Data Preprocessing for Sub-Bottom Profiler-Based Object Detection

Original Scientific Paper

Bruno Zorić
J. J. Strossmayer University of Osijek,
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, Kneza Trpimira 2B
bruno.zoric@ferit.hr

Mario Dudjak*
J. J. Strossmayer University of Osijek,
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, Kneza Trpimira 2B
mario.dudjak@ferit.hr

Filip Garmaz
J. J. Strossmayer University of Osijek,
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, Kneza Trpimira 2B
filip.garmaz@ferit.hr

Dražen Bajer
J. J. Strossmayer University of Osijek,
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, Kneza Trpimira 2B
drazen.bajer@ferit.hr

*Corresponding author

Abstract *– Hydrographic research projects are often described as costly, data-intensive, high-stakes operations. This holds especially true for sub-bottom object detection where decisions regarding route planning, dredging or hazard assessment are often reliant upon sub-bottom profiler (SBP) data. Today, this data is commonly processed and interpreted manually. Whilst deep learning (DL)-based object detectors are increasingly being utilised to automate processing and interpretation, preparation of raw SEG-Y data for optimal exploitation by various models remains underexplored. Because proprietary subsurface datasets are scarce, to say that training DL models from scratch is impractical would be an understatement. Consequently, researchers must rely on transfer learning and models pre-trained on large datasets of out-of-domain general digital images. This reliance underscores the critical importance of the transformation techniques used to translate raw acoustic-response data to digital image pixels. This paper explores and evaluates several SEG-Y-to-digital image transformation techniques, i.e., scaling or normalisation techniques on real-world data in conjunction with two well-known object detection frameworks. Image normalisation is positioned not as just another preprocessing step, but as a critical experimental component of the SBP detection pipeline. Evaluations were conducted on a single-system SBP dataset, representing a specific acoustic frequency range and environmental boundary condition. It was demonstrated that contrast enhancement and outlier clipping increase downstream model convergence, yielding a more robust target recovery performance despite severe background clutter.*

Keywords: *buried object detection, deep learning, hyperbolic shapes, image preprocessing, sub-bottom profiler*

Received: July 12, 2026; Received in revised form: October 16, 2026; Accepted: October 16, 2026

1. INTRODUCTION

Sub-bottom profilers are a standard tool in marine geophysics for imaging the shallow subsurface beneath the seabed. They transmit acoustic pulses and record reflected energy along consecutive traces, producing profiles that reveal sediment layers, buried interfaces, and local anomalies that are not directly visible through bathymetric or optical observations [1, 2]. This information is essential when engineering or environmental decisions depend on sub-seabed conditions. Typical applications include offshore route planning, dredging, archaeological inspection, defence surveys, and the assessment of sub-seabed hazards [3, 4]. In these settings, the practical goal is not only to visualise the seabed, but also locate objects or hazardous structures whose acoustic response may be weak, partly masked, or distorted due to acquisition difficulties.

Object detection from SBP data is challenging because a sub-bottom profile is an acoustic record rather than a conventional image. Each trace contains amplitude samples indexed by travel time. Buried infrastructure, natural objects, and sediment anomalies appear in the data through the acoustic responses they produce. These responses may take the form of hyperbolic reflec-

tions, disrupted sediment horizons, or local changes in acoustic intensity [2–4]. Their visibility is also affected by acquisition conditions and post-acquisition processing [1,5,6]. Manual interpretation, therefore, remains slow and dependent on expert judgement, particularly for long survey lines and repeated monitoring campaigns.

Deep learning has increasingly been used for automated subsurface interpretation. In ground penetrating radar (GPR), convolutional detectors have been applied to B-scan images for tasks such as buried object detection, archaeological interpretation, road defect diagnosis, and rebar estimation [7–10]. This work is relevant to SBP since both modalities can contain hyperbolic or arc-like responses from compact targets. SBP-specific studies are few, but recent work has addressed the automatic detection of buried shipwrecks in 3-D SBP data and shallow mud features in submarine channels [3,4]. These results suggest that learning-based detection can support marine inspection workflows. At the same time, they point to a practical issue addressed in this paper, i.e., performance depends not only on the detector, but also on how acoustic amplitudes are represented as image intensities.

Image normalisation therefore represents an important part of the SBP image representation. Although often treated as a minor preprocessing step, it can affect which acoustic patterns remain visible to the detector. When SBP data are used as image-like inputs, their amplitudes must be transformed to a suitable intensity range [11,12]. This transformation influences the contrast between object-related reflections and the surrounding sediment structure. Global normalisation may preserve profile level amplitude differences, but can be dominated by strong reflections. Trace-wise normalisation can produce a more balanced image, whilst suppressing amplitude trends that may carry physical information. Contrast enhancement can improve the visibility of weak responses, but may also amplify noise. Similar effects have been noted in GPR visualisation, where normalisation and contrast mapping influence interpretation and may introduce observer bias [11].

Despite this importance, image normalisation is rarely evaluated as an explicit design choice in SBP object detection pipelines. Most existing studies apply generic display settings without considering how amplitude compression reshapes subtle acoustic signatures. This gap is especially problematic given the scarcity of annotated marine SBP benchmarks compared to mature computer vision tasks or broader GPR literature [3,4,7,8]. Rather than introducing another incremental network architecture, this paper presents a data-centric paradigm shift: a systematic empirical taxonomy and evaluation framework that uncovers how raw SEG-Y transformation mechanics influence feature preservation in deep neural networks. Faster R-CNN [13] and YOLO [14] are employed to test whether the observed effects are tied to a particular type of detector or remain present across different detection architectures. In this way, the paper treats image normalisation as an experimental component of the SBP detection pipeline, a functional hyperparameter of sorts, rather than as a fixed display setting establishing the first standardised evaluation framework mapping SEG-Y dynamic range compression to acoustic hyperbola retention. The core novelty of this work lies in the first systematic, unified empirical evaluation of how raw SEG-Y amplitude transformations directly dictate the downstream feature extraction capabilities of structurally distinct deep learning architectures. By shifting the focus from architectural tuning to data-centric preprocessing optimisation, relationships between amplitude normalisation techniques and model-specific detection sensitivities in real marine environments are uncovered. The core contributions and empirical findings of this work are the systematic evaluation of 13 transformation techniques on real-world SBP data, demonstrating that global percentile clipping and trace-based scaling provide good feature distribution for generic vision models, bridging the gap between hyperbola enhancement and background noise suppression. The empirical findings are translated into practical utility, since scenario-specific recommendations mapping specific preprocessing pipelines to operational guidelines are formulated based on them.

The rest of the paper is organised as follows. Section 2 reviews object detection approaches in seismic imaging, with emphasis on methods relevant to SBP data. It also provides an overview of commonly employed scaling-related preprocessing approaches. The conducted experimental evaluation is described in Section 3, whilst Section 4 summarises both the key observations and avenues for future work.

2. OBJECT DETECTION AND SEISMIC DATA PREPROCESSING

2.1. OBJECT DETECTION IN SEISMIC IMAGING

Object detection in seismic imaging differs from detection in natural images because targets rarely have closed visual boundaries. Instead, they are identified from the responses produced by wave propagation. In 2-D SBP data, compact buried targets often appear as hyperbolic or arc-shaped responses (or simply hyperbolae, in general). In sediment-related inspection, the relevant cue may be a local intensity change or a disturbance in otherwise continuous reflectors [2, 3]. Detection in such a setting requires, therefore, both physical interpretation and pattern recognition. Generally, the task difficulty depends on signal quality, background clutter, and target geometry [2–4].

Classical methods address subsurface detection by exploiting the geometry of the recorded responses. Hyperbola detection in GPR and seismic profiles has commonly relied on geometric or energy-based rules, such as template matching, curve fitting, and migration-based analysis [12,15]. Furthermore, techniques based on the Ra-

don and Hough transforms have been widely employed to map spatial hyperbolic signatures into localized focal points within a transformed parameter space, facilitating threshold-based target extraction [16–18]. These rules are useful as they encode physical prior knowledge and can work with limited labelled data. However, they depend on assumptions about the medium, reflector geometry, and background structure. Field data often depart from these assumptions, as reflections can overlap, sediment properties can vary laterally, and weak targets may be partly hidden by clutter or striping artifacts. In such complex environments, traditional mathematical detectors often become brittle, requiring extensive manual parameter tuning by expert geophysicists. This limitation has motivated the increasing use of DL-based detection. Whilst generic deep learning models are designed for optical images, they overcome classical limitations by extracting robust, hierarchical spatial features capable of identifying deformed or partial hyperbolic signatures. However, successfully adapting these architectures to acoustic data relies heavily on the quality of the input feature distribution, underscoring the critical need for optimized preprocessing pipelines.

