Physically Motivated Distributions for Bayesian Deep Learning of Synthetic Aperture Sonar Imaging Artifacts

Marko Orescanin, *Senior Member, IEEE*, Derek R. Olson, *Senior Member, IEEE*, Dalton Duvio, Haley Jennings, Marc Geilhufe, Roy E. Hansen, *Senior Member, IEEE*, and Narada Warakagoda

***Abstract*—We demonstrate that the choice of weight prior in Bayesian neural networks (BNNs) is a consequential and nontrivial design decision for synthetic aperture sonar (SAS) imaging artifact classification. Motivated by established statistical models of acoustic seafloor backscatter, we replace the standard univariate Gaussian prior on the first convolutional layer with physically motivated alternatives: a multivariate normal (MVN) prior that captures empirical pixel covariance, and asymmetric priors comprising the Gumbel model, which corresponds to a log-transformed Weibull distribution. All priors are implemented via local reparameterization with closed-form Kullback–Leibler (KL) divergence, preserving computational tractability. On a dataset of 54 geographically diverse SAS images comprising 3×10^5 300×300 pixel patches, the Gumbel prior yields 97.6% macro $F1$ and 0.7% expected calibration error (ECE) on ResNet20, versus 87.1% $F1$ and 9.7% ECE for the Gaussian baseline. On ResNet50, Gumbel achieves 99.2% accuracy, surpassing the deterministic model. Therefore, prior specification deserves domain-specific consideration in BNN deployments.**

***Index Terms*—Bayesian deep learning, Bayesian neural network (BNN), image statistics, synthetic aperture sonar (SAS).**

I. Introduction and Motivation

SONAR systems with high resolution are frequently employed in seabed remote sensing and target detection [1], [2]. Synthetic aperture sonar (SAS) is often used as a remote sensing tool for seabed classification [3], [4]. In these applications, features are extracted from regions of seafloor imagery and used with clustering methods [5], [6]. Deep neural networks are known to be highly sensitive to small image perturbations [7], and therefore, downstream tasks such as seabed classification and target detection can be highly sensitive to imaging artifacts.

Poor navigation or wave speed estimates are observed to cause imaging errors or artifacts such as blurring, grating lobes, and low signal-to-noise ratio (SNR) [8]. Therefore, successful remote sensing using SAS requires image quality control metrics, such as image sharpness [9]. Previous work used image perturbations in beamforming to recognize images corrupted by sound speed, yaw error, or poor SNR [10]. This method was largely successful and had well-calibrated entropy (meaning that high entropy was associated with misclassification of the image artifact). However, this model tended to perform poorly on featureless seafloors and at times confused yaw error with good quality images.

A key limitation shared across most Bayesian deep learning deployments, including our previous work, is the use of isotropic Gaussian weight priors. Fortuin [11] has shown that default Gaussian priors may cause a "cold posterior" effect in which the tempered posterior outperforms the true Bayesian posterior, and they ignore structural properties that are available a priori. The motivation for this work is that prior misspecification is not uniformly distributed across network layers. Glorot and Bengio [12] showed that the variance of activations propagates as a product of input variance and layer-wise weight variances, implying that the statistical coupling between data and weights is strongest and most direct at the first layer, and becomes weaker at deeper layers where nonlinearities progressively decorrelate the signal. Atanov et al. [13] showed empirically that structured, domain-informed priors over convolutional filters outperform univariate Gaussians, particularly under limited data. We identify two independent axes along which the univariate Gaussian prior is inadequate for SAS artifact classification.

The first concerns marginal distribution shape. For SAS imagery, the pdf of acoustic backscatter is well-characterized in literature by heavy-tailed, asymmetric distributions such as the Weibull [14], K [15], [16], [17], and log-normal [18]. When log-transformed (as is common in machine learning with sonar and radar input pipelines), a Gaussian model corresponds with a log-normal distribution for the image intensity. This statistical mismatch between the default prior, independent Gaussian, and a more accurate distribution motivates the first modification proposed in this letter. We propose to use the Weibull distribution for the image intensity, which corresponds to a Gumbel extreme value (GEV) model for the log intensity. While other models, such as the K, are often a better fit to data [19] and are physically motivated [15], the extreme value model has a closed-form solution for Kullback–Leibler (KL) divergence. This choice balances accurate modeling of the input layer model with computational tractability.

