	17.01±0.03
Attention UNet	7.85M	13.35±0.08	23.10±0.16	17.97±0.12	10.14±0.07	7.08±0.06	8.26±0.07	17.03±0.09
CFunet	7.35M	13.35±0.71	23.01±1.02	21.72±0.23	9.94±0.04	6.50±0.10	8.11±0.70	17.07±0.53
ANet	7.05M	14.73±0.43	21.61±1.74	21.58±0.59	11.23±0.37	7.30±0.56	9.67±0.33	18.44±0.62
CA-Unet	7.79M	**15.28±0.06**	26.00±0.67	**22.63±0.20**	**11.62±0.10**	**7.95±0.10**	**10.06±0.09**	**19.43±0.06**
WRDL	35.83M	**15.31±0.09**	**28.01±0.73**	**22.42±0.08**	**11.74±0.16**	**8.23±0.18**	**10.17±0.17**	**19.53±0.16**

Table 34: SEGC3 Continuous Missing 20 Traces.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
ANet	7.05M	14.61±0.42	16.39±0.75	16.34±0.28	13.02±0.45	10.49±0.63	11.55±0.40	16.06±0.50
DnCNN	0.14M	14.88±0.10	16.49±0.16	16.28±0.14	13.16±0.07	11.72±0.07	11.83±0.02	16.06±0.15
CFunet	7.35M	15.12±0.10	18.22±0.07	16.49±0.08	13.53±0.17	11.99±0.06	11.92±0.03	16.51±0.18
SPNet	0.67M	15.45±0.05	18.27±0.09	17.08±0.04	13.59±0.08	11.96±0.00	12.02±0.05	17.06±0.07
Attention UNet	7.85M	15.50±0.09	17.62±0.28	16.81±0.03	13.76±0.17	12.04±0.13	12.11±0.03	16.78±0.15
ResUNet	8.11M	15.56±0.05	19.01±0.04	16.93±0.06	13.87±0.04	12.26±0.08	12.05±0.08	17.13±0.05
UNet	7.76M	15.88±0.08	19.00±0.15	17.13±0.11	14.26±0.04	12.50±0.05	12.29±0.02	17.42±0.05
WRDL	35.83M	16.04±0.32	20.21±1.32	**17.73±0.03**	14.66±0.08	12.83±0.10	**12.60±0.28**	17.88±0.39
DnCNN-L	0.56M	16.14±0.02	19.43±0.02	**17.58±0.05**	14.44±0.02	12.77±0.03	12.49±0.02	17.84±0.02
ResUNet-L	32.44M	16.18±0.10	**20.40±0.32**	17.46±0.13	14.61±0.07	13.00±0.05	12.47±0.02	**17.95±0.23**
Attention UNet-L	31.39M	16.21±0.02	19.71±0.23	17.38±0.01	**14.70±0.06**	13.01±0.13	12.50±0.08	17.91±0.06
UNet-L	31.04M	**16.25±0.07**	19.94±0.07	17.47±0.08	**14.70±0.05**	**13.03±0.09**	12.55±0.02	**17.95±0.11**
CA-Unet	7.79M	**16.39±0.03**	**20.42±0.35**	17.57±0.06	**15.00±0.02**	**13.45±0.08**	**12.65±0.08**	**18.33±0.11**

Table 35: SEGC3 Continuous Missing 30 Traces.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
ANet	7.05M	11.08±2.05	12.55±2.96	12.23±2.25	9.87±1.77	7.69±2.35	8.68±1.29	11.94±2.39
CFunet	7.35M	11.41±0.13	14.09±0.07	12.26±0.12	10.22±0.16	9.61±0.05	9.26±0.05	12.15±0.17
DnCNN	0.14M	11.42±0.06	12.70±0.18	12.23±0.08	10.27±0.04	9.53±0.02	9.32±0.02	12.10±0.09
SPNet	0.67M	12.06±0.05	14.16±0.07	13.31±0.06	10.48±0.05	9.52±0.07	9.35±0.03	13.14±0.07
DnCNN-L	0.56M	12.29±0.04	14.44±0.08	13.28±0.09	10.97±0.01	10.05±0.09	9.64±0.03	13.25±0.08
Attention UNet	7.85M	12.50±0.03	14.77±0.22	13.73±0.05	10.86±0.05	9.65±0.11	9.46±0.02	13.58±0.06
ResUNet	8.11M	12.64±0.03	15.44±0.07	13.79±0.04	11.15±0.02	9.82±0.07	9.58±0.05	13.82±0.09
UNet	7.76M	12.75±0.08	15.62±0.06	13.87±0.09	11.27±0.06	9.89±0.07	9.64±0.06	13.96±0.09
WRDL	35.83M	13.00±0.06	**16.89±0.24**	**14.20±0.02**	11.61±0.14	9.89±0.17	9.87±0.14	14.49±0.01
Attention UNet-L	31.39M	13.05±0.12	16.24±0.10	14.08±0.13	11.72±0.09	10.21±0.18	9.83±0.14	14.39±0.10
UNet-L	31.04M	13.07±0.08	16.33±0.10	14.08±0.07	11.77±0.11	10.18±0.13	9.83±0.11	14.42±0.09
ResUNet-L	32.44M	**13.10±0.03**	**16.52±0.04**	**14.12±0.04**	**11.81±0.03**	**10.32±0.02**	**9.89±0.04**	**14.51±0.08**
CA-Unet	7.79M	**13.21±0.04**	16.26±0.20	**14.20±0.06**	**12.03±0.07**	**10.51±0.06**	**9.99±0.04**	**14.66±0.05**