Deep learning-based object detectors reduce the reliance on handcrafted rules by learning features directly from annotated samples. An early adaptation of Faster R-CNN to GPR B-scans showed that two-stage detectors can localise buried object signatures when real training data are complemented with simulated radargrams [7]. This is relevant to SBP since labelled samples are difficult to obtain and target appearance varies with acquisition conditions. Faster R-CNN first generates candidate regions and then refines object classification and bounding-box localisation [13]. This structure can be useful when targets occupy a small part of the profile or appear in cluttered backgrounds. Such conditions are common in SBP and GPR data, where object related reflections may be narrow, weak, or incomplete. The drawback is slower inference since candidate generation and refinement are performed sequentially. This trade-off may be acceptable for careful offline interpretation, but becomes restrictive when large survey archives have to be screened.

Single-stage detectors take the opposite approach. YOLO predicts bounding boxes and class probabilities directly from a grid-based representation of the input image [14]. Its speed has made the YOLO family attractive for inspection workflows that involve large numbers of profiles. As mentioned, YOLO-based methods found broad use in GPR applications [9,10,19]. These studies indicate that single-stage detectors can capture hyperbolic and locally structured radar responses. They also motivate closer attention to the input representation, alongside detector architecture and training strategy. RetinaNet addresses a related challenge by using focal loss to reduce the influence of numerous easy background samples [20]. This is relevant when object responses occupy only a small part of the profile. In archaeological GPR data, learned detectors have been shown to provide useful hyperbola candidates, although additional fitting or interpretation may still be needed to estimate target properties [8].

SBP-specific studies are still more target dependent than the related GPR literature. Buried shipwreck detection in 3-D SBP data has been addressed using geometric reflection contours and prior hull shape information, rather than by generic image detectors [3]. Shallow submarine mud detection has been studied with an improved YOLOv5s model applied to SBP-derived acoustic images, showing that DL-based detectors can be useful in operational marine channel inspection [4]. These studies indicate that SBP data can support automated detection, but they also highlight an important practical aspect—bounding boxes may be sufficient for a quick survey, whereas engineering or geological interpretation often requires closer analysis of the detected responses. This reinforces the importance of how acoustic responses are represented as image intensities before applying various DL-based detection models.

2.2. PREPROCESSING ACOUSTIC RESPONSES THROUGH SCALING AND CONTRAST MANIPULATION TECHNIQUES

A key aspect of learning in deep object detection models is gradient-based weight adjustment. When processing SBP data, the raw acoustic response values often span several orders of magnitude, which can lead to volatile gradients, sub-optimal convergence or model instability. To alleviate the aforementioned issues, data needs to be adequately scaled. Furthermore, annotated SBP data is scarce, to put it mildly, so concepts such as pre-training and knowledge transfer become essential, as has been shown for [21] GPR data. However, using models pre-trained on large-scale out-of-domain dataset is typically possible only if the acoustic responses produced by the SBP are converted into digital images, as the models are often optimised for 8-bit digital images. Bridging this gap warrants a transformation which would preserve geophysically significant features, whilst conforming to the characteristics of common computer vision benchmarks. The goal is then to find such a transformation or scaling technique amongst the available choices, not merely to serve as a preprocessing step, but as a critical factor in ensuring the model's ability to generalise across diverse sub-bottom environments.

The most straightforward technique for mapping acoustic intensities into the unit interval [0,1] is global min-max (GMM) scaling. Each sample $x_{i,j}$, where the index i denotes the sample and j the trace, respectively, is linearly transformed by [22]

$$x^s_{i,j} = \frac{x_{i,j} - x_{\min}}{x_{\max} - x_{\min}},$$

where $x_{\max}$ and $x_{\min}$ minimum and maximum sample values contained in the entire seismic image. Due to its simplicity, it is easy to implement and use, but may be

sub-optimal for SBP data characterised by high dynamic ranges. High amplitudes, generated by, e.g. the seabed reflector, cause a significant loss of visual contrast by effectively squashing lower amplitude stratigraphic details towards zero. In an attempt to combat this, percentile clipping (PC) is often implemented, in which the n-th and $(100 - n)$-th percentiles are determined, denoted as P_{min} and P_{max} and used as thresholds [23]. Samples exceeding them are clipped to the boundary values, after which standard min-max scaling is performed. To effectively apply this technique, an appropriate choice for the percentiles must be made. Despite the improvement offered by clipping, global scaling techniques often cannot account for lateral variations in signal attenuation or changes in reflectivity across a survey. The use of global minimum and maximum amplitude-values is therefore sometimes discouraged in favour of using local, trace-based ones. The minimum and maximum values are computed for each column of data and samples are transformed as

$$x^s_{i,j} = \frac{x_{i,j} - x_{\min,j}}{x_{\max,j} - x_{\min,j}} ,$$

where $x_{\min,j}$ and $x_{\max,j}$ represent j-th trace minimum and maximum values. Beyond the normalisation techniques, standardisation techniques are also commonly employed. Z-score standardisation, where samples are transformed by subtracting the mean and dividing by the standard deviation so they have zero mean and unit standard deviation (σ) represents one such technique [24]. Clipping is incorporated as an additional step, utilising the number of allowed standard deviations as a cut-off threshold, where 2σ or 3σ represent sensible choices. A common alternative to peak-based and distribution-based transformation techniques is root mean square (RMS) normalisation [25], which scales the dataset based on the total power of the signal. The RMS value represents the quadratic mean of the amplitudes and serves as a measure of the average energy level. For a given trace or window of N samples, the RMS value is calculated as

$$x_{RMS} = \sqrt{(\frac{1}{N}\sum_{i=1}^{N} x_i^2)} .$$

The normalised samples are then simply computed as

$$x^s_i = x_i / x_{RMS} .$$

This makes it significantly more robust to isolated noise. When applied on a trace-by-trace basis, it acts as a gain-balancing tool that equalises the energy levels across the survey, ensuring that traces with weaker penetration are amplified to be comparable with those showing stronger returns. As linear scaling and statistical normalisation adjust the range and centre of the data, they do not alter the relative relationship between high and low amplitudes which can span several orders of magnitude. To bridge this gap, non-linear mapping functions can be used to compress or expand specific portions of the dynamic range. Logarithmic and exponential transformations represent viable choices in this regard [11].

Although the previously discussed techniques effectively stabilise the amplitude range across the dataset, they remain fundamentally "point-wise" or "trace-wise" operations, lacking awareness of the local spatial context within the 2-D seismic section. In sub-bottom profiling, this often results in an effect where deeper stratigraphic features remain suppressed relative to the high-energy seabed, even after normalisation. This warrants a move towards local spatial transforms, amongst which local contrast normalisation (LCN) [26] and contrast limited adaptive histogram equalisation (CLAHE) [27] stand out as prominent candidates. LCN is a non-linear scaling technique that operates on a local neighbourhood. For each sample the local mean ($\mu_{i,j}$) and standard deviation ($\sigma_{i,j}$) are calculated within a sliding window. The normalised samples are then computed as

$$x^s_{i,j} = \frac{x_{i,j} - \mu_{i,j}}{\max\{\epsilon, \sigma_{i,j}\}},$$

where ϵ represents a small constant preventing division by zero. CLAHE is essentially a histogram equalisation method operating on small non-overlapping regions called tiles. The histogram of each tile is computed and a transformation function is derived to redistribute the intensity values. To prevent the over-amplification of noise it introduces a clip-limit, redistributing excess values across other bins before performing equalisation.

As all of the described transformation techniques operate under different assumptions and provide different visual outputs, there is no clear choice to be made. This is corroborated by a visual comparison of the considered techniques, provided in Fig. 1, which illustrates distinct trade-offs between target hyperbola enhancement and background noise amplification. As can be seen in sub-figure (a), Global Min-Max scaling the buried hyperbolae remain visually subdued due to the dominance of the high-amplitude seabed reflection and the figure suffers from severe low contrast issues, effectively rendering hyperbolae indistinguishable from the background. The percentile based clipping approaches shown in sub-figures (b), (c) and (d) stretch the dynamic range, forcing faint targets across the visual threshold. This comes at the cost of heavily amplifying high-frequency artifacts. Sub-figure (h) highlights the boundary challenges of Local Contrast Normalisation (LCN), exhibiting a distinct noise artifact in the upper water-column region above the seafloor. This is a common consequence of local spatial filters processing low-variance acoustic regions. Z-score normalisation, shown in sub-figure (f), offers only marginal contrast improvements over Global Min-Max. Finally, techniques that achieve a balanced dynamic range, such as RMS in sub-figure (e), Per-Trace scaling in sub-figure (g), and CLAHE in sub-figure (i), appear to provide the most sensible representation. It is clear that a comparative analysis of different transformation techniques, as a preprocessing step in a DL-based object detection pipeline, is by all accounts a worthwhile endeavour.

