ONE OR MORE? PHYSICS-GUIDED SUPER-RESOLUTION DECOMPOSITION OF UNDERWATER MULTISPECTRAL LIDAR ECHOES

Anonymous authors
Paper under double-blind review

ABSTRACT

Underwater multispectral LiDAR jointly captures range and spectral information, providing multidimensional observations for detailed underwater sensing. However, echoes from closely spaced targets can overlap within a single waveform, while water-induced temporal broadening further reduces peak separability. These effects make accurate target counting and physical spacing estimation challenging. To study this problem under realistic conditions, we develop an underwater multispectral LiDAR system and use it to acquire two real-world full-waveform datasets: a single-target reference library spanning target ranges, reflectances, and laser intensities, and a multiband multi-target library covering different degrees of echo overlap with physical spacing annotations. We propose Aquatic Super-Resolution (AquaSR), a physics-guided framework for underwater echo super-resolution and decomposition. AquaSR incorporates a backbone-agnostic prior-constrained joint multiband decomposition (PC-JMD) module. Its central principle is to share target count and physical spacing across bands while retaining band-specific echo shapes, amplitudes, first-target delays, and backgrounds. A frozen decoder pretrained on the single-target library provides a continuous single-echo shape prior that constrains component reconstruction. The module determines target count by comparing reliability-weighted reconstruction errors under single- and dual-target hypotheses and estimates shared spacing by fusing cross-band geometric evidence. Physical-attribute experts then refine the spacing using echo morphology and energy features, followed by component reconstruction. Experiments on measured data show that AquaSR outperforms representative classical waveform decomposition methods and deep learning baselines and recovers target range structure at sub-pulse-width scales. Experiments with multiple backbone configurations further demonstrate the cross-architecture compatibility of PC-JMD. These results support precise ranging and structural analysis of closely spaced underwater targets.

1 INTRODUCTION

Oceans cover approximately 71% of Earth's surface, yet direct, high-resolution observations of their interior remain limited. Large areas in global seafloor datasets still rely on sparse survey lines or indirect estimates. Systematic, detailed seafloor mapping therefore remains an ongoing foundational task (Wölfl et al., 2019; Mayer et al., 2018). Coral reef monitoring, underwater infrastructure inspection, archaeological site documentation, and autonomous underwater navigation and operations further require sensors to recover fine-scale geometry and physical properties at close range (Remmers et al., 2024; Nauert & Kampmann, 2023; Wright et al., 2020; Paull et al., 2014).

Current mainstream underwater sensing modalities each have their own limitations. Visible-light cameras provide high-resolution texture and color, but absorption and scattering rapidly reduce contrast, color fidelity, and effective range (Jian et al., 2021; Shuang et al., 2024). Sonar is less sensitive to water transparency and supports wide-area search and geometric sensing. Its spatial resolution is typically lower than that of close-range optical imaging, and it provides limited direct access to material-related information such as color or spectra (Ferreira et al., 2016; Sørensen et al., 2023). Active multispectral ranging combines narrow-beam illumination with time-of-flight measurements of absolute range. It also records return intensities and waveform features at multiple

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.

wavelengths, capturing both fine geometry and wavelength-dependent reflectance (Chen et al., 2024; Ma et al., 2025; Hakala et al., 2012; Halimi et al., 2021). A basic problem arises before these measurements can support structural analysis: does a broadened peak represent one target or several closely spaced targets? Water scattering and the receiver response can broaden a single echo, create a shoulder, and prolong its tail (Jaffe, 2015; Mobley, 1994). Superposition of neighboring returns can produce similar changes. Consequently, counting peaks is not equivalent to counting targets, and locating peaks is not sufficient to recover their spacing.

This ambiguity is particularly acute when two returns merge into a single observed peak. Conventional decomposition fits Gaussian or more flexible component families, often using visible waveform features to initialize their locations (Hofton et al., 2000; Mallet & Bretar, 2009; Shen et al., 2017; Wagner et al., 2006; Zhou et al., 2026). Yet an overly flexible single-echo model can absorb a genuine second target into its width or tail, while an overly restrictive model can split ordinary underwater response variation into spurious targets. Neither a second local maximum nor a small reconstruction error, considered alone, establishes the correct target structure.

Our starting observation is that spacing information can survive peak merging. In measured underwater waveforms, changing target spacing alters shoulders, peak flatness, curvature, and trailing energy before a second peak becomes apparent. Different wavelengths respond differently to the same configuration. These complementary observations motivate the central modeling principle of this work: geometry is shared across wavelengths, while echo morphology is wavelength-specific. Cross-band consistency should constrain target count and physical spacing, not force all bands to have the same waveform shape. The curvature argument and measured morphology evidence underlying this factorization are given in Appendix A; the formal multiband observation model is stated in Section 3.1.

We develop Aquatic Super-Resolution (AquaSR) around this principle. A measured single-target library defines a continuous single-echo prior, constraining which shapes may explain an individual target. A parameter-shared encoder extracts evidence from each band; robust consensus on a physical spacing grid links their geometric estimates. Band-specific latent shapes and amplitudes generate complete waveforms under one- and two-target hypotheses, and their residuals determine target count. Finally, a physical-attribute correction uses local slope, curvature, and energy cues to refine spacing while preserving the count decision. This separates response variation, shared geometry, and count selection within a generative prior-constrained decomposition framework.

Our contributions are:

1. We construct two complementary underwater multispectral LiDAR waveform datasets: a single-target reference library for learning response priors, and a multi-target multi-band library with spacing annotations for evaluating target count, physical spacing, and overlapping-echo decomposition.
2. AquaSR introduces a backbone-agnostic PC-JMD module that combines a frozen generative prior with cross-band geometric consensus to infer shared target count and spacing, refined by six physical-attribute experts.
3. We benchmark AquaSR on measured data against classical and learning-based methods, with multi-seed PC-JMD ablations and backbone replacements. AquaSR achieves highly accurate single-/dual-target classification on the current test set and the lowest mean spacing error.

2 Related Work

Waveform decomposition and multispectral constraints. Gaussian decomposition (Hofton et al., 2000; Wagner et al., 2006), B-spline modeling (Shen et al., 2017), and adaptive asymmetric Gaussian decomposition (AAGD) (Zhou et al., 2026) represent individual returns with prescribed response families. Classical fitting uses least-squares, variable-projection, and nonlinear optimization (Golub & Pereyra, 1973; Lawson & Hanson, 1974; Marquardt, 1963; Nocedal & Wright, 2006); information criteria and skewed families address model order and asymmetry (Akaike, 1974; Schwarz, 1978; Azzalini, 1985). MSWD (Song et al., 2019) and the Rclonte family (Bai et al., 2024b;a) use cross-wavelength information to recover weak or overlapping components. Spectral range-resolution enhancement also studies neighboring returns that appear as one pulse (Xia et al., 2023). AquaSR

instead learns its response family from measured underwater single echoes and imposes a common physical spacing *before* component generation, while retaining band-specific shapes.

Learning-based decomposition and generative priors. AFD-Net classifies component count and predicts waveforms using attention and a U-Net, with training primarily based on synthetic data (Liu & Ke, 2022). AquaSR ties counting to competing generative explanations of measured multiband observations. Underwater imaging and neural rendering model attenuation, backscatter, and scattering (Akkaynak & Treibitz, 2018; 2019; Levy et al., 2023), while acoustic–optical fusion complements geometry (Qadri et al., 2024). Architectural, generative, and variational priors constrain inverse problems (Ulyanov et al., 2018; Bora et al., 2017; Kingma & Welling, 2014); Wasserstein autoencoders and maximum mean discrepancy provide the latent-distribution framework used for our single-echo prior (Tolstikhin et al., 2018; Gretton et al., 2012). Underwater and multispectral Bayesian reconstruction similarly combines physical observation models with signal structure (Halimi et al., 2021).

Structured super-resolution. Super-resolution aims to recover fine-scale latent structure from measurements whose native resolution is insufficient to separate closely spaced sources (Candès & Fernandez-Granda, 2014). Related model-based and optimization-inspired approaches incorporate explicit forward models, structural priors, and iterative inference into learned reconstruction (Gregor & LeCun, 2010; Adler & Öktem, 2018; Aggarwal et al., 2019; Ongie et al., 2020; Beck & Teboulle, 2009), while set-based formulations naturally represent multiple latent components (Carion et al., 2020). In full-waveform hyperspectral LiDAR, closely spaced targets can produce strongly overlapping returns that merge into an apparently unimodal "pseudo-monopulse," with spectral diversity providing additional cues for resolving the underlying targets (Xia et al., 2023). AquaSR addresses this unresolved-overlap regime by using a learned single-echo prior and cross-band shared geometry to infer target multiplicity, estimate inter-target spacing, and decompose individual echo components even when distinct waveform peaks are not directly observable.

3 METHOD

3.1 OVERVIEW

AquaSR couples a measured single-echo prior with prior-constrained joint multiband decomposition (PC-JMD). Figure 1 shows the full architecture and Figure 2 details PC-JMD: infer common geometry, generate wavelength-specific components, compare one- and two-target explanations, and refine dual-target spacing. The central factorization is

$$\Theta_{\text{shared}} = \{K, g\}, \qquad \Theta_m = \{t_{m1}, (z_{mk}, a_{mk})_{k=1}^{K}, b_{m0}, b_{m1}\}, \tag{1}$$

so target count and physical spacing are shared, while delay, morphology, amplitude, and background remain band-specific. A compact multiband forward model is

$$\begin{aligned} x_m[n] &= \sum_{k=1}^{K} a_{mk}\, \mathcal{S}_{t_{mk}-t_0}[D_\psi(z_{mk})][n] + b_{m0} + b_{m1} u_n + \varepsilon_m[n], \\ t_{m2} &= t_{m1} + g/s_m, \qquad s_m = v_m \Delta t/2, \qquad g > 0 \quad (K = 2), \end{aligned} \tag{2}$$

where centers use sample coordinates and s_m converts physical spacing to a band-specific sample offset. Appendix A gives the curvature-based identifiability motivation for this factorization, and Appendix C gives the detailed parameterizations and training objectives.