Table 36: SEGC3 Continuous Missing 40 Traces.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
CFunet	7.35M	9.01±0.05	10.82±0.08	9.46±0.01	8.32±0.09	8.12±0.07	7.61±0.03	9.46±0.06
DnCNN	0.14M	9.28±0.04	10.16±0.09	9.80±0.05	8.49±0.03	8.11±0.10	7.76±0.02	9.72±0.05
ANet	7.05M	9.51±1.17	11.01±1.65	10.59±1.23	8.21±1.08	6.46±1.41	7.20±0.77	10.30±1.35
DnCNN-L	0.56M	9.79±0.04	11.06±0.04	10.41±0.06	8.93±0.01	8.40±0.02	7.95±0.03	10.38±0.06
SPNet	0.67M	9.88±0.05	11.53±0.04	10.76±0.06	8.68±0.04	7.99±0.06	7.75±0.01	10.63±0.06
Attention UNet	7.85M	10.43±0.03	12.37±0.15	11.53±0.04	8.95±0.04	8.07±0.08	7.76±0.00	11.31±0.06
ResUNet	8.11M	10.56±0.09	12.99±0.11	11.57±0.10	9.23±0.08	8.22±0.09	7.85±0.04	11.57±0.10
UNet	7.76M	10.65±0.05	12.95±0.03	11.67±0.05	9.28±0.07	8.29±0.06	7.90±0.06	11.64±0.05
UNet-L	31.04M	10.90±0.04	13.49±0.12	11.82±0.06	9.69±0.00	**8.48±0.11**	8.13±0.01	11.96±0.06
ResUNet-L	32.44M	10.93±0.01	**13.67±0.04**	11.85±0.02	9.74±0.01	8.42±0.02	8.10±0.02	12.05±0.03
Attention UNet-L	31.39M	10.94±0.04	13.49±0.10	11.86±0.07	9.71±0.06	8.42±0.12	8.11±0.06	12.02±0.06
WRDL	35.83M	**10.94±0.13**	**13.93±0.47**	**11.99±0.15**	**9.75±0.09**	8.24±0.10	**8.15±0.05**	**12.25±0.17**
CA-Unet	7.79M	**11.06±0.02**	13.40±0.11	**11.93±0.05**	**10.01±0.05**	**8.61±0.08**	**8.27±0.08**	**12.23±0.04**