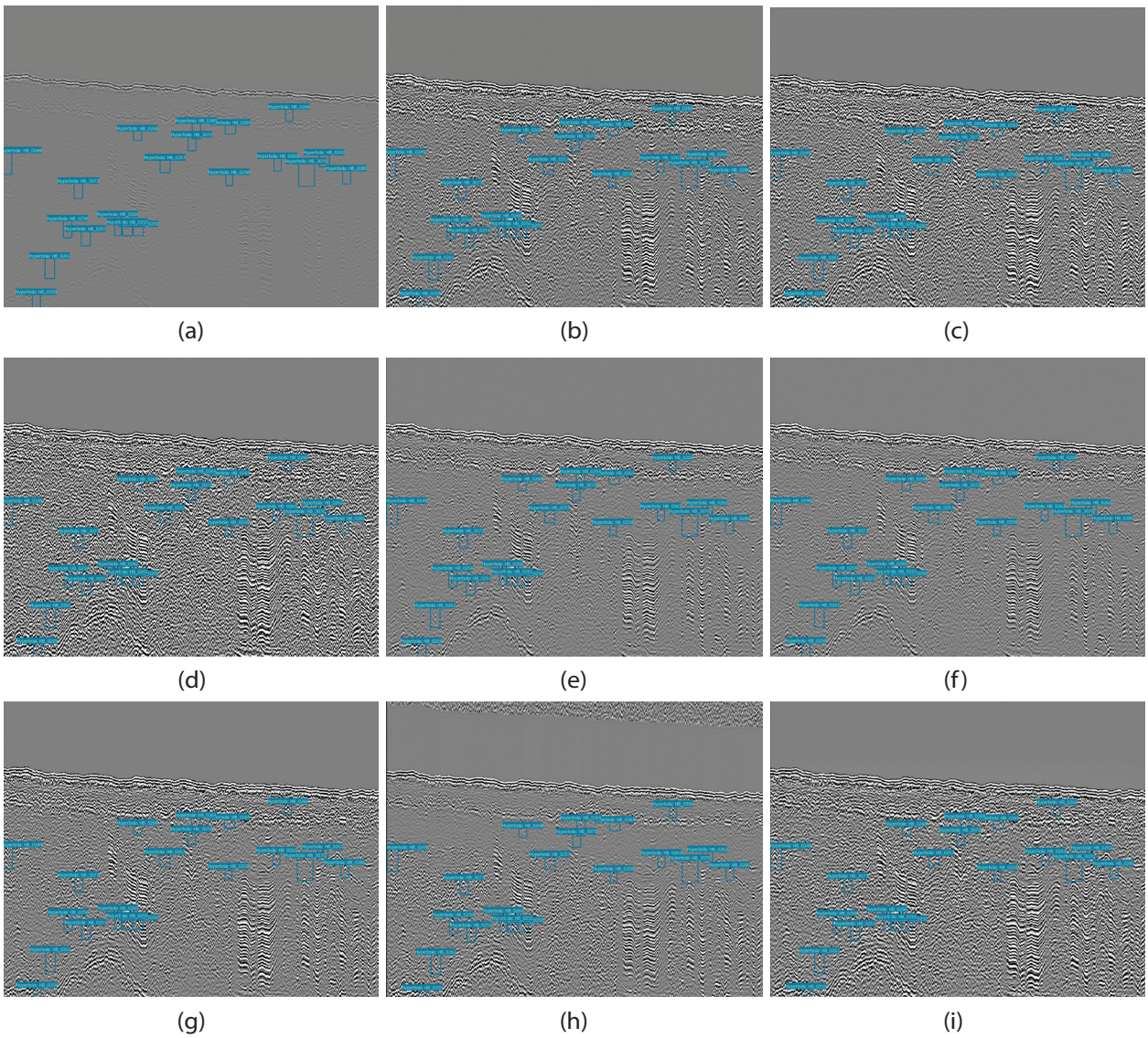


Fig. 1. Examples of transformation techniques on a SEG-Y file patch: (**a**) global min-max, (**b**, **c**, **d**) percentile-based clipping (1-99, 2-98, and 5-95, respectively), (**e**) RMS, (**f**) Z-score, (**g**) per-trace min-max, (**h**) LCN, and (**i**) CLAHE

3. EXPERIMENTAL EVALUATION OF DIFFERENT TRANSFORMATION TECHNIQUES

3.1. EXPERIMENTAL METHODOLOGY AND SETUP

For the purpose of training and evaluating the performance of multiple detector models, real marine data was acquired using an Innomar SES-2000 compact and standard parametric SBPs, which operate at the 2kHz–22kHz low frequency band with 96kHz sampling-rate and 24-bit sample resolution [28]. The dataset was collected on several survey sites in the North sea, with water depths spanning the 10-30m range. The seabed substrate consists predominately of fine to coarse sand and gravelly sand. Typical targets consist of shallowly buried hyperbolic features (e.g., boulders, pipelines, and infrastructure) with dimensions spanning typical SBP acoustic resolutions. The collected SBP data was preprocessed and manually annotated by experts from Hidrocibalae d.o.o. [29]. Preprocessing was performed using standard seismic preprocessing algorithms, including band-pass filtering with 5kHz–11kHz cut-off frequency range at 8kHz central frequency, and automatic gain control (AGC) [30, 31], and stored in the SEG-Y file format. It is important to note that whilst AGC applies a sliding trace-wise scaling to equalize depth-dependent attenuation, it operates on raw 32-bit floating-point amplitude data before any dynamic range conversion to 8-bit image matrices. Each SEG-Y file contains multiple traces of a single survey pass. All traces within a single SEG-Y file have the same number of samples and are stacked, thereby creating a 2-D scan of sub-seabed terrain. These 2-D amplitude matrices were subsequently normalised to a $[0,255] \subset Z$ grayscale range and exported as BMP images to be compatible with standard computer vision workflows. The positions of buried objects, i.e., hyperbolae were marked as points representing apices using the original SEG-Y files by an experienced annotator and then reviewed by another expert annotator. The rectangular bounding boxes were created using the LabelMe annotation tool [32] where the annotator utilised the

previously annotated apex coordinates and tagged the physical bounding-box boundaries (limb extent) based on phase continuity and geometric shape. The final output was a dataset consisting of 141 SEG-Y files, with 9348 labelled hyperbolae. The dataset was partitioned into nine groups, denoted blocks, each one representing a single geological site containing geographically different seismic features influencing the shapes and sizes of buried objects.

As seismic images contain noise, and with no ideal seismic images for comparison to obtain true signal-to-noise ratio (SNR), the effectiveness of the transformation techniques is assessed indirectly, by evaluating the performance of the downstream hyperbolic object detection models. Detection metrics can provide a good insight in DL model performance, whilst DL model performance in turn verifies preprocessing effectiveness. Well-established computer vision performance metrics for object detection are utilised for evaluation purposes, specifically the number of true positive (TP) detections, number of false positive (FP) detections, the number of false negative (FN) detections, precision given as

$$precision = TP/(TP+FP),$$

recall given as

$$recall = TP/(TP+FN),$$

F1- and F2-score (F1 and F2) given as

$$F_\beta = (1+\beta^2)\cdot\frac{precision\cdot recall}{(\beta^2\cdot precision)+recall},$$

where β=1 for F1 and β=2 for F2 respectively, and the mean average precision (mAP). Two details should be elaborated upon. First, the F2-score, although not so common in the literature, was utilised due to the emphasis it places on recall and its relevance to human-in-the-loop workflows. Second, two most common mAP variants were employed for assessing object detection model performance internally, mAP@50 and mAP@50-95. The mAP@50 corresponds to the average AP when the intersection over union (IoU) threshold is set to 0.5 and was selected and utilised for final evaluation due to simple interpretation. The mAP@50-95 is the average mAP across multiple IoU thresholds, "50-95" indicating the use of thresholds in [0.5,0.95] with 0.05 (or 5%) increments. In this paper, the established MS COCO 101-point interpolation is used to approximate the Precision-Recall (PR) curve for efficient mAP calculation with a single class of interest. Accordingly, mAP computation boils down to

$$mAP = \frac{1}{101}\sum_{r\in\{0,0.01,\ldots,0.99,1\}}\max_{\hat{r}\geq r} p(\hat{r}),$$

where $p(\hat{r})$ is the precision measured at recall $\hat{r}$.