Received 8 May 2026; revised 2 July 2026; accepted 13 July 2026. Date of publication 20 July 2026; date of current version 3 August 2026. This work was supported by the Office of Naval Research (ONR) under Grant N0001421WX00575. *(Corresponding author: Marko Orescanin.)*

Marko Orescanin, Derek R. Olson, Dalton Duvio, and Haley Jennings are with Naval Postgraduate School, Monterey, CA 93943 USA (e-mail: marko.orescanin@nps.edu).

Marc Geilhufe, Roy E. Hansen, and Narada Warakagoda are with Norwegian Defence Research Establishment, 2027 Kjeller, Norway.

Digital Object Identifier 10.1109/LGRS.2026.3715375

The second modification is to the spatial correlation structure. The independent Gaussian prior imposes a diagonal covariance on convolutional kernel weights, implying that neighboring pixels are statistically independent. This assumption is inconsistent with seafloor texture, which arises from spatially structured acoustic scattering, and pixels in proximity exhibit measurable statistical dependence [20]. This spatial correlation is a consequence of the seafloor environment, and discarding it in the prior represents a second axis of misspecification that may be addressable from the data directly. Employing matched distribution families for prior and posterior enables a closed-form calculation of the KL divergence, preserving the computational tractability of the standard Gaussian case while replacing the distributional assumption with one that is physically motivated. The physically motivated prior is applied exclusively to the first convolutional layer, where weights operate directly on pixel intensity values; thus, the statistical connection between the prior distribution and the input data distribution is direct.

We increased the dataset of perturbed images to approximately 3.4×10^3 patches (compared to 10×10^3 patches used previously), which helps protect against overfitting when using large neural networks. The data were physically perturbed by alterations to the beamforming and standard data augmentations, which resulted in a dataset size of approximately 377 000 patches.

II. Methodology

A. Bayesian Neural Networks

We adopt the variational inference (VI) framework for Bayesian neural networks (BNNs) as described in [21]. A prior $p(\omega)$ is placed over network weights $\omega \in \Omega$, and the intractable posterior is approximated by $q_\theta(\omega) \approx p(\omega|\mathcal{D})$ via minimization of the negative evidence lower bound (ELBO) [21], [22]

$$\mathcal{L}_q = \mathrm{KL}\left(q_\theta(\omega) \;\|\; p(\omega)\right) - \mathbb{E}_q\left[\log p(\mathcal{D} \mid \omega)\right] \tag{1}$$

where the KL term penalizes divergence from the prior. At inference, predictions are formed by Bayesian model averaging over T weight samples, and predictive uncertainty is quantified via normalized entropy [10]

$$H_p(y^* \mid x^*) = -\sum_{c=1}^{C} \bar{p}(y^* = c \mid x^*) \frac{\log \bar{p}(y^* = c \mid x^*)}{\log C}. \tag{2}$$

This work investigates nondiagonal multivariate normal (MVN) and asymmetric nonnormal priors, implemented via local reparameterization [22], selected for its compatibility with non-Gaussian, asymmetric distributions.

B. Nondiagonal MVN Prior

Standard Bayesian convolutional layers assume diagonal covariance in both prior and posterior weight distributions, implying statistical independence among pixels within the receptive field. Here, we replace the diagonal prior with an MVN prior whose covariance is estimated directly from the training data.