H.3 GROUND-ROLL SUPPRESSION

Table 37: SEGC3 Ground-Roll Noise 1.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
cDDPM	50.61M	13.50±1.39	6.09±1.72	17.56±0.88	3.15±1.45	1.44±1.73	2.03±1.05	18.00±0.75
Enhanced Atten-UNet	8.2M	14.64±0.40	6.48±0.80	16.67±0.17	8.64±1.38	5.70±1.76	7.58±0.64	15.02±0.38
Pix2Pix cGAN	48.84M	15.94±0.42	6.77±1.39	18.47±0.35	8.42±0.31	9.33±0.93	3.73±0.63	17.87±0.38
Physics CNN	7.27M	18.91±0.33	10.19±0.50	20.56±0.24	14.83±0.68	13.62±1.00	10.20±0.43	19.63±0.27
DnCNN	0.14M	19.73±0.33	10.56±0.34	22.50±0.37	11.78±0.30	11.00±0.23	12.62±0.61	20.15±0.29
SANet	0.23M	21.49±0.30	10.92±0.38	25.37±0.28	13.66±0.31	12.17±0.22	12.06±0.35	22.32±0.29
ResUNet-L	32.44M	25.80±0.13	16.00±0.26	27.90±0.07	20.71±0.48	20.75±0.17	15.18±0.29	26.89±0.19
UNet-L	31.04M	26.56±0.45	16.30±0.57	28.81±0.61	22.18±0.42	21.74±0.05	14.97±0.54	28.02±0.45
Attention UNet-L	31.39M	28.01±1.33	17.48±0.97	30.87±1.85	22.39±1.18	21.96±0.34	17.80±2.42	29.03±1.00
UNet	7.76M	28.39±1.09	17.08±1.98	32.00±0.97	**22.60±0.51**	22.20±0.48	**18.77±0.86**	29.23±1.16
Attention UNet	7.85M	**29.48±1.35**	**18.80±1.01**	**32.64±1.86**	21.77±1.45	**23.70±0.67**	17.72±2.10	**31.04±1.04**
ResUNet	8.11M	**31.62±0.57**	**19.78±0.14**	**35.79±1.24**	**25.52±0.90**	**25.68±0.89**	**20.63±0.65**	**32.87±0.61**

Table 38: SEGC3 Ground-Roll Noise 3.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
cDDPM	50.61M	7.73±1.42	0.39±1.87	12.16±1.05	-3.04±1.35	-4.70±1.67	-3.75±1.37	12.59±0.79
Pix2Pix cGAN	48.84M	8.93±0.31	0.73±0.48	11.44±0.32	0.19±0.30	0.55±0.72	-3.18±0.40	10.77±0.28
Enhanced Atten-UNet	8.2M	11.75±0.93	5.56±0.19	13.66±0.81	5.40±2.69	1.44±1.23	6.09±2.07	12.02±0.84
Physics CNN	7.27M	14.44±0.14	5.67±0.32	16.15±0.15	10.77±0.15	9.62±0.73	5.83±0.22	15.23±0.14
SANet	0.23M	17.11±0.25	7.93±0.44	20.25±0.20	8.96±0.16	7.20±0.24	6.61±0.28	18.26±0.25
DnCNN	0.14M	17.24±0.23	9.80±0.23	18.81±0.24	11.38±0.19	9.91±0.25	10.57±0.15	17.62±0.22
ResUNet-L	32.44M	21.63±0.14	13.45±0.04	22.94±0.18	18.23±0.30	17.24±0.27	12.00±0.32	22.46±0.11
ResUNet	8.11M	22.65±0.93	9.82±2.82	**28.36±1.08**	**21.96±0.95**	**21.03±0.23**	10.41±1.91	**24.63±0.68**
Attention UNet-L	31.39M	22.89±0.49	**13.66±0.49**	24.84±0.55	18.36±1.11	17.17±1.19	13.30±1.09	23.75±0.37
UNet	7.76M	23.07±0.10	**13.97±0.37**	24.92±0.23	18.86±0.75	17.58±1.01	**13.93±0.34**	23.81±0.11
UNet-L	31.04M	**23.07±0.18**	13.59±0.12	25.08±0.31	18.94±0.09	17.47±0.18	12.93±0.22	24.06±0.22
Attention UNet	7.85M	**24.39±1.22**	13.25±2.13	**27.33±1.13**	**20.73±0.58**	**19.76±0.80**	**13.86±1.97**	**25.53±1.09**