3.2. EXPERIMENT DESIGN

To compare the transformation techniques, an experimental framework was designed, a high-level outline of which is given in Fig. 2. SEG-Y files were firstly preprocessed and converted to digital images by employing the different transformation techniques. This resulted in preprocessed high-resolution seismic images, ranging from a thousand to tens of thousands of pixels in both, width and height. Upon statistical inspection of different targets, i.e., hyperbola bounding boxes, the bounding boxes of the smallest hyperbolae were found to be less than 20 pixels in both width and height, whilst the size of the largest hyperbolae were not greater than 100 pixels. Most hyperbola sizes were around 32 pixels, and were mostly scattered around, with background clutter dominating the overall image. This poses a significant challenge to object detection models, as the object of interest occupies a minuscule portion of the image. Furthermore, the seismic images came in different resolutions, and DL-based object detectors combat this by resizing images to a common resolution, usually significantly smaller than the seismic image. This raises another issue, as aggressive seismic image resizing would result in hyperbolic features to fade into dominating background clutter. To solve the aforementioned issues, offline slicing aided hyper inference (SAHI) is partly utilised (patch extraction) to break a large seismic image into patches of the same size [33]. To balance hyperbola size whilst still leaving enough context for the model to effectively distinguish the target from background, a patch resolution of 640×640 was used. During SAHI slicing, not every patch contained full bounding box of hyperbolae, and most of patches contained background clutter. These patches were discarded, leaving only ones which fully captured one or more hyperbolae with labelled bounding box coordinates local to that patch. Finally, the saved dataset contained altogether 6704 input patches, preserving a total of 9344 hyperbolae, as some were discarded due to being on the patch boundaries. This was repeated for each transformation technique, keeping the sliced regions the same for all.

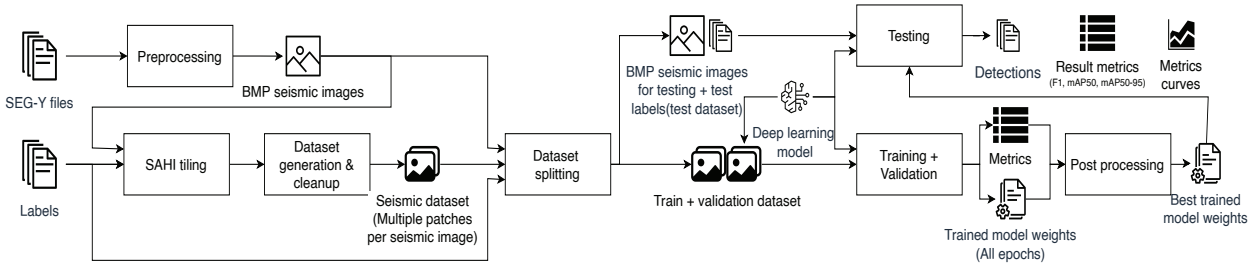


Fig. 2. Graphical outline of the experimental framework

To build an object detection model and evaluate its performance, a dataset containing the full size transformed seismic images, SAHI sliced input patches and labels (bounding boxes) was prepared. The dataset was split into training, validation, and test subsets targeting the ratio of 70%:15%:15%, respectively. Special care was taken to prevent any data leaks from test data into training data by splitting at the full-image level so images included in the test subset (including SAHI patches) were isolated and not used in any way in the training data. Because individual seismic images contain varying densities of annotated hyperbolic targets, although the random image-level selection targeted a ~15% test ratio, it inherently resulted in slight variation in total target counts per subset from the target. It is important to note that since data is acquired at different survey sites (blocks described earlier), it was ensured that each of the subsets contains samples with a proportional number of hyperbolae from each block as in the original dataset. This way, any hyperbolae from one subset will never be present in any other subset, whilst balancing the amount of samples from all blocks amongst subsets. Input patches were directly used for training and validation, whereas full seismic images were used for testing. During model training, after each epoch model performance was evaluated on the validation data and early stopping was applied, i.e., stopping occurred after 30 epochs without improvement. Finally, after training the models for all epochs, the model yielding the best mAP@50-95 metric on the validation data was selected and used for testing.

During model testing, SAHI-tiled inference was fully utilised. Tiled hyper-inference was run across the whole seismic image, and the seismic images were divided into overlapping patches. The size of patch used was the same as patch size used in the creation of the training dataset, i.e., 640×640, whilst patches overlapped at half that size, i.e., 320×320 pixels. After inference on every patch, the detection bounding boxes are transformed back to the full seismic image and non-max suppression (NMS) and confidence thresholding are applied to clean redundant detections. For NMS, the IoU threshold of 50% was chosen. Regarding confidence thresholding, the threshold of 50% was applied for metric calculation, although all detections with confidence of ≥5% were kept during inference to facilitate further analysis. After inference, the detections were saved and then used to compare against labeled ground truths to assess model performance.

Two DL-based object detectors frequently employed in the literature for similar problems (as shown in Section 2) were utilised, YOLOv11n and Faster R-CNN. Table 1 shows the hyper-parameter configurations used. Each model was trained and evaluated on data generated by all of the transformation techniques. All default online data augmentations (e.g., Mosaic, MixUp, random geometric transformations, jitter etc.) were explicitly disabled across all training runs for both detection models. All metrics were taken at a confidence threshold of 0.5 and IoU threshold of 0.5. It is important to acknowledge here the inherent performance trade-off between transfer learning and training models from scratch in the context of SBP data. Preliminary evaluations conducted during the design phase of this study heavily indicated that training from scratch yielded inferior overall performance. Conversely, initialising the models with pre-trained weights substantially improved the TP detections and minimised FN, albeit at the cost of an elevated FP rate, even though the pre-training was done on general natural images, not specific problem data containing hyperbolic signatures. As maximising recall is a critical priority in geophysical workflows with human-in-the-loop, transfer learning was universally adopted for evaluated models.

3.3. RESULTS AND DISCUSSION

The comprehensive set of evaluation metrics for each transformation technique are summarised in Table 2 for the YOLOv11n model and in Table 3 for the Faster R-CNN model where unweighted arithmetic means over three separate spatial splits of data are reported. The exact same three spatial image splits were held strictly constant across all 13 preprocessing transformations and both detection architectures to guarantee identical, controlled evaluation conditions. The assessed techniques include GMM, per-trace min-max (PT-MM), PC with three different low and high percentiles ($PC_{1\text{-}99}$, $PC_{2\text{-}98}$, $PC_{5\text{-}95}$), RMS normalisation, z-score standardisation (Z-score), LCN with three different window sizes ($LCN_{30\times30}$, $LCN_{100\times100}$, $LCN_{400\times60}$) and CLAHE with three different tile sizes and clip limits ($CLAHE_{32\times32\text{-}1}$, $CLAHE_{64\times64\text{-}1}$, $CLAHE_{128\times128\text{-}1}$). For the CLAHE implementations, a composite variant suggested in [34] utilising a 3×3 pixels Gaussian filter was applied, tackling high-frequency intensity artefacts which would otherwise be disproportionately amplified by the adaptive equalisation algorithm, as was inferred during the experimental design phase of this study.

Table 1. Deep learning models and hyper-parameters used in the experiments

Model	YOLO	Faster R-CNN
Variant	YOLOv11n	MobileNet v3 Large + FPN
Pretraining	COCO	
Optimiser	AdamW	SGD
Learning rate	0.002	0.00005
Momentum	/	0.9
Weight decay	0.00005	
Epochs	100	
Batch size	16	

Whilst the above-mentioned techniques represent a curated set of potential choices, they serve to highlight that the method used to transform raw seismic amplitudes into digital images is a critical determinant of downstream object detection success.

This is underscored by the variance in detection performance across experimental runs, despite the fact that training hyper-parameters, local architectures, and ground truth annotations remained completely identical. When looking at the results overall, it is noticeable that both models perform the best with techniques that bound or mitigate amplitude outliers, such as percentile clipping and variants of CLAHE. Specifically, these techniques yield the highest TP counts and conversely the lowest FN counts, meaning they are the most effective at preventing real hyperbolic traces from being overlooked. This hints that the outliers present in the dataset squashing hyperbolae indeed represent an issue and eliminating or penalising them directly facilitates better convergence.