The prior is parameterized as $p(\omega) = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is a 16×16 empirical pixel covariance matrix computed by applying a 4×4 kernel to each training image and averaging across the dataset. The 4×4 kernel size was selected to capture spatial correlation at twice the image grid resolution of 2×2 cm. Because $\mathbf{\Sigma}$ is symmetric positive definite, it admits a Cholesky factorization $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^\intercal$, which is used to instantiate a `MultivariateNormalTriL` prior distribution in TensorFlow Probability (TFP), with scale initialized to $\mathbf{L}$ and location initialized to $\mathbf{0}$.

TFP does not provide a registered analytical KL divergence between a nondiagonal `MultivariateNormalTriL` prior and a diagonal Gaussian posterior. We resolve this by representing the diagonal posterior as a degenerate `MultivariateNormalTriL` with zero off-diagonal entries, which is analytically equivalent and enables use of TFP's registered closed-form $\mathrm{KL}(\mathcal{N}_{\mathrm{TriL}}(\gamma) \| \mathcal{N}_{\mathrm{TriL}}(\delta))$ computation, where γ and δ are the parameters of each distribution. This modification was applied to the TFP `Conv2DReparameterization` layers and maintains the full pixel covariance structure of the prior throughout training.

C. Asymmetric Nonnormal Priors

SAS amplitude statistics are well-characterized by asymmetric, heavy-tailed distributions [14], [15], [18], motivating a physically grounded alternative to the Gaussian model. Asymmetric distributions prevent the use of the computationally efficient flipout estimator, since it requires a symmetric pdf around the mean [23]. All experiments, therefore, employ the local reparameterization estimator exclusively [22].

A custom `Conv2DReparameterization` layer was implemented, in which both prior and posterior are set to the same distribution family. This matched design is a deliberate constraint that enables TFP's natively registered, closed-form KL divergence for same-family pairs [24], avoiding Monte Carlo estimation of the KL term and preserving gradient stability throughout training.

Distribution selection was motivated by the requirements of both fitting acoustic data collected in the field and having a closed-form KL divergence. Empirically, we found that the Weibull model fit better than a log-normal model (corresponding to a Gumbel model and normal model for the log intensity). The GEV model is described by

$$f_{GEV}(x; \beta, \alpha) = e^{-\tilde{x} - e^{-\tilde{x}}} \tag{3}$$

where x is the log-scaled intensity and $\tilde{x} = (x - \alpha)/\beta$ is a version of x modified by the location parameter α and scale parameter β. A Weibull distribution for the image intensity is given by

$$f_W(y; \sigma, k) = \begin{cases} \dfrac{k}{\sigma}\left(\dfrac{y}{\sigma}\right)^{k-1} e^{-(y/\sigma)^k}, & y \geq 0 \\ 0, & y < 0 \end{cases} \tag{4}$$

where y is the image intensity, σ is the scale parameters, and k is the shape parameter.

The normal and Gumbel models are compared to a histogram of a single 300×300 pixel patch in Fig. 1. Visually,

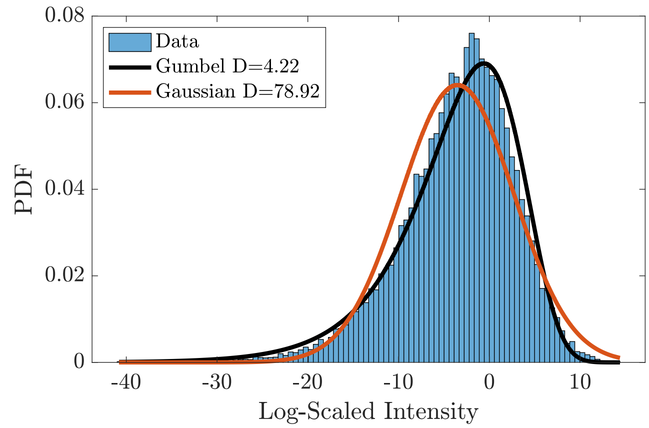

Fig. 1. Distribution of log-compressed pixel intensities (dB) from unperturbed training images, with fit Gumbel and normal distributions. Pixels at the 60 dB dynamic range floor were excluded from fitting. The Gumbel provides a substantially better fit as quantified by the KS test statistic (D = 4.2) than the normal (D = 78.9).