Table 39: SEGC3 Ground-Roll Noise 5.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
cDDPM	50.61M	4.47±1.44	-2.96±1.86	9.31±1.19	-6.53±1.28	-8.22±1.59	-7.03±1.28	9.65±0.83
Pix2Pix cGAN	48.84M	5.70±0.61	-2.74±0.84	8.68±0.68	-3.36±0.44	-3.55±0.71	-6.52±0.69	7.63±0.58
Enhanced Atten-UNet	8.2M	10.16±0.83	4.65±0.11	11.91±0.74	3.58±2.86	0.29±0.91	4.81±2.37	10.41±0.73
Physics CNN	7.27M	12.95±0.19	2.77±0.43	15.07±0.27	9.41±0.28	8.47±0.32	2.97±0.31	14.02±0.24
SANet	0.23M	14.26±0.14	5.25±0.34	17.62±0.13	5.49±0.07	3.73±0.08	3.23±0.14	15.61±0.15
DnCNN	0.14M	15.57±0.22	9.13±0.39	16.65±0.21	10.90±0.38	9.40±0.50	9.83±0.33	15.84±0.23
ResUNet	8.11M	18.11±2.12	4.96±3.67	**23.94±1.49**	**20.25±1.45**	**19.23±0.52**	5.59±2.49	20.16±1.95
ResUNet-L	32.44M	19.52±0.29	**11.98±0.15**	20.64±0.38	16.39±0.00	14.99±0.13	10.57±0.27	20.23±0.31
UNet	7.76M	20.22±0.32	11.46±0.17	21.76±0.27	16.78±1.04	15.37±1.25	**11.25±0.30**	20.92±0.31
Attention UNet-L	31.39M	20.23±0.33	11.55±0.12	21.92±0.35	15.73±1.29	14.21±1.39	11.09±0.27	20.98±0.36
UNet-L	31.04M	**20.39±0.13**	11.44±0.31	22.12±0.24	16.33±0.21	14.49±0.22	10.81±0.30	**21.25±0.17**
Attention UNet	7.85M	**21.96±1.06**	**12.06±2.17**	**24.22±0.92**	**18.43±0.45**	**17.55±0.36**	**11.97±1.79**	**22.96±0.92**

Table 40: SEGC3 Ground-Roll Noise 7.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
cDDPM	50.61M	2.15±1.48	-5.35±1.83	7.18±1.30	-8.94±1.32	-10.67±1.59	-9.40±1.12	7.56±0.91
Pix2Pix cGAN	48.84M	3.44±0.32	-4.91±0.96	6.43±0.17	-5.79±0.38	-5.78±0.49	-8.85±0.45	5.38±0.26
Enhanced Atten-UNet	8.2M	9.24±1.19	3.82±0.15	10.87±1.20	2.89±3.70	0.06±0.86	3.66±3.17	9.52±1.04
Physics CNN	7.27M	11.28±0.55	0.19±0.59	13.88±0.47	7.66±0.83	7.69±1.41	0.60±0.72	12.66±0.50
SANet	0.23M	12.10±0.19	3.08±0.48	15.78±0.17	2.88±0.16	1.10±0.25	0.77±0.27	13.56±0.16
DnCNN	0.14M	15.17±0.13	8.66±0.28	16.34±0.10	10.34±0.34	8.29±0.40	9.28±0.36	15.46±0.10
ResUNet	8.11M	16.60±1.44	7.08±3.82	19.72±1.38	**17.21±1.52**	**16.28±1.14**	5.72±2.07	17.91±1.30
UNet	7.76M	17.90±0.28	9.62±0.11	19.31±0.31	14.37±1.32	12.70±1.31	**9.33±0.34**	18.54±0.29
Attention UNet	7.85M	17.94±2.80	7.41±5.50	**21.13±2.00**	**15.35±2.07**	**14.41±2.80**	7.47±4.27	**19.18±2.36**
ResUNet-L	32.44M	18.29±0.40	**10.92±0.21**	19.43±0.53	14.65±0.06	13.21±0.13	**9.57±0.30**	18.95±0.43
Attention UNet-L	31.39M	**18.31±0.36**	**10.07±0.09**	19.92±0.24	13.42±1.34	11.78±1.40	9.27±0.26	19.05±0.38
UNet-L	31.04M	**18.70±0.09**	9.76±0.39	**20.45±0.21**	14.32±0.30	12.28±0.47	9.03±0.25	**19.58±0.15**

Table 41: SEGC3 Ground-Roll Noise 9.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
Pix2Pix cGAN	48.84M	0.04±2.66	-9.46±4.49	4.06±1.74	-9.06±2.16	-9.24±1.78	-12.92±3.68	2.47±2.15
cDDPM	50.61M	0.39±1.22	-7.19±1.61	5.50±1.14	-10.79±1.02	-12.53±1.28	-11.29±1.38	5.89±0.76
Enhanced Atten-UNet	8.2M	8.49±1.54	3.47±0.22	9.87±1.80	2.69±3.61	0.01±0.91	3.95±3.21	8.69±1.46
Physics CNN	7.27M	10.34±0.29	-0.71±0.41	13.40±0.32	5.57±0.65	5.38±1.08	-0.64±0.35	12.01±0.29
SANet	0.23M	10.58±0.14	1.53±0.51	14.45±0.16	1.14±0.11	-0.56±0.40	-0.92±0.14	12.13±0.15
ResUNet	8.11M	13.89±1.24	3.28±4.08	18.10±2.11	**15.40±0.84**	**14.20±1.43**	2.84±2.97	15.52±0.58
DnCNN	0.14M	14.87±0.18	8.28±0.39	16.12±0.12	9.80±0.59	7.56±0.23	**8.79±0.37**	15.18±0.16
Attention UNet	7.85M	15.78±2.76	4.80±5.52	**19.32±1.40**	**13.39±2.25**	**12.18±3.17**	5.01±4.50	17.13±2.13
UNet	7.76M	16.60±0.46	7.80±0.52	18.13±0.41	13.19±1.29	11.35±1.41	7.60±0.58	17.33±0.45
Attention UNet-L	31.39M	17.06±0.27	**8.94±0.37**	18.72±0.18	11.81±1.27	9.98±1.36	7.98±0.17	17.81±0.29
UNet-L	31.04M	**17.32±0.08**	8.30±0.17	**19.18±0.11**	12.53±0.25	10.39±0.42	7.57±0.17	**18.23±0.12**
ResUNet-L	32.44M	**17.38±0.28**	**10.13±0.17**	18.58±0.35	13.22±0.15	11.60±0.15	**8.59±0.32**	**18.07±0.28**