3.2 SINGLE-ECHO PRIOR

A fixed template cannot cover all measured underwater widths, shoulders, and tails, whereas unrestricted components can explain incorrect geometry. We therefore train a denoising MMD-WAE on measured single-target echoes (Tolstikhin et al., 2018; Gretton et al., 2012). Baseline removal, peak alignment, and unit-energy normalization separate morphology from background, delay, and amplitude. The decoder combines an analytic negative-polarity core, a smooth causal tail, and a

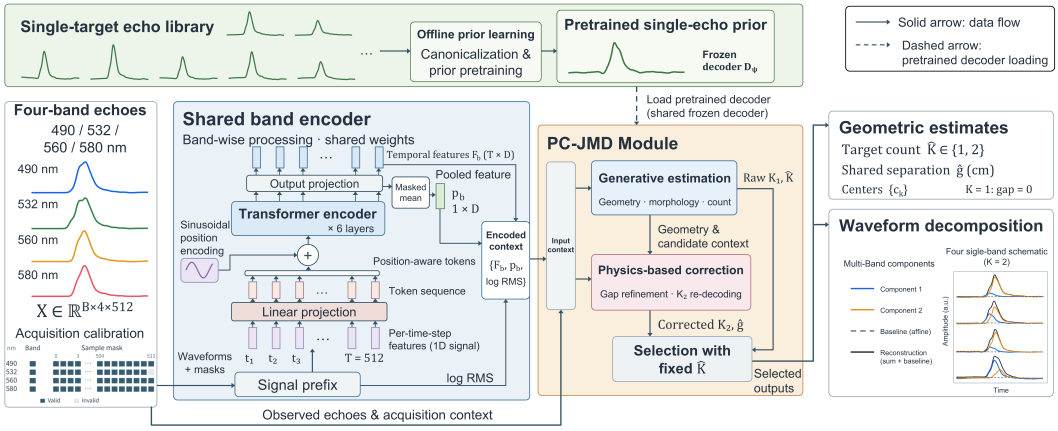

Figure 1: Full AquaSR architecture. A parameter-shared encoder processes each band independently. PC-JMD combines common physical spacing with band-specific shapes and amplitudes. Solid arrows denote data flow; the dashed arrow loads the frozen, pretrained single-echo decoder. Waveforms illustrate the computation.

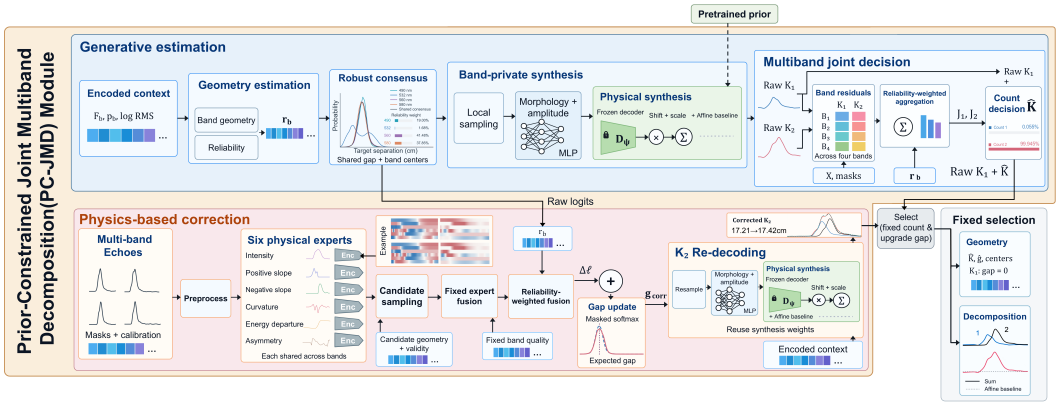

Figure 2: PC-JMD internal structure. The estimation branch fuses spacing candidates, generates band-specific components, and compares reconstruction residuals to select target count. Six physical-attribute experts then refine spacing and regenerate dual-target components, keeping the count and single-target hypothesis fixed.

bounded learned residual,

$$D_\psi(z) = \mathcal{N}\Big[W \odot \Big(- \big[(1-\rho(z))G_{\sigma(z)} + \rho(z)C_{\sigma(z)}\big] + \alpha_R \tanh R_\psi(z)\Big)\Big], \tag{3}$$

where $\mathcal{N}$ canonicalizes polarity, baseline, center, ringing, and energy. The decoder parameters ψ are then frozen: decomposition learns to select plausible shapes rather than changing the shape family to absorb localization errors. Gradients still pass through D_ψ to the shape head. Appendix C.2 gives the detailed normalization and shift operators.

3.3 Shared Cross-Band Geometry Estimation

A parameter-shared temporal encoder processes each band separately using a multiscale convolutional prefix and Transformer layers (Vaswani et al., 2017). From the masked pooled feature $\overline{F}_m$ and log-RMS amplitude η_m, a reliability head gives

$$r_m = \frac{\exp f_r([\overline{F}_m, \eta_m])}{\sum_{b\in\mathcal{V}} \exp f_r([\overline{F}_b, \eta_b])}, \qquad \ell_{mj} = F^{\text{geo}}_{m,2}(t_{m1} + g_j/s_m), \tag{4}$$

where $g_j \in \mathcal{G}$ is a physical spacing candidate. Sampling the second-center evidence at $t_{m1} + g_j/s_m$ maps band-dependent propagation scales to a common physical coordinate.

To limit an overconfident but inconsistent band, per-band log probabilities are clipped around their cross-band median and disagreement is penalized. If $\overline{u}_j$ and V_j are the reliability-weighted mean and variance of the clipped scores, consensus gives

$$c_j = \overline{u}_j - \lambda_v V_j, \qquad p_j = \mathrm{MaskedSoftmax}_j(c/T_g), \qquad \widehat{g} = \sum_j p_j g_j. \tag{5}$$

All bands use $\widehat{g}$ in the shared-spacing observation model of Equation (2); their first centers need not coincide. Appendix C.3 specifies clipping and validity handling.

3.4 Band-Specific Morphology Reconstruction

The same target can have different widths, shoulders, and trailing energy at different wavelengths. Local features at the inferred centers and pooled waveform context form ξ_{mk}, from which the shape head predicts an independent latent code and nonnegative amplitude,

$$z_{mk} = a_z \tanh f_z(\xi_{mk}), \qquad a_{mk} = \mathrm{softplus}\, f_a(\xi_{mk}), \tag{6}$$

with a separate affine background for each band. The frozen decoder and fractional shifts then generate

$$\widehat{x}_m^{(K)}[n] = \sum_{k=1}^{K} a_{mk}\, \mathcal{S}_{t_{mk}-t_0}[D_\psi(z_{mk})][n] + b_{m0} + b_{m1} u_n, \tag{7}$$

where dual-target locations obey the shared $\widehat{g}$. Geometry is therefore coupled without tying shape codes, amplitudes, or backgrounds across bands.

3.5 Reconstruction-Based Target Counting

Target count is selected by comparing complete-waveform explanations rather than by a standalone feature classifier. Let $\beta_m[n]$ be a masked affine baseline fit and $\mu_m[n]$ the sample-validity mask. We normalize the residual by baseline-removed observation energy,

$$J_m^{(K)} = \frac{\sum_n \mu_m[n](x_m[n] - \widehat{x}_m^{(K)}[n])^2}{\max(\sum_n \mu_m[n](x_m[n] - \beta_m[n])^2, \epsilon)}, \qquad J_K = \sum_m r_m J_m^{(K)}. \tag{8}$$

The count probability is

$$P(K = 2 \mid x) = \mathrm{sigmoid}\left(\frac{J_1 - J_2 - \gamma}{T_K}\right), \qquad \gamma = \mathrm{softplus}(\gamma_0), \quad T_K = \mathrm{softplus}(\tau_0) + \epsilon. \tag{9}$$

The positive complexity increment γ offsets the fitting advantage of an extra component. The larger hypothesis probability determines $\widehat{K}$, coupling counting supervision to centers, shapes, amplitudes, and residual evidence. Appendix C.4 gives numerical safeguards.

3.6 Physical-Attribute Correction

Residual geometric error may remain after consensus. After baseline removal, RMS normalization, and smoothing, six experts use intensity, positive and negative slopes, curvature, cumulative-energy deviation, and local energy asymmetry. Denoting the normalized expert residual by δ_{mej}, their cross-band correction is

$$\Delta_{mj} = \frac{q_m}{6} \sum_{e=1}^{6} \delta_{mej}, \qquad \Delta_j = \frac{\sum_m r_m v_{mj} \Delta_{mj}}{\max(\sum_m r_m v_{mj}, \epsilon)}, \tag{10}$$

where q_m is a training-derived band-quality constant and v_{mj} is candidate validity. The corrected spacing is

$$c'_j = \mathrm{sg}(c_j) + \Delta_j, \qquad p'_j = \mathrm{MaskedSoftmax}_j(c'/T_g), \qquad \widehat{g}' = \sum_j p'_j g_j. \tag{11}$$

For events classified as dual-target, the model retains each first center, updates the second center through $\widehat{g}'$, and regenerates shapes and amplitudes. Single-target outputs and count probabilities remain unchanged. Appendix C.5 defines all six attributes and support masks.

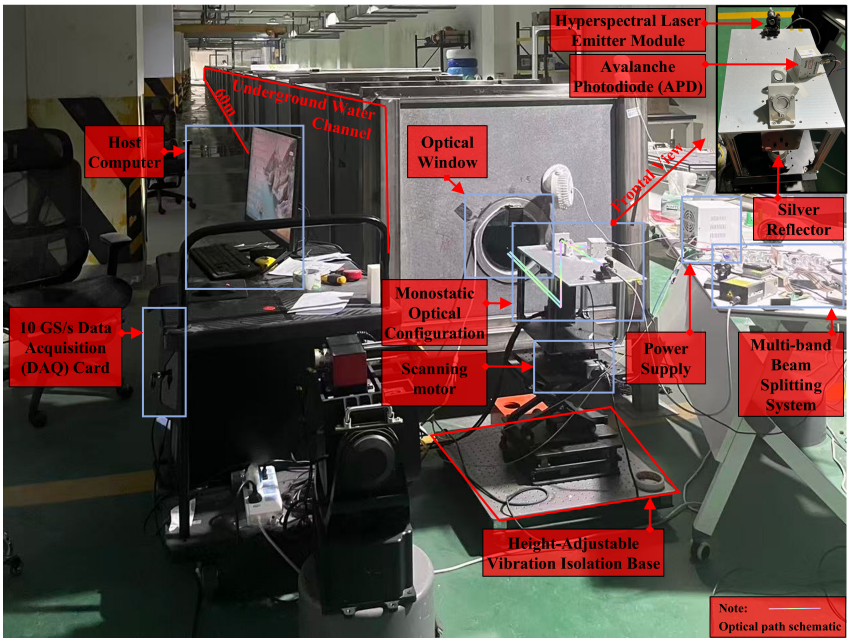

Figure 3: Underwater multispectral LiDAR acquisition system.

3.7 Staged Training

Training has three stages: learn the single-echo prior; freeze that decoder and train the generative estimator; then freeze the estimator and train only the physical experts. The estimator objectives are

$$\mathcal{L}_{\text{syn}} = \mathcal{L}_K + \mathcal{L}_t + \mathcal{L}_g + \mathcal{L}_{\text{rec}} + \lambda_{\text{win}}\mathcal{L}_{\text{win}} + \lambda_{\text{comp}}\mathcal{L}_{\text{comp}} + \lambda_a\mathcal{L}_a + \lambda_z\mathcal{L}_z, \tag{12}$$

$$\mathcal{L}_{\text{real}} = \mathcal{L}_K + \mathcal{L}_g + \lambda_t\mathcal{L}_t + \lambda_{\text{rec}}\mathcal{L}_{\text{rec}} + \mathcal{L}_{\text{win}}, \tag{13}$$

with candidate cross-entropy and cumulative-distribution supervision added to both. After freezing the estimator, the corrector uses

$$\mathcal{L}_B = \mathcal{L}_g + \lambda_{\text{CE}}\mathcal{L}_{\text{CE}} + \mathcal{L}_{\text{CDF}} + \lambda_{\text{band}}\mathcal{L}_{\text{band}} + \lambda_{\text{con}}\mathcal{L}_{\text{con}}. \tag{14}$$

Estimator and corrector training each begin with synthetic data and continue with measured-data fine-tuning. Reconstruction supervision uses the hypothesis associated with the true target count; re-decoding after correction is a no-gradient output step. Appendix C.6 gives the loss definitions and optimization details.