the Gumbel model matches the peak, lower tail, and upper tail better than the Gaussian model. The Kolmogorov–Smirnov (KS) test statistic, D, is used to quantify goodness of fit and is defined as the maximum absolute difference between the model cdf and the empirical data cdf, scaled by $(N_s)^{1/2}$, where $N_s = 9 \times 10^4$ is the number of samples [25]. D is bounded from below by 0 and has no upper limit, small values corresponding to a good match between the model and data, while large values corresponding to a poor match. Raw model-data cdf differences are of magnitude 10^0 or less, and the scaled version can be converted directly to the p-value [25].

The D value (shown in the figure legend) is lower for the Gumbel than for the Gaussian, indicating a better match. For all the 300×300 pixel patches in the dataset, we observed a median KS test statistic for the Gumbel distribution of 0.53, with a 25th to 75th percentile interval of (0.49, 0.56). For the Gaussian model, we observed a median of 35.9 with a percentile interval of (35.5, 36.3), indicating a very poor fit, considering the large number of samples used. These statistics indicate that the Gumbel model is more appropriate than a Gaussian for modeling the input layer prior distribution. In other tests by the authors, the log-transformed K distribution [26] outperformed both Gumbel and Normal (with a median D value of 0.38) but was not chosen due to its lack of analytical tractability in VI and calculation of the KL divergence. The Weibull, Gumbel, and GEV distributions were each evaluated as priors, with all three implemented via TFP's native distribution objects [24]. Using the p-value, the Gaussian model was rejected at the 5% significance level, whereas the Gumbel and log-scaled K models were not rejected.

D. Data

This work builds on our previous dataset [10] by increasing the number of SAS images available for analysis. The data were acquired by the Norwegian Defense Research Establishment (FFI) using the HISAS 1032 interferometric SAS mounted on a HUGIN autonomous underwater vehicle (AUV). Surveys were carried out in five geographically and geologically diverse marine environments: four sites in Norwegian waters and one in the Mediterranean Sea off the Italian coast. These locations were chosen to represent a variety of seafloor types and complex terrain features. The AUV altitude was 15–30 m, the center frequency was 100 kHz with a 30 kHz bandwidth. SAS images were generated using a backprojection algorithm, incorporating sidescan-derived bathymetry to define the render plane, and applying micronavigation techniques for precise motion compensation. The resulting images had a theoretical resolution of approximately 3.5×3.5 cm, with images gridded at 2×2 cm. In our previous study, 28 images were selected for analysis (comprising approximately 10×10^3 patches); in the present work, the dataset contained 54 images (comprising 34×10^3 patches). With the physical perturbations and augmentations, the total dataset size was approximately 377×10^3 patches.

TABLE I

Bayesian Method Results on Expanded Dataset

	ResNet20		ResNet50	
Bayesian Method	**Acc**	**MNLL**	**Acc**	**MNLL**
Determin.	0.975	N/A	0.949	N/A
Reparam.	0.868	0.337	0.851	0.358
Reparam. MVN	0.913	0.233	0.792	0.450
Reparam. Gumb.	0.976	0.067	0.992	0.022
Reparam. Weib.	0.286	1.381	0.328	1.343

III. Results

Table I summarizes accuracy and mean negative log-likelihood (MNLL) for all Bayesian methods and prior combinations on the expanded dataset, evaluated on both ResNet20 (0.12M parameters) and ResNet50 (0.31M) architectures. Our backbones are ResNet V1 variants of depth 20 and 50 with a constant width of 16 filters per stage and are therefore substantially smaller than the CIFAR ResNet20 (0.27M) and the ImageNet ResNet50 (25.6M). All Bayesian variants double the parameter count of a deterministic layer by adding a per-weight scale, so the Gaussian and Gumbel models use 2× the parameters of the deterministic baseline, whereas applying a full-covariance MVN posterior to the first convolutional layer raises that layer's penalty to 9.5× through its 16×16 Cholesky factorization. Gumbel and Gaussian models both use the same number of parameters. Since only the first layer is altered in the two proposed methods, the MVN model results in a very slight increase in the number of parameters compared to the univariate Gaussian. Deterministic baselines achieve 97.5% and 94.9% accuracy for ResNet20 and ResNet50, respectively, with the lower ResNet50 performance suggesting overfitting.