Table 42: Field Ground-Roll Noise 1.0.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
SANet	0.23M	14.75±0.14	14.53±0.11	17.51±0.20	11.30±0.24	6.16±0.15	4.20±0.07	17.01±0.21
Physics CNN	7.27M	14.88±0.28	14.54±0.13	19.00±0.82	9.03±0.61	6.32±1.21	4.37±0.14	17.20±0.26
DnCNN	0.14M	15.29±0.14	14.84±0.12	18.25±0.13	11.24±0.31	8.79±0.34	5.18±0.11	17.18±0.18
Pix2Pix cGAN	48.84M	16.48±0.18	15.71±0.21	21.24±0.14	16.26±0.57	13.27±0.67	4.74±0.11	19.82±0.36
Attention UNet	7.85M	19.29±0.32	18.20±0.33	25.37±0.23	19.77±1.19	17.90±0.33	7.57±0.29	22.47±0.36
UNet	7.76M	19.30±0.31	18.24±0.30	25.06±0.20	21.08±1.65	17.28±0.34	7.58±0.27	22.49±0.37
ResUNet	8.11M	19.73±0.13	18.69±0.09	25.70±0.35	19.60±1.12	18.04±0.89	8.06±0.08	22.85±0.19
UNet-L	31.04M	20.49±0.11	19.35±0.10	**27.22±0.47**	**21.65±0.97**	**20.27±0.21**	8.79±0.12	23.67±0.10
Attention UNet-L	31.39M	**20.51±0.13**	**19.36±0.13**	**27.22±0.30**	**22.02±2.56**	**20.30±0.13**	**8.80±0.13**	**23.70±0.13**
ResUNet-L	32.44M	**20.70±0.04**	**19.61±0.02**	**27.18±0.14**	20.74±0.80	19.41±0.29	**9.11±0.05**	**23.74±0.04**

H.4 Multiple attenuation

Table 43: Marine Multiples Attenuation Dataset.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
DnCNN	0.56M	15.90±0.98	14.56±1.07	16.65±1.08	15.82±0.86	11.70±0.58	-16.14±1.35	20.97±0.37
DNNDAT	37.07M	18.51±0.46	19.18±1.12	19.21±0.68	18.71±0.40	15.36±1.19	-13.83±0.77	25.03±0.24
UNet	7.76M	20.52±0.58	22.65±0.44	21.69±0.63	20.05±0.64	15.17±0.35	-11.49±0.72	26.35±0.32
ResUNet	8.11M	20.77±0.71	22.64±0.47	21.92±0.67	20.82±0.86	15.99±1.06	-10.95±0.86	25.93±0.59
Attention UNet	7.85M	20.85±0.88	22.08±0.67	22.08±0.92	20.94±0.94	15.80±0.90	-11.13±1.06	27.03±0.16
ResUNet-L	32.44M	21.61±0.54	**23.91±0.48**	22.49±0.50	21.73±0.49	17.81±0.74	-10.08±0.61	26.98±0.48
Attention UNet-L	31.39M	22.33±0.46	23.77±0.54	23.23±0.54	**22.71±0.86**	**18.53±0.83**	**-8.94±0.89**	**27.44±0.37**
SAGAN	32.34M	**22.44±1.46**	20.27±1.53	**23.75±1.23**	22.29±2.10	16.42±1.71	-9.08±2.13	26.89±0.61
UNet-L	31.04M	**22.75±0.64**	**24.18±0.47**	**23.68±0.57**	**22.93±0.77**	**18.18±0.63**	**-8.37±0.90**	**27.51±0.25**