Distinct architectural preferences are visible, as some techniques appear to be better suited for certain detectors. One such example is the $LCN_{400\times60}$ which proved to be a good performer when combined with Faster R-CNN matching the performance of CLAHE in some aspects, albeit generating a higher number of detections (and thus false positives). The same technique cannot, however, be recommended as a viable tool when YOLOv11n is under consideration as the model of choice. The rectangular shape of the window for this technique, along with competitive performance shown by the PT-MM technique both indicate that scaling data along individual acoustic traces preserves localised hyperbolic waveforms in a manner that favours two-stage region-proposal networks, warranting further research into trace-centric techniques.

Another issue that is easily discernible from the shown results is the presence of a severe trade-off between missing hyperbolae and detecting them where there are none. The techniques which achieved the best results in terms of TP and FN are usually the worst ones in terms of FP (e.g. $CLAHE_{64\times64\text{-}1}$ and $LCN_{400\times60}$ for the Faster R-CNN and $PC_{5\text{-}95}$ and $CLAHE_{32\times32\text{-}1}$ for YOLO), which is reflected in the attained precision and recall values for both models. Here one might argue that precision is arguably low for all of the techniques. Whilst a generalised precision below 0.5 would be problematic in standard computer vision tasks, it reflects a well-documented domain constraint in shallow marine geophysics. Further, it must be noted that even human annotators frequently disagree whether a faint or fragmented hyperbolic shape is relevant and should be included in their report. The data on which the models have been trained reflects these inherent biases. Inter-rater agreement and metrics taking it into account might prove beneficial for this matter, but the annotation process is slow and expensive so annotations by multiple experts is hard to obtain. Another contributing factor to the high false positive rate stems from the patch generation strategy, which purposefully extracted training patches centred on annotated hyperbolae whilst excluding pure-background patches (hard negatives). During sliding-window (SAHI) inference, the detector encounters unannotated background regions that were unrepresented during training, occasionally leading to over-proposed candidate regions. Whilst incorporating a controlled proportion (e.g., 10–20%) of negative background patches during training can ameliorate this, high recall remains preferable in human-in-the-loop engineering workflows. In practical SBP interpretation, where the system serves as a decision-support tool for human geophysicists, rejecting false candidate proposals is significantly less labor-intensive than manually searching for unprompted, faint hyperbolic targets across extended line profiles.

Turning attention to the models, YOLO proved highly sensitive to the technique applied. The number of correctly identified hyperbolae varies significantly across techniques, as does the number of overall detections. This is not the case for the two stage Faster R-CNN, however, as it demonstrated more stability in this regard, albeit at the expense of numerous total detections. This behaviour is structurally tied to Faster R-CNN's region-proposal network (RPN), which heavily samples prospective regions from background clutter. Notably, Faster R-CNN yielded significantly higher average confidence scores for correct decisions, implying that its high false-positive rate could be effectively mitigated in post-processing by raising the confidence threshold above the initial evaluation baseline of 0.5. Consequently, given its superior stability and higher overall recall potential, Faster R-CNN represents the preferred choice for this subsurface application, with future optimisations focused on tuning per-trace variations and adaptive block-size parameters in local contrast techniques.

Finally, the attained results, although obtained on an arguably large dataset for this specific problem, do not render any specific transformation technique a clear choice overall. They do, however point to the importance of carefully choosing which one is to be used depending on the model and that they do indeed have a profound impact on the feature extraction capabilities of deep learning architectures. Moreover, they hint at the usefulness of clipping outliers, highlight the need of looking at trace-based procedures and enhancing contrast.

To accompany the tabular data and serve as visual reference, an example of detection on unseen test data for $CLAHE_{64\times64\text{-}1}$ and Faster R-CNN is provided in Fig. 3. In the example, the blue bounding boxes denote ground truths, the green bounding boxes denote true positives (with IoU=0.5), whilst the red bounding boxes denote false positives. Although only a single seismic image is annotated utilising the tool, the overall outlook is similar regardless of the selected image. The model struggles with detecting hyperbolae surrounded by seabed transitions, some occasions of a smaller bounding boxes leading to false positives whilst overlapping a ground truth bounding box are visible, and multiple instances of hyperbola like shapes are annotated which were not amongst the annotated instances.

Table 2. Performance of the transformation techniques with YOLOv11n. Note: Shaded cells contain the best attained values for specific metrics. Values are reported as unweighted arithmetic means across three splits accompanied by the relative standard deviation [%].

Technique	GT μ±RSD	Det. μ±RSD	TP μ±RSD	FP μ±RSD	FN μ±RSD	Avg. conf. μ±RSD	Avg. IoU μ±RSD	Precision μ±RSD	Recall μ±RSD	F1 μ±RSD	F2 μ±RSD	mAP@50 μ±RSD
GMM	1789.00 ±1.95	1779.67 ±58.77	822.00 ±32.06	957.67 ±82.75	967.00 ±25.98	0.654 ±5.87	0.758 ±1.47	0.510 ±27.64	0.459 ±31.33	0.459 ±8.07	0.453 ±19.40	0.435 ±13.85
PT-MM	1789.00 ±1.95	1478.67 ±11.81	861.33 ±5.37	617.33 ±21.33	927.67 ±6.05	0.653 ±2.33	0.776 ±1.93	0.586 ±7.44	0.482 ±5.58	0.527 ±0.68	0.499 ±3.52	0.544 ±2.12
PC_{1-99}	1789.00 ±1.95	1766.00 ±48.15	889.33 ±25.58	876.67 ±71.87	899.67 ±24.51	0.658 ±3.23	0.770 ±1.07	0.542 ±24.70	0.497 ±25.02	0.500 ±5.08	0.494 ±15.25	0.506 ±10.72
PC_{2-98}	1789.00 ±1.95	1332.33 ±19.30	840.00 ±11.12	492.33 ±33.65	949.00 ±6.65	0.649 ±1.55	0.779 ±0.67	0.637 ±7.79	0.469 ±9.37	**0.538** ±2.22	0.494 ± 6.59	**0.553** ± 2.10
PC_{5-95}	1789.00 ±1.95	1860.33 ±42.30	**956.33** ±18.44	904.00 ±67.68	**832.67** ±20.47	0.653 ±2.78	0.774 ±0.72	0.546 ±21.16	**0.534** ±17.94	0.526 ±3.38	**0.527** ±9.00	0.534 ±2.49
RMS	1789.00 ±1.95	1518.00 ± 30.88	877.00 ±19.67	641.00 ±46.23	912.00 ±15.52	**0.663** ±2.56	0.778 ±0.88	0.595 ±14.16	0.489 ±18.16	0.528 ±5.03	0.502 ±13.42	0.545 ±4.74
$Z\text{-score}_{\sigma=2}$	1789.00 ±1.95	1494.33 ±21.08	844.00 ±13.53	650.33 ±31.12	945.00 ±9.26	**0.663** ±5.08	0.781 ±0.85	0.570 ±7.14	0.471 ±11.91	0.513 ±3.12	0.486 ±8.43	0.521 ±2.04
$LCN_{30\times30}$	1789.00 ±1.95	988.67 ±22.56	688.67 ±15.96	**300.00** ±39.68	1100.33 ±8.38	0.633 ±3.36	**0.784** ±0.88	**0.704** ±8.71	0.385 ±14.69	0.494 ±7.14	0.422 ±12.04	0.535 ±3.98
$LCN_{100\times100}$	1789.00 ±1.95	1154.00 ±17.86	748.67 ±8.26	405.33 ±35.88	1040.33 ±9.10	0.650 ±0.36	0.772 ±1.20	0.656 ±8.99	0.419 ±10.17	0.508 ±2.47	0.450 ±7.32	0.518 ±1.14
$LCN_{400\times60}$	1789.00 ±1.95	1650.67 ±28.87	888.00 ±17.30	762.67 ±43.25	901.00 ±13.23	0.646 ±1.28	0.774 ±1.04	0.550 ±11.76	0.495 ±15.50	0.515 ±4.80	0.502 ±10.86	0.522 ±5.34
$CLAHE_{32\times32-1}$	1789.00 ±1.95	**2023.67** ±30.10	943.33 ±3.37	1080.33 ±54.06	845.67 ±6.61	0.659 ±4.17	0.772 ±1.90	0.490 ±24.70	0.528 ±4.56	0.501 ±11.79	0.514 ±2.82	0.503 ±12.78
$CLAHE_{64\times64-1}$	1789.00 ±1.95	1766.00 ±61.65	855.00 ±30.45	911.00 ±93.33	934.00 ±31.20	0.646 ±6.42	0.772 ±1.58	0.559 ±32.77	0.480 ±32.03	0.480 ±11.58	0.473 ±22.02	0.520 ±4.42
$CLAHE_{128\times128-1}$	1789.00 ±1.95	1923.67 ±71.10	870.67 ±31.20	1053.00 ±104.32	918.33 ±33.37	0.649 ±3.29	0.778 ±1.28	0.537 ±33.35	0.489 ±33.28	0.474 ±5.91	0.472 ±17.51	0.509 ±2.70

Table 3. Performance of the transformation techniques with Faster R-CNN. Note: Shaded cells contain the best attained values for specific metrics. Values are reported as unweighted arithmetic means across three splits accompanied by the relative standard deviation [%].