Among the Bayesian variants, the Gumbel prior achieves the strongest performance across both architectures: 97.6% accuracy and 0.067 MNLL on ResNet20, and 99.2% accuracy and 0.022 MNLL on ResNet50. The Gumbel ResNet50 result is particularly notable, as it exceeds the corresponding deterministic baseline by 4.3% points, suggesting that the physically motivated prior provides regularization beyond that of the deterministic model. On the contrary, the standard Gaussian reparameterization prior yields lower accuracy (86.8% and 85.1%) with higher MNLL (0.337 and 0.358). The MVN model performs as shown in

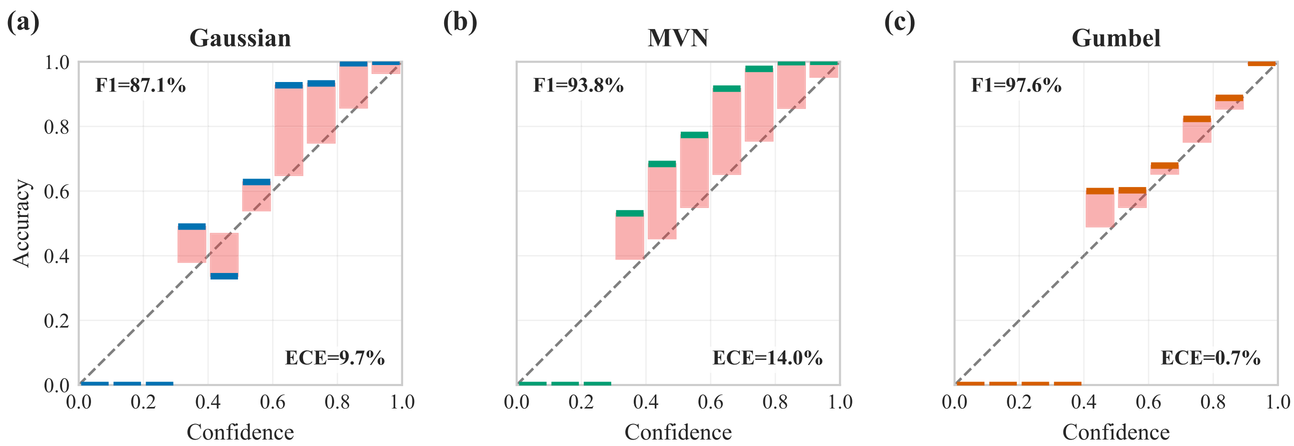

Fig. 2. Reliability diagrams for Bayesian ResNet20 with (a) Gaussian, (b) MVN, and (c) Gumbel priors. Per-bin accuracy is plotted against prediction confidence; the dashed diagonal indicates perfect calibration. Red shading shows calibration gaps. The Gumbel prior achieves ECE = 0.7% with $F1 =$ 97.6%, versus 9.7%/87.1% (Gaussian) and 14.0%/93.8% (MVN).