H.5 Deblending

Table 44: Common-Receiver Deblending T02_mod.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
ResUNet	8.11M	14.11±0.08	11.37±0.04	14.33±0.06	14.89±0.17	10.76±0.19	6.26±0.07	14.86±0.06
QUNet	21.88M	14.48±0.16	11.51±0.12	14.47±0.15	15.28±0.19	11.31±0.13	6.50±0.12	15.16±0.20
Attention UNet	7.85M	14.56±0.14	11.38±0.17	15.10±0.15	15.20±0.17	9.55±0.31	6.85±0.07	15.30±0.18
UNet	7.76M	14.70±0.24	11.37±0.06	15.07±0.54	15.48±0.21	10.10±0.72	6.50±0.33	15.53±0.28
UNet++	9.05M	16.13±0.10	13.24±0.17	15.95±0.07	16.88±0.27	**13.26±0.09**	8.38±0.19	16.71±0.13
UNet-L	31.04M	16.68±0.64	13.61±0.61	16.58±0.71	17.49±0.55	**13.13±0.16**	8.45±0.59	17.38±0.66
FBResNet	0.45M	17.01±0.00	11.41±0.00	18.05±0.00	17.43±0.00	10.44±0.00	9.27±0.00	17.88±0.00
cDDPM	33.3M	17.24±0.23	**15.25±0.23**	**19.50±0.13**	**18.77±0.16**	11.60±0.32	8.80±0.24	**19.51±0.15**
DnCNN	0.56M	**17.40±0.05**	11.44±0.07	18.48±0.06	17.94±0.06	10.60±0.10	**9.97±0.07**	18.19±0.05
SCRN	0.43M	**19.36±0.04**	**14.43±0.28**	**20.16±0.10**	**19.40±0.08**	13.12±0.17	**11.93±0.09**	**19.94±0.03**

Table 45: Common-Receiver Deblending T02_comp.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
ResUNet	8.11M	13.05±0.10	9.81±0.08	13.35±0.07	13.84±0.15	9.61±0.11	4.73±0.07	13.87±0.08
Attention UNet	7.85M	13.45±0.19	9.98±0.08	13.97±0.17	14.15±0.14	8.51±0.21	5.19±0.13	14.21±0.21
QUNet	21.88M	13.51±0.04	10.05±0.05	13.62±0.07	14.27±0.06	10.20±0.10	5.02±0.08	14.26±0.03
UNet	7.76M	13.72±0.03	10.15±0.12	14.19±0.08	14.39±0.02	9.03±0.40	5.23±0.22	14.56±0.04
UNet++	9.05M	15.24±0.27	11.90±0.31	15.24±0.25	15.81±0.24	**12.13±0.21**	7.04±0.23	15.90±0.27
FBResNet	0.45M	15.45±0.00	10.03±0.00	16.30±0.00	15.98±0.00	9.49±0.00	7.32±0.00	16.46±0.00
UNet-L	31.04M	15.58±0.73	12.19±0.72	15.63±0.79	16.21±0.64	11.89±0.29	7.01±0.74	16.34±0.74
DnCNN	0.56M	15.91±0.03	10.11±0.14	16.88±0.05	16.51±0.07	9.43±0.11	**7.99±0.08**	16.89±0.06
cDDPM	33.3M	**16.35±0.27**	**14.34±0.35**	**18.07±0.20**	**17.61±0.15**	11.02±0.44	7.89±0.32	**18.29±0.19**
SCRN	0.43M	**17.90±0.12**	**13.34±0.18**	**18.46±0.10**	**18.15±0.19**	**12.26±0.15**	**9.69±0.10**	**18.60±0.12**