4 Experiments

4.1 Experimental Setup And Main Results

Acquisition platform. We built an integrated underwater multispectral LiDAR platform for transmission, spectral splitting, detection, and synchronized acquisition to acquire our dataset. The overall system architecture is shown in Figure 3. Details regarding the dimensional accuracy and other specifications of the targets are provided in Figure 8 (Appendix D).

Measured data and splits. We collected a single-target dataset covering different distances, different standard target reflectances, and different laser intensities, along with a synchronized multiband library of static target configurations with nominal spacing annotations. Each event contains 490, 532, 560, and 580 nm waveforms. By employing a range-disjoint split, the datasets are partitioned based on target distance, meaning the specific distances used in the test set do not appear in the training or validation sets. Consequently, no test-set data enter prior learning, synthetic-data construction, or fine-tuning. The test set contains both single- and dual-target cases, including closely spaced dual targets with strong overlap as well as cases exhibiting distinct peaks. Appendix D gives acquisition, preprocessing, and target details.

Table 1: Comparison of target counting accuracy and decomposition spacing among various waveform decomposition methods on the same test set. Counting metrics are percentages. D→S is the fraction of true dual targets classified as single. Spacing errors are in cm, are conditional on valid dual-target outputs, and therefore should be interpreted jointly with counting performance. Bias is the mean estimated spacing minus the physical label spacing.

	Target counting				Spacing error			
Method	Overall↑	Single↑	Dual↑	D→S↓	MAE↓	RMSE↓	P80↓	Bias→ 0
Hofton (Hofton et al., 2000)	45.54	54.55	38.60	19.30	12.96	16.67	24.23	−1.61
MSWD (Song et al., 2019)	15.84	6.82	22.81	3.51	21.32	22.24	25.87	14.05
AFD-Net (Liu & Ke, 2022)	54.46	20.45	80.70	**0.00**	8.73	11.45	15.18	−8.62
Rclonte (Bai et al., 2024b)	29.70	15.91	40.35	7.02	17.09	18.10	23.26	15.04
Rclonte-M (Bai et al., 2024a)	22.77	25.00	21.05	1.75	16.66	18.53	24.26	16.66
AAGD (Zhou et al., 2026)	67.33	**100.00**	42.11	57.89	0.68	0.95	1.27	0.32
AquaSR (ours)	**100.00**	**100.00**	**100.00**	**0.00**	**0.48**	**0.68**	**0.85**	**0.25**

Baselines and evaluation. We evaluate our method against Hofton, MSWD, AFD-Net, Rclonte, Rclonte-M, and AAGD (Appendix D.4) on the same test dataset, with the results presented in Table 1. Among these evaluated baselines, only AFD-Net is learning-based, whereas the others are classical waveform decomposition approaches. Baseline spacing errors are conditional on valid dual-target outputs and must be read with per-class counting accuracy. Ablations instead evaluate the dual-target hypothesis on all true dual-target events and report three-seed means ± sample standard deviations.

Main results. Table 1 reveals contrasting decomposition tendencies. AAGD achieves 100.00% single-target accuracy but merges 57.89% of dual targets, the highest rate. This is consistent with adaptive asymmetric fitting and component pruning suppressing spurious splitting while potentially merging overlapping returns. Its conditional MAE (0.68 cm) and bias magnitude (0.32 cm), both lowest among baselines, indicate accurate localization when pairs are resolved. AFD-Net achieves the highest baseline dual-target accuracy (80.70%) with no dual-to-single errors, but tends toward over-decomposition. Its shape-based classifier may interpret underwater shoulders as additional targets, consistent with only 20.45% single-target accuracy; its -8.62 cm bias also indicates systematic spacing underestimation. Hofton's small bias (-1.61 cm) masks a large MAE (12.96 cm), consistent with symmetric Gaussian fitting inadequately representing underwater echo shapes. MSWD has the largest MAE (21.32 cm) and lowest overall counting accuracy (15.84%), while Rclonte and Rclonte-M overestimate spacing (biases: 15.04/16.66 cm), suggesting that cross-band component matching remains vulnerable to underwater waveform distortion. AquaSR combines 100.00% single- and dual-target accuracy with the lowest conditional MAE (0.48 cm) and lowest bias (0.25 cm).

Figure 4 contrasts the same methods in two representative resolution regimes. In the 7 cm case, most observed waveforms remain effectively single-peaked, placing the decomposition problem in the super-resolution regime where the two targets are not directly separable from visible peaks. AAGD collapses the event to one component in every band, whereas AquaSR retains a dual-target solution and estimates a shared physical spacing of 8.40 cm. This behavior is consistent with reconstruction-based counting: although the weak second return does not form a distinct local maximum, its presence is supported by the joint reduction in full-waveform residual, while band-private decoder codes allow each wavelength to explain the overlapping return with its own width and tail. In contrast, the 35 cm case exhibits clearly distinguishable dual-peak structure, representing a more conventionally resolvable regime. AquaSR recovers 35.29 cm, with the shared-spacing constraint mainly enforcing cross-band geometric consistency and preventing wavelength-dependent peak shifts from being interpreted as different physical separations. Together, the two cases illustrate that AquaSR extends decomposition from visibly resolved echoes to strongly overlapped, effectively single-peaked echoes.

4.2 Core Ablations

Table 2 compares PC-JMD structural variants under the same evaluation protocol. When spacing is estimated independently in each band, MAE rises from 0.50 ± 0.03 to 0.67 ± 0.14 cm, RMSE from 0.72 ± 0.05 to 0.93 ± 0.23 cm, and P90 from 1.33 ± 0.07 to 1.63 ± 0.32 cm while count accuracy stays at $100.00 \pm 0.00\%$. This pattern is consistent with the intended role of the shared physical coordinate

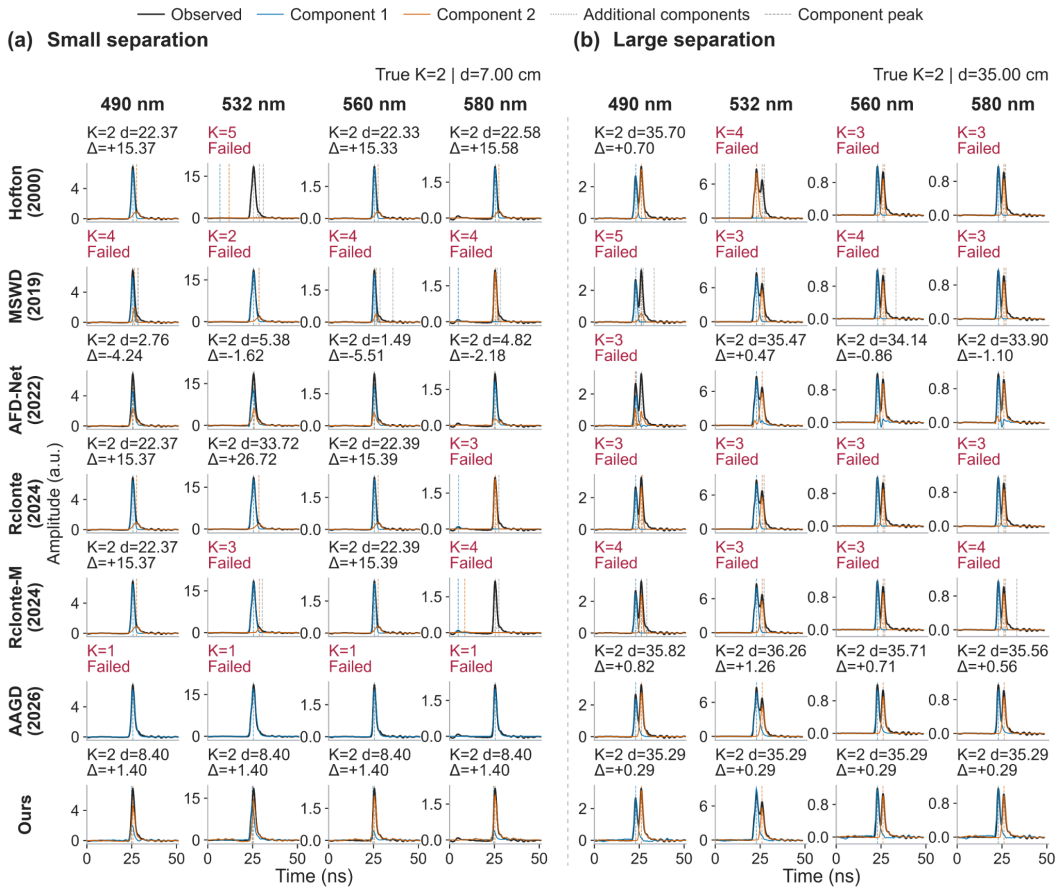

Figure 4: Qualitative four-band decomposition in two overlap regimes. (a) Small separation: nominal 7 cm, where most observed waveforms remain effectively single-peaked. (b) Large separation: nominal 35 cm, where the second return is more visible. Dashed vertical markers indicate component peaks.

Table 2: PC-JMD ablations. Spacing metrics are in cm and count accuracy is in percent. Each entry is mean ± sample standard deviation over three random seeds. Bold and underline mark the best mean in each column, including ties.

Configuration	MAE↓	RMSE↓	P90↓	Count Accuracy↑
Full model	**0.50** ± 0.03	**0.72** ± 0.05	1.33 ± 0.07	**100.00** ± 0.00
Independent spacing in estimation	0.67 ± 0.14	0.93 ± 0.23	1.63 ± 0.32	**100.00** ± 0.00
Fully independent band processing	1.67 ± 0.27	2.51 ± 0.29	3.09 ± 0.44	96.94 ± 3.20
Shared morphology across bands	0.57 ± 0.13	0.77 ± 0.16	**1.31** ± 0.32	**100.00** ± 0.00
Direct feature-based counting	0.63 ± 0.05	0.89 ± 0.05	1.52 ± 0.15	**100.00** ± 0.00
No physical correction	0.55 ± 0.07	0.77 ± 0.10	1.41 ± 0.12	**100.00** ± 0.00

in Equations (4)–(5): wavelength-dependent morphology can shift per-band localization evidence, whereas early consensus constrains those shifts before final localization. Fully independent band processing increases MAE/RMSE/P90 to $1.67 \pm 0.27/2.51 \pm 0.29/3.09 \pm 0.44$ cm and lowers count accuracy to $96.94 \pm 3.20\%$, further supporting a benefit from cross-band evidence under ambiguous waveforms. The shared-morphology variant yields 0.57 ± 0.13 cm MAE and 0.77 ± 0.16 cm RMSE; although its mean P90 is slightly lower than the full model (1.31 versus 1.33 cm), its P90 standard deviation increases from 0.07 to 0.32 cm, about $4.6\times$ larger. This is consistent with the motivation for band-private morphology, because tying z_{mk} across wavelengths can make geometry compensate for genuine band-dependent width, shoulder, and tail differences. Replacing residual-

Table 3: Spacing recovery with six one-dimensional backbones under the same PC-JMD decomposition. Results are mean ± sample standard deviation over three random seeds, in cm. Bold and underline mark the best mean in each column.