TABLE II
PER-CLASS PRECISION (P), RECALL (R), AND $F1$-SCORE FOR GAUSSIAN AND GUMBEL PRIORS ON RESNET20

	Gaussian			Gumbel		
Class	P	R	F1	P	R	F1
No Artifact	1.000	0.950	0.975	1.000	1.000	1.000
Sound Speed	0.989	0.662	0.793	0.992	0.911	0.950
SNR	0.996	0.863	0.925	1.000	1.000	1.000
Yaw	0.658	0.996	0.792	0.918	0.993	0.954
Macro Avg	0.911	0.868	0.871	0.977	0.976	0.976

Table II, which reveals the per-class source of the Gaussian prior's underperformance. The Gaussian model exhibits confusion between sound speed and Yaw artifacts: Yaw recall is high (0.996), but precision is low (0.658), while sound speed shows the inverse pattern (precision 0.989, recall 0.662). Physically, both artifacts manifest as phase errors in beamforming, and the symmetric Gaussian prior apparently provides insufficient distributional flexibility to resolve these signatures. The Gumbel prior largely eliminates this confusion, raising sound speed recall from 0.662 to 0.911 and Yaw precision from 0.658 to 0.918 while achieving perfect classification ($F1 =$ 1.000) on both no artifact and SNR classes.

The MVN prior improves over the diagonal Gaussian on ResNet20 (91.3% accuracy, 0.233 MNLL) but degrades on ResNet50 (79.2%, 0.450 MNLL). This poor performance for larger models suggests overfitting behavior for the MVN. This has been verified by examining the training curves, in which the MVN validation accuracy peaks early and then overfits, while the Gumbel validation accuracy increases more slowly with training epoch.

As a counterexample, we used the Weibull prior on log-scaled data (an intentional mismatch), and it performed poorly on all metrics, achieving near-chance accuracy (~30%). Because the network operates on log-scaled intensity data, the Weibull distribution, which is appropriate for raw intensity, is fundamentally mismatched with the network input. The Gumbel distribution succeeds because it corresponds to a log-transformed Weibull, making it the correct physically motivated prior. This result underscores that prior selection must account for both the underlying physics *and* the data

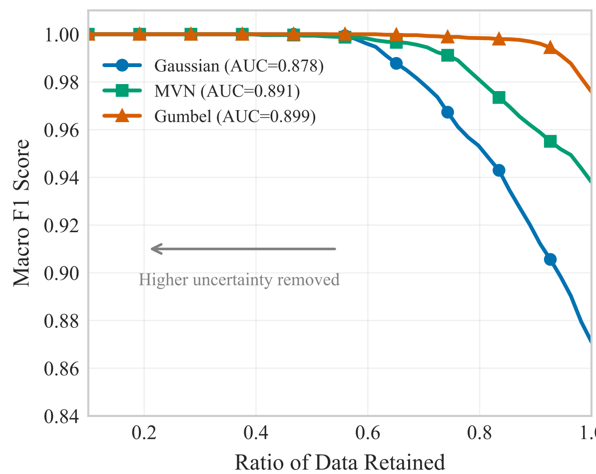

Fig. 3. Macro $F1$ versus ratio of test data retained after entropy-based filtering for three Bayesian ResNet20 models. The Gumbel prior maintains high $F1$ at higher retention ratios (AUC = 0.899) compared to MVN (0.891) and Gaussian (0.878).

preprocessing pipeline. We have previously attempted to train a model using the linear intensity and a Weibull model, but it failed to achieve satisfactory results because of the large dynamic range of the image intensity. This is the primary reason for log-scaling data as part of the processing pipeline.

Fig. 2 presents reliability diagrams for the three viable Bayesian ResNet20 models. The Gumbel prior achieves near-perfect calibration with an expected calibration error (ECE) of just 0.7%, compared to 9.7% for the Gaussian and 14.0% for the MVN prior. The Gumbel model concentrates its predictions in high-confidence bins while maintaining accuracy that closely tracks confidence, indicating that the model is confidently correct on the vast majority of samples. The Gaussian and MVN priors exhibit substantial calibration gaps across multiple confidence levels, with predictions dispersed more broadly.