Table 46: Common-Receiver Deblending T02_simp.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
ResUNet	8.11M	16.55±0.16	13.52±0.07	16.81±0.13	17.66±0.24	13.06±0.27	8.37±0.21	17.35±0.15
Attention UNet	7.85M	16.67±0.16	12.86±0.25	17.35±0.06	17.77±0.16	10.99±0.35	9.24±0.04	17.56±0.19
QUNet	21.88M	16.95±0.37	13.81±0.22	16.92±0.37	18.06±0.43	13.65±0.19	8.88±0.56	17.68±0.35
UNet	7.76M	17.04±0.21	13.36±0.41	17.52±0.08	18.14±0.18	12.05±0.77	8.75±0.08	17.92±0.24
UNet++	9.05M	18.89±0.18	15.48±0.24	18.71±0.17	19.84±0.31	**15.93±0.06**	11.34±0.05	19.46±0.22
UNet-L	31.04M	18.95±0.18	15.35±0.08	18.80±0.20	20.10±0.17	15.53±0.06	10.74±0.10	19.67±0.19
cDDPM	33.3M	19.27±0.43	**17.20±0.41**	**22.45±0.12**	**21.70±0.07**	13.39±0.53	10.95±0.59	**22.27±0.13**
FBResNet	0.45M	20.09±0.00	14.24±0.00	21.00±0.00	20.75±0.00	13.61±0.00	12.33±0.00	20.84±0.00
DnCNN	0.56M	**20.41±0.03**	14.52±0.17	21.27±0.07	21.28±0.10	13.90±0.00	**13.08±0.05**	21.05±0.03
SCRN	0.43M	**22.37±0.16**	**17.33±0.42**	**23.07±0.10**	**22.70±0.21**	**15.93±0.28**	**15.37±0.16**	**22.88±0.15**

Table 47: AVO Common-Receiver Deblending T03_avo_mod.

Method	Params	SNR	FB-FRE				EB-FE	
			Low	Mid	High	VHigh	Med.	Strong
cDDPM	33.3M	17.61±0.61	16.18±0.48	21.79±0.48	17.19±0.68	-0.28±0.73	-2.81±0.50	22.72±0.43
ResUNet	8.11M	19.18±0.42	17.23±0.20	22.65±0.35	18.73±0.65	2.09±0.55	-1.38±0.23	22.68±0.56
UNet	7.76M	19.89±0.69	17.80±0.24	22.65±0.79	19.22±0.97	3.54±0.57	0.69±0.27	21.76±0.70
Attention UNet	7.85M	20.58±0.41	18.21±0.19	23.08±0.48	20.20±0.58	4.74±0.37	1.40±0.08	22.22±0.54
FBResNet	0.45M	20.88±0.00	18.25±0.00	23.03±0.00	20.94±0.00	6.09±0.00	-0.34±0.00	25.61±0.00
DnCNN	0.56M	21.14±0.08	18.14±0.08	23.38±0.08	21.50±0.15	6.46±0.18	-0.34±0.08	26.61±0.05
QUNet	21.88M	21.78±0.44	18.87±0.25	24.24±0.24	22.02±0.66	6.11±0.90	1.87±0.09	23.63±0.70
UNet-L	31.04M	23.93±0.14	**19.97±0.26**	25.68±0.14	25.70±0.21	11.56±0.45	**2.50±0.25**	26.80±0.14
UNet++	9.05M	**24.12±0.24**	19.64±0.25	**25.75±0.24**	**26.63±0.22**	**13.52±0.25**	2.42±0.33	**27.40±0.22**
SCRN	0.43M	**25.49±0.04**	**21.74±0.04**	**26.62±0.03**	**27.05±0.09**	**15.68±0.03**	**4.37±0.04**	**28.41±0.00**

H.6 First-arrival picking

Table 48: First-Break Picking Multi-Dataset.

Method	Params	MAE	RC_{norm}	H@1	H@5	H@9
STUNet	71.55M	**1.72±0.10**	**7.03±0.42**	0.408±0.010	**0.932±0.007**	**0.978±0.003**
Attention UNet-L	31.39M	**5.67±0.18**	12.92±2.02	**0.757±0.020**	**0.932±0.006**	**0.960±0.003**
DSU-Net	1.99M	6.40±0.98	17.74±1.58	0.611±0.017	0.910±0.010	0.949±0.007
HUNet	7.76M	6.40±0.63	**11.14±0.64**	0.709±0.017	0.930±0.006	0.959±0.003
UNet-L	31.04M	6.92±2.01	14.18±5.36	0.719±0.053	0.922±0.020	0.952±0.015
ResUNet-L	32.44M	8.28±4.67	16.16±7.01	**0.735±0.029**	0.919±0.018	0.948±0.021
ResUNet	8.11M	8.82±0.89	15.54±0.86	0.684±0.019	0.912±0.006	0.944±0.004
UNet	7.76M	9.72±2.25	19.28±3.80	0.657±0.016	0.905±0.007	0.938±0.007
Attention UNet	7.85M	12.00±4.04	23.98±9.21	0.667±0.031	0.901±0.021	0.931±0.019
DnCNN Seg	0.56M	64.16±10.19	149.93±14.12	0.397±0.029	0.609±0.046	0.630±0.048

Table 49: Brunswick First-Break Picking.