Technique	GT μ±RSD	Det. μ±RSD	TP μ±RSD	FP μ±RSD	FN μ±RSD	Avg. conf. μ±RSD	Avg. IoU μ±RSD	Precision μ±RSD	Recall μ±RSD	F1 μ±RSD	F2 μ±RSD	mAP@50 μ±RSD
GMM	1789.00 ±1.95	2550.00 ±7.63	1116.33 ±4.05	1433.67 ±11.98	672.67 ±5.22	0.769 ±1.58	0.742 ±1.24	0.439 ±6.43	0.624 ±3.16	0.515 ±3.22	0.575 ±2.02	0.540 ±4.34
PT-MM	1789.00 ±1.95	2813.33 ±18.24	1178.33 ±6.69	1635.00 ±27.23	610.67 ±12.10	0.778 ±1.42	0.748 ±1.21	0.425 ±13.22	0.659 ±6.22	0.514 ±5.91	**0.591** ±1.99	**0.567** ±2.62
PC_{1-99}	1789.00 ±1.95	2557.67 ±12.93	1147.00 ±6.20	1410.67 ±18.87	642.00 ±8.36	0.775 ±0.39	**0.751** ±0.68	0.451 ±7.11	0.641 ±5.08	**0.528** ±2.62	0.590 ±2.09	0.566 ±2.03
PC_{2-98}	1789.00 ±1.95	2631.00 ±24.71	1135.33 ±8.86	1495.67 ±36.84	653.67 ±17.30	0.775 ±1.48	0.749 ±0.97	0.444 ±17.90	0.635 ±9.35	0.517 ±6.06	0.579 ±2.64	**0.567** ±1.45
PC_{5-95}	1789.00 ±1.95	2937.00 ±16.64	1182.33 ±4.94	1754.67 ±24.90	606.67 ±8.47	0.771 ±1.95	0.744 ±0.26	0.408 ±11.93	0.661 ±4.31	0.502 ±5.96	0.586 ±0.93	0.558 ±2.88
RMS	1789.00 ±1.95	2629.00 ±10.85	1139.33 ±7.96	1489.67 ±13.18	649.67 ±9.48	0.768 ±2.03	0.747 ±1.45	0.434 ±3.23	0.636 ±6.35	0.516 ±1.21	0.582 ±3.90	0.552 ±3.32
$Z\text{-score}_{\sigma=2}$	1789.00 ±1.95	2704.00 ±8.41	1165.00 ±5.38	1539.00 ±12.46	624.00 ±7.56	0.769 ±1.91	0.747 ±1.13	0.432 ±6.57	0.651 ±4.33	0.519 ±3.64	0.591 ±2.98	0.566 ±3.58
$LCN_{30\times30}$	1789.00 ±1.95	2450.67 ±14.36	1102.00 ±7.13	**1348.67** ±20.34	687.00 ±8.83	0.760 ±0.42	0.744 ±0.96	**0.453** ±7.01	0.616 ±6.00	0.520 ±1.68	0.573 ±2.56	0.548 ±2.66
$LCN_{100\times100}$	1789.00 ±1.95	3003.33 ±3.21	1176.33 ±2.39	1827.00 ±3.80	612.67 ±2.22	0.771 ±1.53	0.750 ±0.61	0.392 ±1.05	0.658 ±0.92	0.491 ±0.32	0.579 ±0.34	0.551 ±1.40
$LCN_{400\times60}$	1789.00 ±1.95	**3101.33** ±19.53	**1194.00** ±6.52	1907.33 ±27.70	**595.00** ±13.96	0.777 ±1.04	0.745 ±0.53	0.391 ±12.58	**0.668** ±6.60	0.491 ±5.66	0.582 ±0.99	0.561 ±2.13
$CLAHE_{32\times32-1}$	1789.00 ±1.95	2791.00 ±16.79	1167.33 ±4.91	1623.67 ±25.42	621.67 ±11.42	0.772 ±0.55	0.747 ±1.00	0.424 ±12.80	0.653 ±5.46	0.512 ±5.33	0.587 ±0.64	0.566 ±1.71
$CLAHE_{64\times64-1}$	1789.00 ±1.95	2949.67 ±14.93	**1194.00** ±5.03	1755.67 ±22.38	**595.00** ±12.71	**0.784** ±1.74	0.747 ±0.92	0.410 ±12.09	**0.668** ±5.71	0.505 ±5.37	**0.591** ±2.08	0.560 ±2.05
$CLAHE_{128\times128-1}$	1789.00 ±1.95	2588.00 ±14.62	1144.67 ±3.85	1443.33 ±24.34	644.33 ±10.47	0.776 ±1.55	0.747 ±0.97	0.447 ±12.02	0.640 ±5.01	0.524 ±5.81	0.587 ±2.60	**0.567** ±4.08

Tools which might prove useful for dealing with mislabelled instances include hard negative mining, adjustment of the confidence threshold, including more of the hyperbola environment for additional information and training a separate model on the output of the object detector to deal with the false positives. The general guideline here trying to avoid missing hyperbolae, and dealing with the false positives post hoc, as even a semi-automated procedure would be highly beneficial to annotators which would be able to reject suggestions not representing relevant objects.

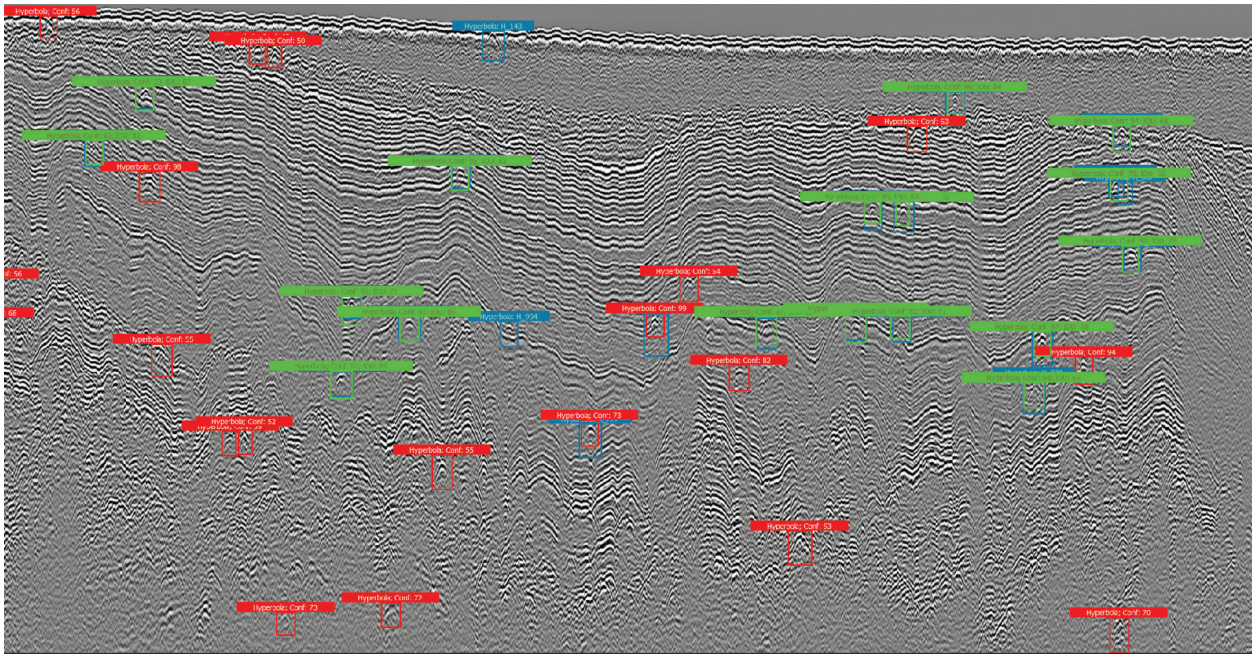


Fig. 3. An example of annotations on unseen test data (Confidence=0.5, IoU=0.5), where ground truths, true positives and false positives are represented by blue, green and red bounding boxes, respectively.