Backbone	MAE↓	RMSE↓	P90↓
Transformer (Vaswani et al., 2017)	$\underline{\mathbf{0.50}} \pm 0.03$	$\underline{\mathbf{0.72}} \pm 0.05$	$\underline{\mathbf{1.33}} \pm 0.07$
ResNet (He et al., 2016)	0.83 ± 0.12	1.14 ± 0.12	1.77 ± 0.43
ConvNeXt V2 (Woo et al., 2023)	0.90 ± 0.12	1.28 ± 0.07	2.30 ± 0.14
Mamba-3 MIMO (Lahoti et al., 2026)	1.37 ± 0.29	2.15 ± 0.40	3.32 ± 0.88
Mamba-3 SISO (Lahoti et al., 2026)	1.44 ± 0.19	2.23 ± 0.39	3.69 ± 1.04
ConvNeXt V2, single-scale	2.14 ± 0.15	3.12 ± 0.33	3.44 ± 0.39

based hypothesis selection with direct feature counting leaves count accuracy at $100.00 \pm 0.00\%$ but increases MAE/RMSE/P90 to 0.63 ± 0.05/0.89 ± 0.05/1.52 ± 0.15 cm; equal count accuracy therefore does not imply equal geometric precision. Finally, removing physical correction changes MAE/RMSE/P90 from 0.50 ± 0.03/0.72 ± 0.05/1.33 ± 0.07 to 0.55 ± 0.07/0.77 ± 0.10/1.41 ± 0.12 cm count accuracy unchanged by design, which is consistent with the six experts in Equations (10)–(11) providing a modest post-count geometric refinement. These ablations support the proposed interpretation.

4.3 Backbone Compatibility

Table 3 keeps the PC-JMD interface fixed while changing the temporal backbone. Transformer gives 0.50 ± 0.03 cm MAE, 0.72 ± 0.05 cm RMSE, and 1.33 ± 0.07 cm P90; compared with ResNet, its MAE is lower by 0.33 cm (39.8%). This result is consistent with a benefit from representing long-range waveform relations when candidate scoring must connect the first center with weak shoulder and tail evidence at possible second-center locations. Multiscale ConvNeXt V2 reaches 0.90 ± 0.12 cm MAE, whereas its single-scale version reaches 2.14 ± 0.15 cm with RMSE/P90 of 3.12 ± 0.33/3.44 ± 0.39 cm. The within-family difference is consistent with overlap signatures being distributed across local and broader temporal scales, although architecture and optimization effects cannot be excluded. Mamba-3 MIMO is slightly better than SISO (1.37 versus 1.44 cm MAE; 2.15 versus 2.23 cm RMSE), while both show larger P90 and cross-seed spread than Transformer under the current candidate interface. Because the frozen single-echo decoder, shared count and spacing, and band-private morphology are held fixed, these comparisons characterize backbone behavior within PC-JMD rather than a general ranking of the architectures. Appendix E.1 reports the corresponding runtime measurements.

5 Conclusion

We presented AquaSR for overlapping-echo decomposition in underwater multispectral LiDAR and constructed a measured single-target reference library covering different propagation conditions together with a multiband multi-target dataset with physical spacing annotations. AquaSR learns a single-echo shape prior from single-target waveforms, shares target count and physical spacing across bands, and retains band-specific echo shapes, amplitudes, and backgrounds. It determines target count by comparing single- and dual-target explanations and then refines spacing with physical attributes. Experiments show that overlapping-echo morphology and complementary multiband information improve the accuracy and stability of close-target range-structure recovery; ablations and backbone replacements further support the key design choices and cross-architecture compatibility. Current validation mainly uses static targets under limited underwater conditions; future work should extend to different water conditions to study generalization and capability limits. Overall, AquaSR combines a single-echo physical prior with multiband geometric consistency for fine geometric recovery from strongly overlapping underwater waveforms under the evaluated conditions. Ultimately, it provides a new solution for fine target geometric recovery in underwater limited-resolution observations, and offers a new research perspective for extracting sub-resolution structural information from degraded waveforms under complex medium conditions.

AI USE STATEMENT

Generative AI tools were used to assist with Chinese-to-English translation, language polishing, manuscript-structure editing, and consistency checking. They were also used to assist with the interpretation of some experimental results; any AI-generated interpretations included in the manuscript were reviewed by the authors and evaluated against the experimental evidence before being retained.

Additionally, generative AI tools were used for auxiliary coding and debugging tasks. These tools were not used to implement the core methods proposed in this work. AI-assisted code and its outputs were manually reviewed and verified by the authors. Generative AI tools were also used to convert the authors' completed Chinese manuscript into LaTeX and to adapt its formatting to the ICLR submission requirements. The underlying manuscript content was written by the authors.

Generative AI tools were not used to formulate or refine research hypotheses, design the research methodology or experiments, develop theoretical or conceptual frameworks, formulate mathematical claims, or provide critical components of mathematical proofs.

The authors reviewed all AI-assisted content and take responsibility for the final manuscript, including its text, interpretations, code, and other artifacts.

REFERENCES

Jonas Adler and Ozan Öktem. Learned Primal-Dual Reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1322–1332, 2018. doi: 10.1109/TMI.2018.2799231. URL https://doi.org/10.1109/TMI.2018.2799231.

Hemant K. Aggarwal, Merry P. Mani, and Mathews Jacob. MoDL: Model-Based Deep Learning Architecture for Inverse Problems. *IEEE Transactions on Medical Imaging*, 38(2):394–405, 2019. doi: 10.1109/TMI.2018.2865356. URL https://doi.org/10.1109/TMI.2018.2865356.

Hirotugu Akaike. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974. doi: 10.1109/TAC.1974.1100705. URL https://doi.org/10.1109/TAC.1974.1100705.

Derya Akkaynak and Tali Treibitz. A Revised Underwater Image Formation Model. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6723–6732, 2018.

Derya Akkaynak and Tali Treibitz. Sea-Thru: A Method for Removing Water from Underwater Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1682–1691, 2019.

Adelchi Azzalini. A Class of Distributions Which Includes the Normal Ones. *Scandinavian Journal of Statistics*, 12(2):171–178, 1985.

Jie Bai, Zheng Niu, Kaiyi Bi, Xuebo Yang, Yanru Huang, Yuwen Fu, Mingquan Wu, and Li Wang. Toward an Advanced Method for Full-Waveform Hyperspectral LiDAR Data Processing. *IEEE Transactions on Geoscience and Remote Sensing*, 62:5702516, 2024a. doi: 10.1109/TGRS.2024.3382481. URL https://doi.org/10.1109/TGRS.2024.3382481.

Jie Bai, Zheng Niu, Yanru Huang, Kaiyi Bi, Yuwen Fu, Shuai Gao, Mingquan Wu, and Li Wang. Full-Waveform Hyperspectral LiDAR Data Decomposition via Ranking Central Locations of Natural Target Echoes (Rclonte) at Different Wavelengths. *Remote Sensing of Environment*, 310:114227, 2024b. doi: 10.1016/j.rse.2024.114227. URL https://doi.org/10.1016/j.rse.2024.114227.

Amir Beck and Marc Teboulle. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009. doi: 10.1137/080716542. URL https://doi.org/10.1137/080716542.

Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G. Dimakis. Compressed Sensing Using Generative Models. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 537–546, 2017. URL https://proceedings.mlr.press/v70/bora17a.html.

Emmanuel J. Candès and Carlos Fernandez-Granda. Towards a Mathematical Theory of Super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014. doi: 10.1002/cpa.21455. URL https://doi.org/10.1002/cpa.21455.

Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-End Object Detection with Transformers. In *Computer Vision – ECCV 2020*, pp. 213–229. Springer, 2020.

Yongqiang Chen, Qihui Luo, Shouchuan Guo, Weibiao Chen, Shanjiang Hu, Jian Ma, Yan He, and Yifan Huang. Multispectral LiDAR-Based Underwater Ore Classification Using a Tunable Laser Source. *Optics Communications*, 551:129903, 2024. doi: 10.1016/j.optcom.2023.129903. URL https://doi.org/10.1016/j.optcom.2023.129903.

Fausto Ferreira, Diogo Machado, Gabriele Ferri, Samantha Dugelay, and John Potter. Underwater Optical and Acoustic Imaging: A Time for Fusion? A Brief Overview of the State-of-the-Art. In *OCEANS 2016 MTS/IEEE Monterey*, pp. 1–6, 2016. doi: 10.1109/OCEANS.2016.7761354. URL https://doi.org/10.1109/OCEANS.2016.7761354.

Gene H. Golub and Victor Pereyra. The Differentiation of Pseudo-Inverses and Nonlinear Least Squares Problems Whose Variables Separate. *SIAM Journal on Numerical Analysis*, 10(2):413–432, 1973. doi: 10.1137/0710036. URL https://doi.org/10.1137/0710036.

Karol Gregor and Yann LeCun. Learning Fast Approximations of Sparse Coding. In *Proceedings of the 27th International Conference on Machine Learning*, pp. 399–406, 2010.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A Kernel Two-Sample Test. *Journal of Machine Learning Research*, 13:723–773, 2012.

Teemu Hakala, Juha Suomalainen, Sanna Kaasalainen, and Yuwei Chen. Full Waveform Hyperspectral LiDAR for Terrestrial Laser Scanning. *Optics Express*, 20(7):7119–7127, 2012. doi: 10.1364/OE.20.007119. URL https://doi.org/10.1364/OE.20.007119.

Abderrahim Halimi, Aurora Maccarone, Robert Lamb, Gerald S. Buller, and Stephen McLaughlin. Robust and Guided Bayesian Reconstruction of Single-Photon 3D Lidar Data: Application to Multispectral and Underwater Imaging. *IEEE Transactions on Computational Imaging*, 7:961–974, 2021. doi: 10.1109/TCI.2021.3111572. URL https://doi.org/10.1109/TCI.2021.3111572.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016. doi: 10.1109/CVPR.2016.90. URL https://doi.org/10.1109/CVPR.2016.90.

Michelle A. Hofton, Jean-Bernard Minster, and J. Bryan Blair. Decomposition of Laser Altimeter Waveforms. *IEEE Transactions on Geoscience and Remote Sensing*, 38(4):1989–1996, 2000. doi: 10.1109/36.851780. URL https://doi.org/10.1109/36.851780.

Jules S Jaffe. Underwater Optical Imaging: The Past, the Present, and the Prospects. *IEEE Journal of Oceanic Engineering*, 40(3):683–700, 2015. doi: 10.1109/JOE.2014.2350751. URL https://doi.org/10.1109/JOE.2014.2350751.

Muwei Jian, Xiangyu Liu, Hanjiang Luo, Xiangwei Lu, Hui Yu, and Junyu Dong. Underwater Image Processing and Analysis: A Review. *Signal Processing: Image Communication*, 91:116088, 2021. doi: 10.1016/j.image.2020.116088. URL https://doi.org/10.1016/j.image.2020.116088.

Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. In *International Conference on Learning Representations*, 2014.

Aakash Lahoti, Kevin Y. Li, Berlin Chen, Caitlin Wang, Aviv Bick, J. Zico Kolter, Tri Dao, and Albert Gu. Mamba-3: Improved Sequence Modeling Using State Space Principles. arXiv preprint arXiv:2603.15569, 2026. URL https://arxiv.org/abs/2603.15569.

Charles L. Lawson and Richard J. Hanson. *Solving Least Squares Problems*. Prentice-Hall, Englewood Cliffs, NJ, 1974.

Deborah Levy, Amit Peleg, Naama Pearl, Dan Rosenbaum, Derya Akkaynak, Simon Korman, and Tali Treibitz. SeaThru-NeRF: Neural Radiance Fields in Scattering Media. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 56–65, 2023.

Gangping Liu and Jun Ke. End-to-End Full-Waveform Echo Decomposition Based on Self-Attention Classification and U-Net Decomposition. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:7978–7987, 2022. doi: 10.1109/JSTARS.2022.3203130. URL https://doi.org/10.1109/JSTARS.2022.3203130.

Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A ConvNet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11976–11986, 2022.

Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*, 2019. URL https://arxiv.org/abs/1711.05101.

Yukai Ma, Hao Zhang, Rui Wang, Fashuai Li, Tingting He, Boyu Liu, Yicheng Wang, and Fei Han. Underwater Sphere Classification Using AOTF-Based Multispectral LiDAR. *Photonics*, 12(10):998, 2025. doi: 10.3390/photonics12100998. URL https://doi.org/10.3390/photonics12100998.

Clément Mallet and Frédéric Bretar. Full-Waveform Topographic Lidar: State-of-the-Art. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(1):1–16, 2009. doi: 10.1016/j.isprsjprs.2008.09.007. URL https://doi.org/10.1016/j.isprsjprs.2008.09.007.

Donald W Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963. doi: 10.1137/0111030. URL https://doi.org/10.1137/0111030.

L. Mayer, M. Jakobsson, G. Allen, B. Dorschel, R. Falconer, V. Ferrini, G. Lamarche, H. Snaith, and P. Weatherall. The Nippon Foundation—GEBCO Seabed 2030 Project: The Quest to See the World's Oceans Completely Mapped by 2030. *Geosciences*, 8(2):63, 2018. doi: 10.3390/geosciences8020063. URL https://doi.org/10.3390/geosciences8020063.

Curtis D Mobley. *Light and Water: Radiative Transfer in Natural Waters*. Academic Press, San Diego, CA, 1994.

Franka Nauert and Peter Kampmann. Inspection and Maintenance of Industrial Infrastructure with Autonomous Underwater Robots. *Frontiers in Robotics and AI*, 10:1240276, 2023. doi: 10.3389/frobt.2023.1240276. URL https://doi.org/10.3389/frobt.2023.1240276.

Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, 2 edition, 2006.

Gregory Ongie, Ajil Jalal, Christopher A. Metzler, Richard G. Baraniuk, Alexandros G. Dimakis, and Rebecca Willett. Deep Learning Techniques for Inverse Problems in Imaging. *IEEE Journal on Selected Areas in Information Theory*, 1(1):39–56, 2020. doi: 10.1109/JSAIT.2020.2991563. URL https://doi.org/10.1109/JSAIT.2020.2991563.

Liam Paull, Sajad Saeedi, Mae L. Seto, and Howard Li. AUV Navigation and Localization: A Review. *IEEE Journal of Oceanic Engineering*, 39(1):131–149, 2014. doi: 10.1109/JOE.2013.2278891. URL https://doi.org/10.1109/JOE.2013.2278891.

Mohamad Qadri, Kevin Zhang, Akshay Hinduja, Michael Kaess, Adithya Pediredla, and Christopher A. Metzler. AONeuS: A Neural Rendering Framework for Acoustic-Optical Sensor Fusion. In *SIGGRAPH 2024 Conference Papers*, 2024. URL https://arxiv.org/abs/2402.03309. arXiv:2402.03309.

Tiny Remmers, Alana Grech, Chris Roelfsema, Sophie E. Gordon, Marine Lechene, and Renata Ferrari. Close-Range Underwater Photogrammetry for Coral Reef Ecology: A Systematic Literature Review. *Coral Reefs*, 43(1):35–52, 2024. doi: 10.1007/s00338-023-02445-w. URL https://doi.org/10.1007/s00338-023-02445-w.

Gideon Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464, 1978. doi: 10.1214/aos/1176344136. URL https://doi.org/10.1214/aos/1176344136.

Xiang Shen, Qing-Quan Li, Guofeng Wu, and Jiasong Zhu. Decomposition of LiDAR Waveforms by B-Spline-Based Modeling. *ISPRS Journal of Photogrammetry and Remote Sensing*, 128: 182–191, 2017. doi: 10.1016/j.isprsjprs.2017.03.006. URL https://doi.org/10.1016/j.isprsjprs.2017.03.006.

X. Shuang, J. Zhang, and Y. Tian. Algorithms for Improving the Quality of Underwater Optical Images: A Comprehensive Review. *Signal Processing*, 219:109408, 2024. doi: 10.1016/j.sigpro.2024.109408. URL https://doi.org/10.1016/j.sigpro.2024.109408.

Shalei Song, Binhui Wang, Wei Gong, Zhenwei Chen, Xin Lin, Jia Sun, and Shuo Shi. A New Waveform Decomposition Method for Multispectral LiDAR. *ISPRS Journal of Photogrammetry and Remote Sensing*, 149:40–49, 2019. doi: 10.1016/j.isprsjprs.2019.01.014. URL https://doi.org/10.1016/j.isprsjprs.2019.01.014.

Fredrik Fogh Sørensen, Christian Mai, Malte von Benzon, Simon Pedersen, and Jesper Liniger. Optical and Acoustic Imaging Comparison in a Controlled Underwater Environment. In *OCEANS 2023 Limerick*, pp. 1–8, 2023. doi: 10.1109/OCEANSLimerick52467.2023.10244506. URL https://doi.org/10.1109/OCEANSLimerick52467.2023.10244506.

Ilya Tolstikhin, Olivier Bousquet, Sylvain Gelly, and Bernhard Schölkopf. Wasserstein Auto-Encoders. In *International Conference on Learning Representations*, 2018.

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep Image Prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9446–9454, 2018.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://papers.nips.cc/paper/7181-attention-is-all-you-need.

Wolfgang Wagner, Andreas Ullrich, Vesna Ducic, Thomas Melzer, and Nick Studnicka. Gaussian Decomposition and Calibration of a Novel Small-Footprint Full-Waveform Digitising Airborne Laser Scanner. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(2):100–112, 2006. doi: 10.1016/j.isprsjprs.2005.12.001. URL https://doi.org/10.1016/j.isprsjprs.2005.12.001.

Anne-Cathrin Wölfl, Helen Snaith, Sam Amirebrahimi, Colin W. Devey, Boris Dorschel, Vicki Ferrini, Veerle A. I. Huvenne, Martin Jakobsson, Jennifer Jencks, Gordon Johnston, Geoffroy Lamarche, Larry Mayer, David Millar, Terje Haga Pedersen, Kim Picard, Anja Reitz, Thierry Schmitt, Martin Visbeck, Pauline Weatherall, and Rochelle Wigley. Seafloor Mapping – The Challenge of a Truly Global Ocean Bathymetry. *Frontiers in Marine Science*, 6:283, 2019. doi: 10.3389/fmars.2019.00283. URL https://doi.org/10.3389/fmars.2019.00283.

Sanghyun Woo, Shoubhik Debnath, Ronghang Hu, Xinlei Chen, Zhuang Liu, In So Kweon, and Saining Xie. ConvNeXt V2: Co-Designing and Scaling ConvNets With Masked Autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16133–16142, 2023. doi: 10.1109/CVPR52729.2023.01548. URL https://doi.org/10.1109/CVPR52729.2023.01548.

Anne E. Wright, David L. Conlin, and Steven M. Shope. Assessing the Accuracy of Underwater Photogrammetry for Archaeology: A Comparison of Structure from Motion Photogrammetry and Real Time Kinematic Survey at the East Key Construction Wreck. *Journal of Marine Science and Engineering*, 8(11):849, 2020. doi: 10.3390/jmse8110849. URL https://doi.org/10.3390/jmse8110849.

Yuhao Xia, Shilong Xu, Hui Shao, Ahui Hou, Jiajie Fang, Fei Han, Youlong Chen, Jiaqi Wen, Fashuai Li, Yuwei Chen, and Yihua Hu. Range Resolution Enhanced Method With Spectral Properties for Hyperspectral LiDAR. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 5703517, 2023. doi: 10.1109/TGRS.2023.3301896. URL https://doi.org/10.1109/TGRS.2023.3301896.

Xiang Zhou, Xujia Xie, and Guoqing Zhou. Linear Transfer-Driven Adaptive Asymmetric Gaussian Decomposition for Full-Waveform LiDAR. *Optics Express*, 34(1):675–696, 2026. doi: 10.1364/OE.581680. URL https://doi.org/10.1364/OE.581680.

A Physical Evidence and Identifiability Motivation

This appendix groups the key observation and the physical evidence that motivate AquaSR's factorization. It first shows how target spacing can alter measured multiband morphology before a second peak becomes visible, then separates that observation from the independent broadening and frozen-template discrimination evidence. These results describe the recorded observations and motivate a restricted model; they do not establish unrestricted identifiability under unknown echo shapes.

A.1 Key observation and local curvature intuition

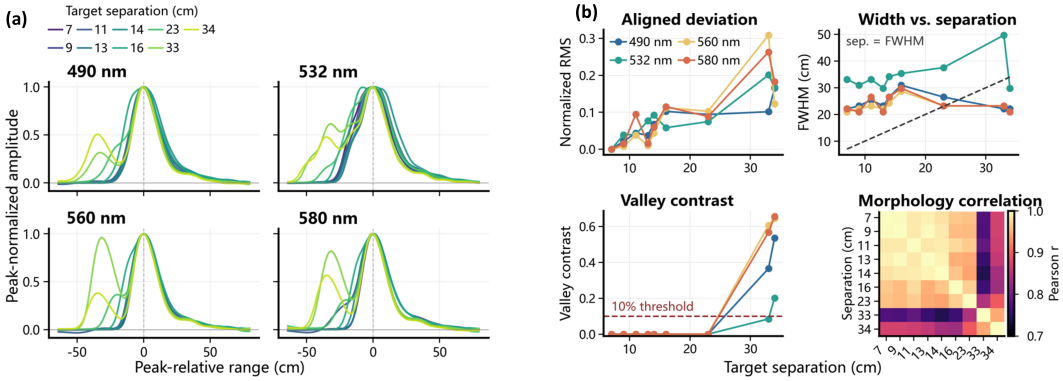

Figure 5: Spacing evidence before peak separation. (a) Peak-aligned four-band measurements at 8 m change morphology with spacing. (b) Shape diagnostics show complementary wavelength responses; the 7–23 cm cases remain predominantly single-peaked. The water-broadening and frozen-template discrimination evidence supporting this observation is reported below in this appendix.