Fig. 3 shows macro $F1$ score as a function of data retained after entropy-based filtering. All three priors have near-perfect classification when only the most confident predictions are retained, confirming that predictive entropy is a meaningful uncertainty proxy. However, the Gumbel prior maintains high $F1$ at substantially higher retention ratios. At full retention, it achieves 0.976 $F1$ versus 0.871 for Gaussian. The area under the retention curve (AUC) summarizes this advantage: Gumbel (0.899) outperforms MVN (0.891) and Gaussian (0.878). The Gumbel prior produces fewer uncertain predictions overall.

IV. Discussion and Conclusion

These results demonstrate that prior selection in BNNs is a consequential design choice for SAS artifact classification. The Gumbel prior, physically motivated by the literature on backscatter intensity statistics [15], achieves near-perfect calibration and outperforms the standard Gaussian baseline by over 10% points in $F1$ on ResNet20. The failure of Weibull prior highlights that the prior must match the data representation (log-scaling), not merely the underlying physical process.

The MVN prior's inconsistent performance across architectures suggests that capturing spatial correlation through the prior covariance is promising but requires further investigation into architecture-dependent interactions. Future work could explore learned covariance structures or use other random process models.

A limitation of this study is the use of a single sensor system (HISAS 1032) across 54 images in two geographic regions. Validation on additional SAS platforms and other geographic areas would further establish the approach. Because physically motivated priors impose domain structure on the weights, they act as implicit regularizers that lessen the reliance on large training sets; this is reflected in the smaller generalization gap and lower ECE of the Gumbel prior relative to the Gaussian baseline and is consistent with the demonstrated advantage of domain-informed priors under limited data [13]. Nonetheless, our method is likely specific to the geographic regions and sensors with which the model was trained. However, transfer learning has been demonstrated as a viable method to use a model trained on a different sensor or different targets [27]. Additional training data that fully captures image textures and sensor characteristics, coupled with a model with larger capacity, may result in more generalizable model results.

In summary, this letter establishes that physically motivated priors, specifically the Gumbel distribution applied to the first convolutional layer, substantially improve both accuracy and calibration for Bayesian SAS artifact classification. The approach preserves computational tractability through closed-form KL divergence and requires modification to only a single network layer, making it straightforward to adopt in existing Bayesian deep learning pipelines.

References

[1] C. Heinrich, P. Feldens, and K. Schwarzer, "Highly dynamic biological seabed alterations revealed by side scan sonar tracking of Lanice conchilega beds offshore the island of Sylt (German Bight)," *Geo-Marine Lett.*, vol. 37, no. 3, pp. 289–303, Jun. 2017.

[2] D. P. Williams, "Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 2497–2502.

[3] J. T. Cobb and J. Principe, "Autocorrelation features for synthetic aperture sonar image seabed segmentation," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2011, pp. 3341–3346.

[4] B. Gips, "Texture-based seafloor characterization using Gaussian process classification," *IEEE J. Ocean. Eng.*, vol. 47, no. 4, pp. 1058–1068, Oct. 2022.

[5] M.-B. Naghi, L. Kovacs, and L. Szilagyi, "A review on advanced C-means clustering models based on fuzzy logic," in *Proc. IEEE 21st World Symp. Appl. Mach. Intell. Informat. (SAMI)*, Jan. 2023, pp. 000293–000298.

[6] M. Perkonigg, D. Sobotka, A. Ba-Ssalamah, and G. Langs, "Unsupervised deep clustering for predictive texture pattern discovery in medical images," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2020, pp. 1–11.

[7] V. Veerabadran et al., "Subtle adversarial image manipulations influence both human and machine perception," *Nature Commun.*, vol. 14, 2023, Art. no. 4933, doi: 10.1038/s41467-023-40499-0.

[8] D. A. Cook and D. C. Brown, "Analysis of phase error effects on stripmap SAS," *IEEE J. Ocean. Eng.*, vol. 34, no. 3, pp. 250–261, Jul. 2009.