3.4. GENERALISATION TEST

An additional insight into the effect of the considered techniques on the generalisation capabilities of downstream models can be gained from Table 4. The results of a separate experiment are reported in which the trained Faster R-CNN model from the first experiment was employed, as it was the better performing one amongst the two considered, on an unseen, geographically separate block of seismic data. The utilised block consists of 13 SEG-Y files with 171 annotated hyperbolae in total. Two important points must be clarified. Firstly, the entire block (region) of seismic images is previously unseen by the model, i.e., no data from this entire block can be found in the training subset. Secondly, the data annotations differ significantly in the amount of annotated hyperbolae when compared to the original data utilised in the first experiment, highlighting the differences between annotator operation. Both of the aforementioned facts render this experiment a challenging real-world test of the included transformation techniques and their influence. The first thing that stands out when considering the obtained results is the number of false positives, which is unreasonably high, even accounting for the (expected) large number of detections generated by Faster R-CNN. The reasons for this should primarily be sought after in the expert annotations, as the number of ground truths is considerably lower for the images contained in this new block than in similarly sized images in previously used blocks, even though they remain visually similar. What can be observed, despite this discrepancy, is the fact that the best results overall on this new data are obtained when utilising percentile clipping, and that the PT-MM has shown overall the highest number of TP. Since these approaches are followed by the CLAHE with a small window and LCN with a rectangular one, the conclusions reached in the first experiment still hold, at least to a degree. Utilising techniques which scale data along individual acoustic traces and remove outliers do indeed seem to be favourable choices. Certainly, a single block is not enough to make any bold claims, but the outcome is symptomatic and requires further investigation on yet more separate blocks of different geographical properties.

3.5. ACTIONABLE GUIDANCE

To summarise, based on the described empirical findings, actionable guidance for practitioners deploying deep learning on SBP data can be offered. For general-purpose analysis, percentile clipping approaches, especially trace-based ones (e.g., PT-MM), are recommended as a default baseline, as they standardise amplitudes without introducing high-frequency artifacts. In high-noise environments or in the case of faint targets, composite local pipelines such as CLAHE-based ones would be preferable.

Although they might increase the false positive rate, Gaussian smoothing enables them to reduce background speckle and better isolate hyperbolic signatures in these complex environments. Whilst they do require careful parametrisation, based on the conducted analysis it seems that utilising a square window of 32×32 or 64×64 (roughly the size or double the size of the smallest targets) and a small window size for the Gaussian filter (e.g. 3×3 pixels) represents a good starting point. Furthermore, both PT-MM and CLAHE pipelines seem suitable for human-in-the-loop workflows where the false-positive rejection is left to the human operator. For resource constrained environments, where real-time inference is expected, complex operators should be avoided in favour of (computationally) simpler PT-MM or PC techniques.

Table 4. Performance of the transformation techniques with Faster R-CNN on a previously unseen block of data

Technique	GT	Det.	TP	FP	FN	Avg. conf.	Avg. IoU	Precision	Recall	F1	F2	mAP@50
GMM	171	1175	83	1092	88	0.703	**0.710**	0.071	0.485	0.123	0.223	0.147
PT-MM	171	1415	**118**	1297	**53**	0.713	0.694	0.083	**0.690**	0.149	0.281	0.194
$PC_{1\text{-}99}$	171	1130	113	1017	58	0.724	0.700	0.100	0.661	0.174	**0.311**	**0.262**
$PC_{2\text{-}98}$	171	1344	114	1230	57	0.710	0.699	0.085	0.667	0.151	0.281	0.235
$PC_{5\text{-}95}$	171	1199	106	1093	65	0.732	0.677	0.088	0.620	0.155	0.281	0.213
RMS	171	**387**	51	**336**	120	0.702	0.675	**0.132**	0.298	**0.183**	0.238	0.093
$Z\text{-score}_{\sigma=2}$	171	1363	107	1256	64	0.707	0.705	0.079	0.626	0.140	0.261	0.185
$LCN_{30\times30}$	171	882	96	786	75	0.698	0.700	0.109	0.561	0.182	0.307	0.253
$LCN_{100\times100}$	171	1387	102	1285	69	0.712	0.699	0.074	0.596	0.131	0.246	0.197
$LCN_{400\times60}$	171	1595	108	1487	63	**0.735**	0.701	0.068	0.632	0.122	0.237	0.210
$CLAHE_{32\times32\text{-}1}$	171	1330	107	1223	64	0.730	0.692	0.080	0.626	0.143	0.266	0.158
$CLAHE_{64\times64\text{-}1}$	171	1509	103	1406	68	**0.735**	0.685	0.068	0.602	0.123	0.235	0.151
$CLAHE_{128\times128\text{-}1}$	171	1416	100	1316	71	0.725	0.691	0.071	0.585	0.126	0.238	0.175

4. CONCLUSION

This paper investigates the critical role of image transformation techniques when applying deep learning object detection models to seismic data acquired via sub-bottom profiling. Their importance is emphasised by data scarcity and operational costs of acquiring additional data, warranting the use of transfer learning and models pre-trained on general out-of-domain digital image data. Through an experimental analysis on real-world datasets, several established scaling techniques such as percentile clipping, LCN, and CLAHE are investigated as critical components of an end-to-end detection pipeline featuring SAHI. Two distinct DL architectures (an anchor-free single-stage YOLOv11n and a two-stage Faster R-CNN) were employed to serve as proxies for the effectiveness of the considered transformation techniques.

The obtained results demonstrate a clear imperative for mitigating high-amplitude acoustic outliers. Percentile clipping and CLAHE showcased favourable performance regardless of the model utilised, maximising true positive target recovery and minimising missed hyperbolae. The results highlight a synergy between trace-wise scaling techniques and the two-stage region-proposal framework. Whilst the two-stage Faster R-CNN emerged as the more robust and stable model across all input variations, it generated an inflated volume of false positive detections. However, given the model's high average confidence scores on valid targets, these false positives could be managed via post-processing confidence thresholds. Finally, this work highlights a critical domain-specific limitation: the detection metrics, taken at face value, could be misleading, as the annotation is a subjective process (probably more so than in other areas of research or industry). This is emphasized by the second part of the experimental analysis where a new, previously unseen block of data was utilised, and which had discernible differences in annotated object count, when compared to the original data. Measuring inter-rater agreement would be beneficial for establishing the baseline for model performance evaluation (and correspondingly the transformation techniques) but is rendered difficult by the incurred costs.

Whilst this study offers valuable practical insights for the marine geophysics and applied machine learning communities, several dataset and methodological constraints and generalisation limitations remain to be addressed in future work. First, due to computational constraints, our results are based on three separate splits per configuration. Although the broader performance trends remain consistent across architectures, minor metric variations should be interpreted with caution, as they do not fully account for deep learning training stochasticity or random seed variability. Second, because the utilised dataset originates from a single acquisition system under similar survey conditions (e.g. parametric SBP data acquired in shallow-water environments), further research is required to fully assess how these preprocessing techniques generalise to alternative sensors, varying acoustic frequencies, and geological environments of significant diversity, such as deep-water datasets, chirps/sparkers, and heterogeneous seabed geologies.

Future research will focus on establishing statistical robustness through additional multi-seed training and cross-validation across more distinct geological blocks, utilising other evaluation approaches (such as leave-one-block-out). The utilisation of other trace-wise

techniques, post-processing methods and alternative DL architectures is also planned. More complex image preprocessing techniques, such as edge detection or the exploitation of the frequency domain for noise removal also represent viable avenues of research, as well as utilising FP reduction techniques such as hard negative mining or post inference classification of true and false detections.

ACKNOWLEDGMENT

This research was supported in part by the project "Development of spatial data processing models and machine learning in water surface research projects" (IP.1.1.03.0022). The authors would like to thank the anonymous reviewers for their valuable comments and constructive suggestions, which helped improve the quality of this manuscript.

DECLARATION OF THE USE OF AI TOOLS

The authors state that no AI tools were used for content generation.