Figure 5(a) shows four-band echo morphology for static targets at 8 m. Most waveforms at spacings of 7–23 cm have no clear pair of peaks, yet their shoulders, peak flatness, local curvature, and post-peak energy exhibit measurable differences. Secondary lobes become more apparent at 33–34 cm. Figure 5(b) summarizes morphological deviation, full width at half maximum, valley contrast, and morphological correlation, showing wavelength-dependent responses to these changes. This creates an opportunity: observations with different response shapes can provide complementary evidence for the same geometry, even when a weak component lacks a visible peak in some bands.

To see why spacing information can survive peak merging, consider two copies of a smooth response h with total amplitude A, amplitude fractions w and $1-w$, and temporal spacing δ. Around their amplitude-weighted center,

$$\frac{y(u)}{A} = h(u) + \frac{w(1-w)}{2}\delta^2 h''(u) + O(\delta^3). \tag{15}$$

The first-order shift cancels, leaving a spacing-dependent curvature perturbation. This term can survive before peak separation, but weakens quadratically as spacing decreases and can resemble single-target width variation. Appendix B gives the derivation and its assumptions. The measured broadening evidence and the separate frozen-template discrimination experiment below further motivate this formulation.

The multiband forward model used by AquaSR is given in Equation (2). The key modeling restriction is that target count and physical spacing are shared across bands, while first-target delay, component morphology, amplitude, background, and noise remain band-specific. Validity masks exclude unavailable samples and bands.

A.2 Water-induced broadening

Under the approximation of short-pulse, incoherent intensity detection, a single-target echo is determined jointly by the transmitted pulse, the receiver-chain response, and the path-delay distribution of

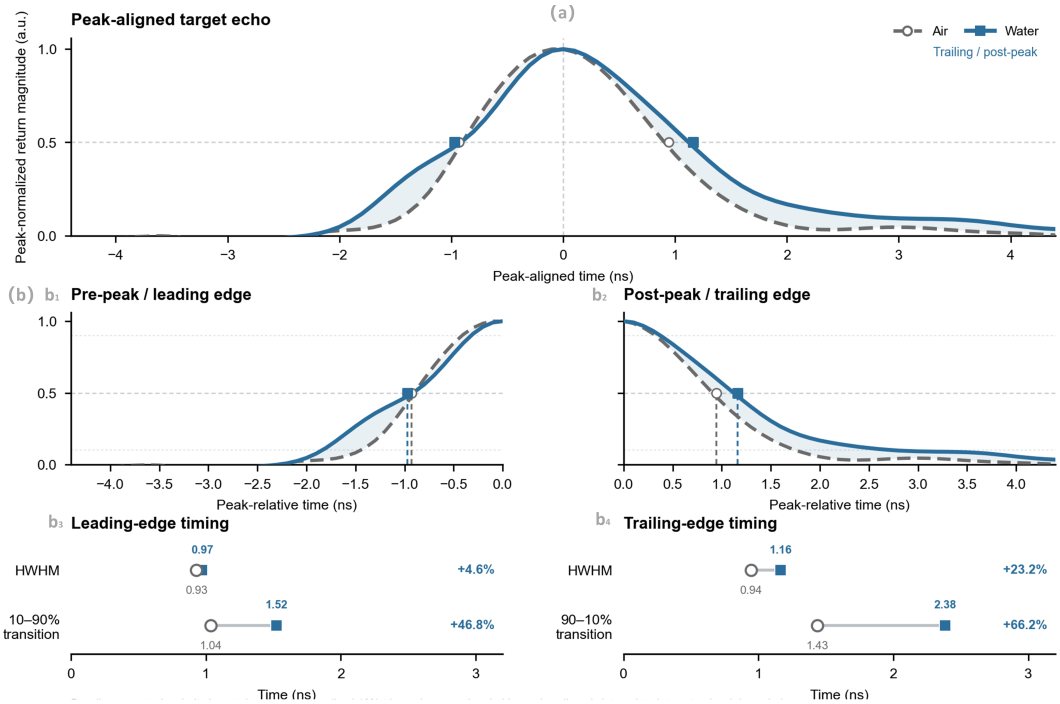

Figure 6: Water propagation changes echo morphology. (a) Peak-aligned, amplitude-normalized 532 nm echoes in air and water at the same 10 m range. (b) Panels b_1 and b_2 show the leading and trailing edges. Panel b_3 compares the leading-edge half-width at half-maximum and 10–90% rise time; b_4 compares the trailing-edge half-width and 90–10% decay time. HWHM denotes half-width at half-maximum. Underwater echoes exhibit more pronounced trailing-edge broadening and prolonged decay.

the water and target. Scattering lengthens the paths of some photons and spreads energy toward later times (Jaffe, 2015; Mobley, 1994). Figure 6(a) compares 532 nm echoes in air and water at the same 10 m range and transmitted energy. Figure 6(b_1–b_4) compares their leading and trailing edges and temporal metrics. After peak alignment and amplitude normalization, the leading-edge half-width at half-maximum changes only from 0.93 to 0.97 ns. The trailing-edge half-width increases from 0.95 to 1.16 ns, and the 90–10% decay time increases from 1.43 to 2.38 ns. Underwater changes are concentrated in the broader shoulder and longer post-peak recovery, where a weak neighboring echo can readily be absorbed.

A.3 Frozen-template target-count discrimination

Learned decomposition relies on the waveform retaining information that distinguishes one target from two when closely spaced returns lack distinct peaks. We test this premise with a frozen-template model comparison using only the 532 nm channel. We compare how single- and dual-echo hypotheses explain measured waveforms independently of the neural network. This experiment evaluates target-count discriminability; later experiments assess component recovery and geometric accuracy.

We construct an empirical response template h_{532} from single-target acquisitions at 14, 16, 18, and 20 m and freeze it before evaluation. The evaluation set contains measured single-target full waveforms at 10 and 12 m, together with dual-target full waveforms whose annotated physical spacings range from 7 to 23 cm. Each independent acquisition is one statistical unit. Repeated waveforms within an acquisition are used only for robust aggregation and noise estimation. We fit the following models to the same 512-sample observation:

$$K = 1: \quad y(t) = a h_{532}(t - \tau) + b + \varepsilon(t),$$
$$K = 2: \quad y(t) = a_1 h_{532}(t - \tau_1) + a_2 h_{532}(t - \tau_2) + b + \varepsilon(t).$$

The $K = 1$ model uses one shifted and scaled measured template; the $K = 2$ model superimposes two templates. Amplitudes are nonnegative, dual-echo delays satisfy $\tau_1 < \tau_2$, and both models include a constant background b. Every acquisition uses the same 7–23 cm spacing search range, with no nominal spacing supplied to the fit. BIC (Schwarz, 1978) penalizes the additional parameters. We define $\Delta\mathrm{BIC} = \mathrm{BIC}(K = 1) - \mathrm{BIC}(K = 2)$, the horizontal coordinate in Figure 7(c). A larger value indicates stronger support for two echoes. The vertical coordinate is the fractional reduction in raw residual energy after adding the second echo. Both models use the same AR(1)-whitened objective.

Figure 7(a,b) shows a representative acquisition at 12 m with a nominal spacing of 13 cm. The measured waveform has one prominent peak. The single-echo fit deviates near the peak and falling edge, while the dual-echo fit follows the overall contour more closely. The $K = 1$ residual has larger positive and negative excursions; the $K = 2$ residual is smaller and closer to zero. Raw residual energy decreases by 74.2%. Overlapping echoes thus retain measurable differences in the peak and shoulder, even when these differences produce no second local maximum.

Figure 7(c) compares these fit differences across single- and dual-target acquisitions. Blue open markers cluster in the lower-left region, indicating limited improvement from adding a second echo to single-target data. Orange-red open markers lie in the upper-right region, indicating a larger improvement for closely spaced dual targets. The horizontal and vertical dashed lines mark the numerical midpoints of the between-class gaps in the two scores for this cohort. The complete waveform provides measurably different support for the two hypotheses even without distinct peaks.

Quantitatively, raw residual energy decreases by at most 7.2% for single targets and by at least 45.4% for closely spaced dual targets. Their $\Delta\mathrm{BIC}$ ranges are also disjoint, and this separation holds within both the 10 m and 12 m groups (Table 4). The score separation is interpreted as cohort-level evidence rather than a formal significance claim.

Table 4: Range-stratified target-count discrimination scores. The upper and lower bounds are the maximum single-target and minimum dual-target $\Delta\mathrm{BIC}$, respectively. Empirical AUROC is computed from continuous $\Delta\mathrm{BIC}$ scores and summarizes post hoc discriminability in this cohort.

Range	Single / dual n	Single-target upper bound	Dual-target lower bound	AUROC
10 m	11 / 6	128.10	419.80	1.00
12 m	15 / 5	79.70	316.30	1.00
Pooled cohort	26 / 11	128.10	316.30	1.00

A second component can also absorb variation within the single-target response, so fit improvement alone should not be treated as a universal target-count rule. Discrimination depends on the size of the fit improvement relative to the range of single-target variation in this cohort. The separation in Figure 7(c) shows that closely spaced overlapping echoes retain usable target-count information under these experimental conditions. Complete-waveform structural differences remain available to learning-based methods when local peaks become indistinct. This observational experiment is distinct from AquaSR's learned count rule and from the held-out-range benchmark in Section 4.

A.4 Scope of the evidence

The cohort contains 26 single-target and 11 closely spaced dual-target independent acquisitions at two ranges. Repeated traces within an acquisition are not treated as independent experimental units. The score gaps and range-stratified AUROCs describe this cohort. The main paper uses this experiment as evidence that the measured single-peaked waveforms retain target-count information, not as a generalization guarantee. The dashed score thresholds in Figure 7 are cohort-gap midpoints, not the learned threshold.