Method	Params	MAE	RC_{norm}	H@1	H@5	H@9
STUNet	71.55M	**0.83±0.04**	**3.51±0.62**	0.603±0.010	**0.988±0.002**	**0.995±0.001**
HUNet	7.76M	**2.02±0.45**	5.33±1.19	0.807±0.022	**0.984±0.004**	**0.989±0.003**
ResUNet-L	32.44M	2.19±0.76	**4.74±0.69**	**0.840±0.013**	0.983±0.005	0.988±0.004
UNet-L	31.04M	2.25±0.30	5.12±1.62	0.838±0.014	0.983±0.002	0.988±0.002
Attention UNet-L	31.39M	2.26±0.52	5.79±1.04	**0.853±0.014**	0.983±0.003	0.988±0.003
UNet	7.76M	2.30±0.83	5.52±1.50	0.803±0.033	0.981±0.005	0.987±0.004
ResUNet	8.11M	2.86±0.87	6.11±1.17	0.820±0.031	0.979±0.005	0.984±0.004
Attention UNet	7.85M	3.00±0.70	5.78±1.03	0.822±0.022	0.979±0.004	0.984±0.003
DSU-Net	1.99M	5.23±0.76	13.70±1.28	0.720±0.032	0.964±0.006	0.971±0.004
DnCNN Seg	0.56M	37.16±9.74	81.72±24.68	0.564±0.047	0.764±0.056	0.771±0.056

Table 50: Halfmile First-Break Picking.

Method	Params	MAE	RC_{norm}	H@1	H@5	H@9
STUNet	71.55M	**1.04±0.07**	**5.25±0.13**	0.402±0.006	**0.965±0.004**	**0.993±0.002**
Attention UNet-L	31.39M	**2.45±0.47**	**11.62±1.69**	**0.728±0.021**	**0.956±0.003**	**0.983±0.003**
ResUNet-L	32.44M	2.58±0.86	13.17±3.89	**0.710±0.009**	0.955±0.004	0.983±0.004
UNet-L	31.04M	2.66±1.00	11.70±4.06	0.699±0.025	0.955±0.007	0.982±0.006
HUNet	7.76M	2.93±1.42	11.73±2.79	0.617±0.022	0.955±0.009	0.981±0.009
ResUNet	8.11M	3.27±0.79	13.63±3.27	0.647±0.028	0.951±0.006	0.978±0.005
Attention UNet	7.85M	3.93±2.29	14.78±7.02	0.639±0.041	0.948±0.012	0.975±0.012
UNet	7.76M	4.23±2.95	16.36±7.56	0.631±0.027	0.947±0.016	0.974±0.016
DSU-Net	1.99M	7.60±2.87	24.49±9.25	0.507±0.041	0.922±0.018	0.950±0.018
DnCNN Seg	0.56M	59.99±4.60	175.18±11.13	0.229±0.012	0.526±0.025	0.545±0.028

Table 51: Lalor First-Break Picking.

Method	Params	MAE	RC_{norm}	H@1	H@5	H@9
STUNet	71.55M	**2.27±0.31**	**18.78±0.94**	0.236±0.022	**0.869±0.021**	**0.962±0.009**
DSU-Net	1.99M	**5.69±0.72**	**21.85±4.43**	0.545±0.011	**0.885±0.006**	**0.949±0.005**
Attention UNet	7.85M	13.18±0.83	31.88±0.44	0.616±0.022	0.860±0.009	0.920±0.004
Attention UNet-L	31.39M	15.63±1.85	41.21±3.26	**0.672±0.042**	0.858±0.011	0.912±0.010
HUNet	7.76M	17.05±8.10	38.08±26.88	0.507±0.120	0.840±0.050	0.904±0.041
UNet	7.76M	17.93±4.32	42.52±11.56	0.544±0.031	0.836±0.012	0.897±0.015
ResUNet-L	32.44M	18.35±1.51	47.30±3.71	**0.633±0.042**	0.842±0.011	0.898±0.009
UNet-L	31.04M	19.37±12.79	50.52±35.13	0.583±0.103	0.836±0.058	0.893±0.055
ResUNet	8.11M	21.73±9.68	54.66±26.84	0.488±0.036	0.812±0.038	0.877±0.040
DnCNN Seg	0.56M	89.85±12.61	158.74±18.16	0.265±0.020	0.487±0.057	0.521±0.061