[9] M. Geilhufe, D. R. Olson, R. E. Hansen, and S. A. V. Synnes, "A wavelet shrinkage approach to detect candidate point scatterers in synthetic aperture sonar images for resolution estimation," in *Proc. 5th Int. Conf. Synth. Aperture Sonar Synth. Aperture Radar (SAS/SAR)* (Proceedings of the Institute of Acoustics), vol. 45, G. Heald, Ed., Lerici, Italy,, Sep. 2023.

[10] M. Orescanin et al., "Classification of imaging artifacts in synthetic aperture sonar with Bayesian deep learning," *IEEE J. Ocean. Eng.*, vol. 50, no. 3, pp. 2280–2295, Jul. 2025.

[11] V. Fortuin, "Priors in Bayesian deep learning: A review," *Int. Stat. Rev.*, vol. 90, no. 3, pp. 563–591, Dec. 2022.

[12] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.

[13] A. Atanov, A. Ashukha, K. Struminsky, D. Vetrov, and M. Welling, "The deep weight prior," in *Proc. Int. Conf. Learn. Represent*., 2019. [Online]. Available: https://openreview.net/forum?id=ByGuynAct7

[14] M. Gensane, "A statistical study of acoustic signals backscattered from the sea bottom," *IEEE J. Ocean. Eng.*, vol. 14, no. 1, pp. 84–93, Jan. 1989.

[15] A. P. Lyons and D. A. Abraham, "Statistical characterization of high-frequency shallow-water seafloor backscatter," *J. Acoust. Soc. Amer.*, vol. 106, no. 3, pp. 1307–1315, Sep. 1999.

[16] D. R. Olson, A. P. Lyons, D. A. Abraham, and T. O. Sæbø, "Scattering statistics of rock outcrops: Model-data comparisons and Bayesian inference using mixture distributions," *J. Acoust. Soc. Amer.*, vol. 145, no. 2, pp. 761–774, Feb. 2019.

[17] D. R. Olson and M. Geilhufe, "Model selection techniques for seafloor scattering statistics in synthetic aperture sonar images of complex seafloors," *IET Radar, Sonar Navigat.*, vol. 18, no. 11, pp. 2044–2056, Jun. 2024.

[18] N. P. Chotiros et al., "Acoustic backscattering at low grazing angles from the ocean bottom. Part II. Statistical characteristics of bottom backscatter at a shallow water site," *J. Acoust. Soc. Amer.*, vol. 77, no. 3, pp. 975–982, Mar. 1985.

[19] A. P. Lyons, D. R. Olson, and R. E. Hansen, "Modeling the effect of random roughness on synthetic aperture sonar image statistics," *J. Acoust. Soc. Amer.*, vol. 152, no. 3, pp. 1363–1374, Sep. 2022.

[20] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images (Scitech Radar and Defense Series)*. Rijeka, Croatia: SciTech, 2004. [Online]. Available: https://books.google.com/books?id=IeGKe40S77AC

[21] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *J. Amer. Stat. Assoc.*, vol. 112, no. 518, pp. 859–877, Apr. 2017.

[22] D. P. Kingma, T. Salimans, and M. Welling, "Variational dropout and the local reparameterization trick," in *Proc. 29th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 2. Cambridge, MA, USA: MIT Press, 2015, pp. 2575–2583.

[23] Y. Wen, P. Vicol, J. Ba, D. Tran, and R. Grosse, "Flipout: Efficient pseudo-independent weight perturbations on mini-batches," in *Proc. Int. Conf. Learn. Represent.*, 2018. [Online]. Available: https://openreview.net/forum?id=rJNpifWAb

[24] J. V. Dillon et al., "TensorFlow distributions," 2017, *arXiv:1711.10604*.

[25] A. Papoulis and S. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed., New York, NY, USA: McGraw-Hill, 2002.

[26] K. D. Ward, "Compound representation of high resolution sea clutter," *Electron. Lett.*, vol. 17, no. 16, pp. 561–563, Aug. 1981.

[27] D. P. Williams, "Transfer learning with SAS-image convolutional neural networks for improved underwater target classification," in *Proc. IGARSS - IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 78–81.