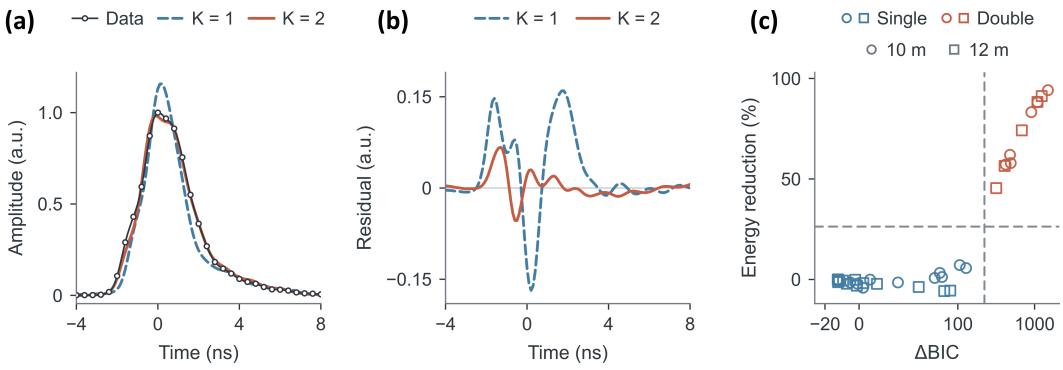

Figure 7: Target-count information retained in single-peaked overlapping echoes. (a) A representative acquisition at 12 m with nominal spacing 13 cm, fitted by single-echo ($K = 1$) and dual-echo ($K = 2$) models. (b) Residuals on a common scale. In (a,b), blue dashed curves denote $K = 1$ and orange-red solid curves denote $K = 2$. Time zero is the measured main peak; all curves are sign-inverted and normalized by its absolute amplitude. The display covers -4 to 8 ns, while statistics use the full 512-sample window. The measured curve is a pointwise 10% trimmed mean within the acquisition. (c) ΔBIC and raw residual-energy reduction for 26 single-target and 11 closely spaced dual-target independent acquisitions. Blue open markers denote single targets and orange-red open markers denote dual targets; circles and squares indicate 10 and 12 m, respectively. Dashed lines mark the numerical midpoints of the between-class gaps in residual reduction and ΔBIC in this cohort. The horizontal axis is linear within ± 40 and logarithmic outside, with ticks labeled in the original units; the vertical axis is linear. All observations are shown, with some overlapping markers. A few negative values arise from differences between the whitened fitting objective and raw residual energy.

B Derivation of the Overlapping-Echo Approximation

B.1 Assumptions and centered superposition

Consider two echoes with a common smooth response h, total amplitude $A > 0$, amplitude fractions w and $1 - w$ with $0 < w < 1$, and separation δ. Their centers, relative to the amplitude-weighted center, are $-(1 - w)\delta$ and $w\delta$. Thus

$$\frac{y(u)}{A} = wh\big(u + (1 - w)\delta\big) + (1 - w)h(u - w\delta). \tag{16}$$

This is a local equal-shape model used for intuition. AquaSR's actual observation model permits different shapes for different targets and bands. Assume first that h is three times differentiable with locally bounded derivatives near u.

B.2 Cancellation and the second-order term

Taylor expansion of the two shifted responses gives

$$h\big(u + (1 - w)\delta\big) = h(u) + (1 - w)\delta h'(u) + \tfrac{1}{2}(1 - w)^2\delta^2 h''(u) + O(\delta^3),$$
$$h(u - w\delta) = h(u) - w\delta h'(u) + \tfrac{1}{2}w^2\delta^2 h''(u) + O(\delta^3).$$

After weighting, the coefficient of $\delta h'(u)$ is $w(1 - w) - (1 - w)w = 0$. The coefficient of $\delta^2 h''(u)$ is

$$\tfrac{1}{2}\big[w(1 - w)^2 + (1 - w)w^2\big] = \tfrac{1}{2}w(1 - w).$$

Consequently,

$$\frac{y(u)}{A} = h(u) + \frac{w(1 - w)}{2}\delta^2 h''(u) + O(\delta^3), \tag{17}$$

which is Equation (15). Centering eliminates a global first-order translation, not the second-order morphological perturbation. For equal amplitudes, $w = 1/2$, its coefficient is $\delta^2/8$.

B.3 Higher-order terms and equal amplitudes

With a locally bounded fourth derivative, expansion through third order yields

$$\frac{y(u)}{A} = h(u) + \frac{w(1-w)}{2}\delta^2 h''(u) + \frac{w(1-w)(1-2w)}{6}\delta^3 h^{(3)}(u) + O(\delta^4). \tag{18}$$

The cubic term vanishes for equal amplitudes. In that symmetric case the superposition is even in δ, so odd-order terms cancel. For $h \in C^4$, the explicit fourth-order form is

$$\frac{h(u+\delta/2) + h(u-\delta/2)}{2} = h(u) + \frac{\delta^2}{8}h''(u) + \frac{\delta^4}{384}h^{(4)}(u) + o(\delta^4). \tag{19}$$

In particular, truncation after the second-order term has an $O(\delta^4)$ remainder under this additional smoothness assumption. Odd-order cancellation does not generally hold for unequal amplitudes.

B.4 Interpretation and limitations

In band m, $\delta_m = g/s_m$, so the same physical spacing perturbs each band's response through its own temporal scale and local curvature. This explains how complementary morphology can precede visible peak separation. It also explains why recovery is difficult: the leading perturbation decreases quadratically with spacing and its coefficient decreases when either component becomes weak. Variation in single-target width can mimic part of this perturbation. Noise, unknown shapes, and finite sampling are not bounded by this calculation, and no uniqueness or estimation-error guarantee follows from it. The measured single-echo prior and shared geometry constrain this ambiguity empirically.

C AquaSR Architecture and Implementation Details

C.1 Forward-model conventions and architecture

Each event contains M bands, L samples per band, sample-validity masks $\mu_m[n]$, and a valid-band set $\mathcal{V}$. The shape code is $z_{mk} \in \mathbb{R}^{d_z}$. Equation (2) is the full forward model; it separates a unit-ℓ_2 component generator from nonnegative amplitude coefficients, fractional-sample placement, and an affine background. The physical conversion $s_m = v_m \Delta t/2$ maps a spacing g in centimeters to a band-specific sample offset g/s_m. Only K and g are shared; first centers are allowed to differ across bands. The default temporal encoder uses a signal prefix, a linear projection, six Transformer layers, and sinusoidal positional encoding. Band-wise processing shares encoder parameters, not activations. Figure 2 summarizes the initial decomposition and correction flow.

C.2 Single-echo prior and decoder

A fixed template has limited coverage of underwater variations in width, shoulders, and tails, whereas unconstrained components can absorb localization errors. We therefore learn a low-dimensional denoising MMD-WAE from measured single-target responses (Tolstikhin et al., 2018; Gretton et al., 2012). During prior training, baseline removal, peak alignment, and unit-energy normalization separate background, global delay, and amplitude from morphology. A ConvNeXt-style encoder preserves temporal order (Liu et al., 2022); decomposition later uses only the frozen decoder.

The decoder is given in Equation (3). Its analytic core contains a compact Gaussian component G and a smooth causal component C with a longer recovery; W tapers the edges and R_ψ learns repeatable residual shape deviations. The canonicalization operator $\mathcal{N}$ sequentially removes the edge baseline, centers the soft negative peak, standardizes polarity, limits positive ringing energy, and applies ℓ_2 normalization. Bounded parameterizations of the analytic core's FWHM and tail fraction restrict the main width and delayed energy to preset response ranges, while α_R limits the residual contribution.

A shape code is predicted separately for each target in each band. The codes share one generative coordinate system but retain independent values. During decomposition training, ψ is fixed while gradients pass through D_ψ to the shape-estimation head. The generator uses Fourier shifts for canonical centering; target placement in the observation window uses linear interpolation and zero padding.

C.3 TEMPORAL ENCODING AND SPACING CONSENSUS

The shared encoder processes bands separately so that wavelength-dependent morphology is preserved during early feature extraction. A multiscale depthwise-convolutional prefix extracts local features, followed by local and channel gates and the Transformer stack (Vaswani et al., 2017). The output is $F_m \in \mathbb{R}^{C\times L}$ with masked mean-pooled context $\overline{F}_m$.

Reliability and candidate sampling follow Equation (4). The first geometric score sequence is converted to t_{m1} by a softargmax within a preset valid time window. The second score sequence is interpolated at $t_{m1} + g_j/s_m$ for each $g_j \in \mathcal{G}$. Candidate validity v_{mj} requires both a valid band and an in-window sampled location. Thus all bands vote in a common physical coordinate even though their temporal scales differ.

Before the consensus in Equation (5), per-band log probabilities are clipped around their cross-band median to limit the effect of a sharply peaked but inconsistent band. Let $u_{mj} = \log \mathrm{softmax}_j(\ell_m)$, and let d_j be the lower median across valid bands. With clipping radius κ,

$$\widetilde{u}_{mj} = d_j + \mathrm{clip}(u_{mj} - d_j, -\kappa, \kappa), \qquad w_{mj} = r_m v_{mj}. \tag{20}$$

The weighted mean and variance of $\widetilde{u}_{mj}$ define the $\overline{u}_j$ and V_j used in Equation (5). The variance term penalizes disputed candidates, and the expectation over the candidate distribution preserves sub-grid spacing information.

C.4 BAND-SPECIFIC GENERATION AND RESIDUAL COUNTING

Given the geometric centers, the shape head samples local features around each target and concatenates them with pooled waveform context, log RMS, normalized center, spacing condition, and a slot indicator. The mappings for z_{mk} and a_{mk} are Equation (6). For a single target, the spacing condition and slot indicator are zero; for two targets, the slot indicators are -1 and $+1$. Initial dual-target positions use the shared $\widehat{g}$, whereas the shape head is additionally conditioned on each band's candidate-distribution expectation $\widehat{g}_m$. This preserves shared component locations while allowing local band evidence to influence morphology. A separate head predicts the affine background from $[\overline{F}_m, \eta_m]$.

Target counting uses Equations (8)–(9). The masked affine fit β_m in the residual denominator provides scale conditioning only; the reconstruction background is still generated by the network. The positive complexity increment γ and temperature T_K are learned jointly with the estimator. The counting loss therefore acts on a probability derived from the competing reconstructions rather than from an independent classifier.

C.5 PHYSICAL-ATTRIBUTE EXPERTS

Shared geometry determines the overall decomposition, but local curvature and energy shifts can remain informative under strong overlap. Each waveform therefore undergoes masked affine-baseline removal, RMS normalization, and Gaussian smoothing to obtain a preprocessed waveform s. Six attributes are computed on valid support and normalized independently. Derivatives use the median of central differences at multiple temporal scales.

Each expert uses channel projection, multiscale depthwise convolution, and LayerNorm. Features are sampled at the candidate second centers $t_{m1} + g_j/s_m$, and a zero-initialized output projection produces candidate-wise correction responses. Let e_{mej} be the response of expert e in band m to candidate j, and h_{mej} its valid-support mask. Centering and standardizing over valid candidates gives

$$\delta_{mej} = h_{mej} \tanh\left(\frac{e_{mej} - \overline{e}_{me}}{\sqrt{\mathrm{Var}(e_{me}) + \epsilon}}\right). \tag{21}$$

These residuals are fused by Equation (10). The fixed band-quality constant q_m is the mean sigmoid-transformed standardized log signal-to-noise ratio of unaugmented training waveforms; event reliability r_m remains input-dependent. The corrected candidate distribution and spacing follow Equation (11). For initially dual-target events, the first center is kept fixed, the second center is

Table 5: Six physical attributes and their decomposition cues. Attributes are normalized independently. Derivatives and local windows are evaluated only where their full support is valid.

Attribute	Construction and morphological meaning
Intensity	$\log(1 + s^2)$: local response intensity and energy distribution near the peak.
Positive slope	Positive part of the central difference: local change during recovery of a negative-polarity echo.
Negative slope	Magnitude of the negative part of the central difference: local change along the descending edge of a negative-polarity echo.
Second derivative	Median multiscale second differences: peak flatness, shoulders, and curvature transitions.
Cumulative-energy deviation	Normalized cumulative squared energy minus progress through valid samples: temporal concentration and displacement of energy.
Local energy asymmetry	Left–right energy contrast within a local window: imbalance between local leading and trailing edges and the tail.

updated using $\widehat{g}'$, and shapes and amplitudes are re-predicted. Single-target outputs and count probabilities are unchanged. Re-decoding is used only to form the final output and does not provide a reconstruction gradient to the corrector.

C.6 Staged objectives and optimization

Training follows the three stages summarized in Section 3.7. The synthetic and measured estimator objectives are Equations (12) and (13). Here $\mathcal{L}_K$ is binary cross-entropy; $\mathcal{L}_t$ is a Smooth-L1 center loss in centimeters; $\mathcal{L}_g$ is a scale-normalized Smooth-L1 spacing loss on true dual-target events; $\mathcal{L}_{\mathrm{rec}}$ constrains the full waveform; and $\mathcal{L}_{\mathrm{win}}$ constrains local windows around target centers. Synthetic data additionally supervise component waveforms, amplitudes, and latent codes through $\mathcal{L}_{\mathrm{comp}}$, $\mathcal{L}_a$, and $\mathcal{L}_z$. Reconstruction uses the hypothesis associated with the true target count so that decomposition supervision does not depend on a hard count decision.

Both estimator losses also include candidate-distribution supervision. Soft labels form a discrete Gaussian centered at the true spacing; $\mathcal{L}_{\mathrm{CE}}$ is candidate cross-entropy and $\mathcal{L}_{\mathrm{CDF}}$ is a Smooth-L1 loss between cumulative candidate distributions. After the estimator is frozen, the six experts are optimized with Equation (14). The term $\mathcal{L}_{\mathrm{band}}$ supervises per-band candidate distributions, and $\mathcal{L}_{\mathrm{con}}$ pulls them toward the stop-gradient mean distribution over valid bands. Estimator and corrector training each begin with synthetic optimization and continue with measured-data fine-tuning; the estimator fine-tuning objective is $\mathcal{L}_{\mathrm{real}} + \lambda_{\mathrm{mix}}\mathcal{L}_{\mathrm{syn}}$, with the analogous candidate-supervision mixture for the correction stage. AdamW is used for optimization (Loshchilov & Hutter, 2019).

D Experimental Setup

D.1 Libraries, platform, and targets

We construct a single-target reference response library and a multi-target multiband dataset. The reference library covers different propagation ranges, reflectances, and transmitted intensities for learning the single-echo shape prior. The multi-target dataset contains static target configurations with nominal physical spacings for evaluating target counting and spacing recovery. Single-target events are acquired synchronously. Dual-target events contain synchronized measurements across bands at the same propagation range and static target configuration. The fixed configuration establishes cross-band geometric correspondence. Observations contain four bands—490, 532, 560, and 580 nm—standardized to 512-sample waveforms, with sample-validity and band-validity masks retained.

Figure 3 shows the underwater multispectral LiDAR platform at one end of a 60 m underground water tank. The platform integrates laser transmission, multiband spectral splitting, an optical window, an avalanche photodiode (APD), a scanning motor, and an acquisition computer. The transmitter and receiver are installed at the same measurement position.

Figure 8 shows stepped targets A and B and a multi-spacing target holder C. Target-surface positions are referenced to the base level at 0 cm. Targets A and B are 3D-printed with a precision of ± 0.02 mm.

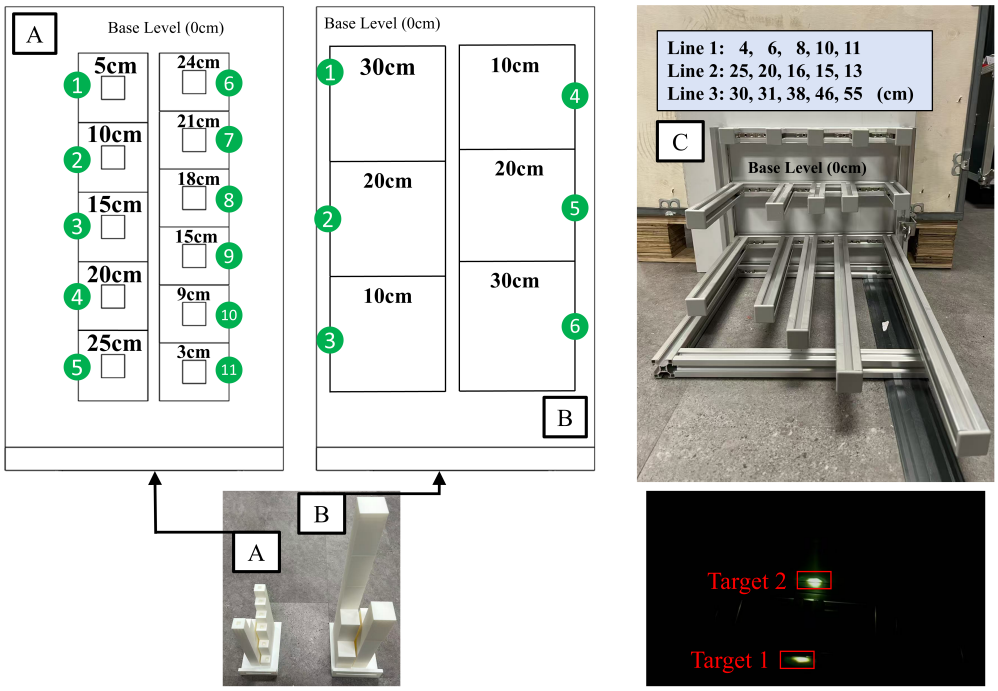

Figure 8: Target configurations and illumination. Stepped targets A and B, multi-spacing target holder C, and the illuminated dual targets. Position dimensions are in centimeters and referenced to the 0 cm base level. Target 1 and Target 2 mark the illuminated target locations.

Target C is machined from aluminum profiles to fixed dimensions with a precision of ± 0.50 mm. During acquisition, the laser spot is adjusted to illuminate the measurement regions on both target surfaces simultaneously and uniformly. The lower-right photograph shows the experimental illumination state. Returns from targets at different ranges combine into an overlapping echo. Experiments are conducted in darkness to avoid stray-light interference.

The hardware photograph in Figure 3 labels a 10 GS/s data-acquisition card. The acquisition hardware rate and the time calibration of the standardized inputs play distinct roles in the processing chain.

D.2 Learning splits and independent physical-evidence cohort

All learning data are split by propagation range: 10 m for independent testing. No 10 m data are used for single-target prior learning, synthetic-data construction, or measured-data fine-tuning. Single-target test cases are included for target-count evaluation. The ablation and backbone experiments report means and sample standard deviations over three random seeds. Their spacing errors use the dual-target hypothesis for every true dual-target event, so architectures are assessed on the same geometric target set.

The frozen-template experiment in Appendix A.3 is separate. It builds a fixed 532 nm template using single-target acquisitions at 14, 16, 18, and 20 m, then evaluates 26 single-target and 11 dual-target independent acquisitions at 10 and 12 m. Those template-construction ranges do not replace the 10 m split used for the learned AquaSR evaluation. The experiment's 7–23 cm search interval is also not a specification of the neural model's full candidate grid or the complete learned-benchmark spacing schedule.

D.3 Preprocessing and supervision

Input waveforms are standardized to 512 samples, retaining sample-validity and band-validity masks. For prior learning, baseline removal, peak alignment, and unit-energy normalization isolate single-echo morphology. Decoder canonicalization uses the operations in Appendix C.2. The counting

residual is normalized by energy after a masked affine baseline fit, whereas reconstructed backgrounds are predicted by the network. Physical attributes use masked affine-baseline removal, RMS normalization, Gaussian smoothing, and support-aware multiscale differences. These are distinct preprocessing roles, not a single interchangeable normalization procedure.

Synthetic supervision includes count, centers, spacing, shape codes, amplitudes, and component waveforms. Measured supervision includes count and geometry, with reconstruction constraining the component explanation.

D.4 Baseline implementations

The evaluated methods are Hofton's Gaussian decomposition (Hofton et al., 2000), MSWD (Song et al., 2019), AFD-Net (Liu & Ke, 2022), Rclonte (Bai et al., 2024b), Rclonte-M (Bai et al., 2024a), and AAGD (Zhou et al., 2026).For AFD-Net, the classifier and attention U-Net are reconstructed from the published architecture diagrams. The 512-sample input is resampled to 256 samples for per-band inference, and outputs are restored to 512 samples. Weights are selected using synthetic validation data. AFD-Net and AquaSR use the same test events, valid bands, and evaluation metrics, excluding 10 m test labels from training and weight selection. AquaSR's measured-data fine-tuning and AFD-Net's synthetic validation selection are reported separately.

D.5 Metrics and reporting scope

Target-count metrics are overall accuracy, single-target accuracy, dual-target accuracy, and the fraction of true dual targets classified as single targets. Balanced accuracy is the mean of the single- and dual-target accuracies. A dual-to-single rate need not equal one minus dual-target accuracy because other counts and failure outputs can occur. Spacing metrics are MAE, RMSE, signed mean error (Bias), and the reported absolute-error quantiles: P80 for baseline comparisons and P90 for structural comparisons. Bias is estimated minus nominal spacing; its ideal value is zero.

Baseline spacing errors are aggregated over each method's valid dual-target outputs, whereas structural comparisons use all true dual-target events. The two protocols therefore quantify different aspects of recovery and should be interpreted separately.



Figure 9: Backbone latency and method processing time. (a) GPU latency of backbones in the same decomposition framework on an RTX 5090, reported as mean ± standard deviation in milliseconds. (b) Mean method-level processing time in milliseconds per four-band event. Circles denote single-threaded CPU timing and diamonds denote RTX 5090 timing. Both panels use logarithmic time axes.

E Additional Experimental Results

The complete counting, spacing, ablation, and backbone metrics are already reported in Tables 1–3 of the main paper. To avoid duplicating those tables and their interpretation, this appendix retains only the additional runtime analysis.

E.1 Runtime and accuracy–latency trade-offs

Figure 9(a,b) reports backbone- and method-level timing under their stated configurations. At method level, AquaSR requires 449.94 ms per four-band event on one Intel Core i7-14700K thread and 35.02 ms on an RTX 5090, corresponding to an approximately $12.85\times$ GPU speedup. Its CPU cost is lower than Hofton and MSWD but higher than Rclonte, Rclonte-M, AAGD, and AFD-Net because the reported AquaSR implementation evaluates constrained component generation, reliability-weighted spacing candidates, one-versus-two reconstruction residuals, and a correction-stage re-decoding. AFD-Net is faster on the GPU at 3.38 ms, while Table 1 reports 20.45% single-target accuracy, 80.70% dual-target accuracy, and 8.73 cm conditional MAE for the evaluated implementation. These numbers show the measured accuracy–latency trade-off under the stated protocols; they do not by themselves isolate which implementation detail causes the runtime difference.

At backbone level, Transformer takes 30.44 ± 0.31 ms and has the lowest geometric errors in Table 3. ResNet18 and the ConvNeXt variants have similar latency with higher errors, while both Mamba-3 adaptations are slower in the reported implementation. Method-level and backbone-level timings use different experimental configurations and should not be numerically combined.