REFERENCES

[1] X. Lurton, "An Introduction to Underwater Acoustics: Principles and Applications", Springer, 2002. doi: 10.1007/978-3-662-05165-8

[2] E. Kozaczka, G. Grelowska, S. Kozaczka, W. Szymczak, "Detection of Objects Buried in the Sea Bottom with the Use of Parametric Echosounder", Archives of Acoustics, Vol. 38, No. 1, 2013, pp. 99-104. doi: 10.2478/aoa-2013-0012

[3] S. Li, Y. Zhang, J. Zhao, Y. Wu, S. Bian, G. Zhai, "A Comprehensive Buried Shipwreck Detection Method Based on 3-D SBP Data", IEEE Journal of Oceanic Engineering, Vol. 49, No. 2, 2024, pp. 458-473. doi: 10.1109/JOE.2023.3318793

[4] J. Hou, C. Zhang, "Shallow Mud Detection Algorithm for Submarine Channels Based on Improved YOLOv5s", Heliyon, Vol. 10, No. 10, 2024, p. e31029. doi: 10.1016/j.heliyon.2024.e31029

[5] M. E. Vardy, M. Vanneste, T. J. Henstock, M. A. Clare, C. F. Forsberg, G. Provenzano, "State-of-the-Art Remote Characterization of Shallow Marine Sediments: The Road to a Fully Integrated Solution", Near Surface Geophysics, Vol. 15, No. 4, 2017, pp. 387-402. doi: 10.3997/1873-0604.2017024

[6] C.-C. Tsai, C.-H. Lin, "Review and Future Perspective of Geophysical Methods Applied in Nearshore Site Characterization", Journal of Marine Science and Engineering, Vol. 10, No. 3, 2022, p. 344. doi: 10.3390/jmse10030344

[7] M.-T. Pham, S. Lefèvre, "Buried Object Detection from B-Scan Ground Penetrating Radar Data Using Faster-RCNN", Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22-27 July 2018, pp. 6804-6807. doi: 10.1109/IGARSS.2018.8517683

[8] T. Wunderlich, D. Wilken, B. S. Majchczack, W. Rabbel, "Hyperbola Detection with RetinaNet and Comparison of Hyperbola Fitting Methods in GPR Data from an Archaeological Site", Remote Sensing, Vol. 14, No. 15, 2022, p. 3665. doi: 10.3390/rs14153665

[9] J. Zhu, D. Zhao, X. Luo, "Evaluating the Optimised YOLO-Based Defect Detection Method for Subsurface Diagnosis with Ground Penetrating Radar", Road Materials and Pavement Design, Vol. 25, No. 1, 2024, pp. 186-203. doi: 10.1080/14680629.2023.2199880

[10] S. Park, J. Kim, K. Jeon, J. Kim, S.-H. Park, "Improvement of GPR-Based Rebar Diameter Estimation Using YOLO-v3", Remote Sensing, Vol. 13, No. 10, 2021, p. 2011. doi: 10.3390/rs13102011

[11] T. X. H. Luo, W. W. L. Lai, Z. Lei, "Intensity Normalisation of GPR C-Scans", Remote Sensing, Vol. 15, No. 5, 2023, p. 1309. doi: 10.3390/rs15051309

[12] X. Bai, Y. Yang, S. Wei, G. Chen, H. Li, Y. Li, H. Tian, "A Comprehensive Review of Conventional and Deep Learning Approaches for Ground-Penetrating Radar Detection of Raw Data", Applied Sciences, Vol. 13, No. 13, 2023, p. 7992. doi: 10.3390/app13137992

[13] S. Ren, K. He, R. Girshick, J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", Advances in Neural Information Processing Systems, Vol. 28, Montreal, Canada, 7-12 December 2015, pp. 91-99.

[14] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27-30 June 2016, pp. 779-788. doi: 10.1109/CVPR.2016.91

[15] W. Lei, F. Hou, J. Xi, Q. Tan, M. Xu, X. Jiang, G. Liu, Q. Gu, "Hyperbola Detection and Fitting in GPR B-Scan Image", Automation in Construction, Vol. 106, 2019, p. 102839. doi: 10.1016/j.autcon.2019.102839

[16] Q. Dou, L. Wei, D. R. Magee, A. G. Cohn, "Real-Time Hyperbola Recognition and Fitting in GPR Data", IEEE Transactions on Geoscience and Remote Sensing, Vol. 55, No. 1, 2017, pp. 51-62. doi: 10.1109/TGRS.2016.2592679

[17] L. Capineri, P. Grande, J. A. G. Temple, "Advanced image-processing technique for real-time interpretation of ground-penetrating radar images", International Journal of Imaging Systems and Technology, Vol. 9, No. 1, 1998, pp. 51-59.

[18] Z. Peng, M. Li, B. Pan, Y. Du, D. Gao, "High-resolution hyperbolic radon transform ground penetrating radar target signal extraction based on time-scale correction", Journal of Applied Geophysics, Vol. 241, 2025, p. 105819. doi: 10.1016/j.jappgeo.2025.105819

[19] X. Zheng, S. Fang, H. Chen, L. Peng, Z. Ye, "Internal Detection of Ground-Penetrating Radar Images Using YOLOX-s with Modified Backbone", Electronics, Vol. 12, No. 16, 2023, p. 3520. doi: 10.3390/electronics12163520

[20] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, "Focal Loss for Dense Object Detection", Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22-29 October 2017, pp. 2980-2988. doi: 10.1109/ICCV.2017.324

[21] X.-S. Tang *et al*. "Automatic detection of single-hyperbola pipeline signatures in urban roads using data-augmented ground penetrating radar profiles", Results in Engineering, Vol. 29, 2026, p. 108478. doi: 10.1016/j.rineng.2025.108478

[22] I. Mesecan, I. O. Bucak, "Efficient underground object detection for ground penetrating radar signals", Defence Science Journal, Vol. 67, No. 1, 2017, pp. 12-18. doi: 10.14429/dsj.67.9063

[23] Z. Huang, G. Xu, J. Tang, H. Yu, D. Wang, "Research on Void Signal Recognition Algorithm of 3D Ground-Penetrating Radar Based on the Digital Image", Frontiers in Materials, Vol. 9, 2022. doi: 10.3389/fmats.2022.850694

[24] P. Asadi, M. Gindy, M. Alvarez, "A machine learning based approach for automatic rebar detection and quantification of deterioration in concrete bridge deck ground penetrating radar B-scan images", KSCE Journal of Civil Engineering, Vol. 23, No. 6, 2019, pp. 2618-2627. doi: 10.1007/s12205-019-2012-z

[25] S. Yu, J. Yu, L. Chen, B. Zhu, X. Liang, Y. Xie, Q. Sun, "Advances and Challenges in Respiratory Sound Analysis: A Technique Review Based on the ICBHI2017 Database", Electronics, Vol. 14, No. 14, 2025, p. 2794. doi: 10.3390/electronics14142794

[26] K. Ishitsuka, S. Iso, K. Onishi, T. Matsuoka, "Object Detection in Ground-Penetrating Radar Images Using a Deep Convolutional Neural Network and Image Set Preparation by Migration", International Journal of Geophysics, Vol. 2018, No. 1, 2018, p. 9365184. doi: 10.1155/2018/9365184

[27] S. Xie, J. Chen, M. Cai, Z. Cheng, S. Wang, Y. Zhang, "GPR Feature Enhancement of Asphalt Pavement Hidden Defects Using Computational-Efficient Image Processing Techniques", Materials, Vol. 18, No. 18, 2025, p. 4400. doi: 10.3390/ma18184400

[28] Innomar Technologie GmbH, "Innomar standard Parametric Sub-Bottom Profiler", 2026. https://innomar.com/innomar-flyer/Innomar-standard.pdf (accessed: 7 May 2026)

[29] Hidrocibalae d.o.o., "Hidrocibalae - Hydrographic data processing centre", 2026. https://hidrocibalae.hr (accessed: 7 May 2026)

[30] Sunjay, "Sub Bottom Profiling Chirp Signal Processing", Proceedings of GeoConvention, Calgary, Canada, 13-17 May 2019.

[31] D. Dondurur, "Acquisition and Processing of Marine Seismic Data", Elsevier, 2018. doi: 10.1016/C2016-0-04353-8

[32] K. Wada, "labelme, Image annotation with Python", 2026. https://github.com/wkentaro/labelme (accessed: 7 May 2026)

[33] F. C. Akyon, S. O. Altinuc, A. Temizel, "Slicing Aided Hyper Inference and Fine-Tuning for Small Object Detection", Proceedings of the International Conference on Image Processing, Bordeaux, France, 16-19 October 2022, pp. 966-970. doi: 10.1109/ICIP46576.2022.9897990

[34] K. Halloum, H. Ez-Zahraouy, "Enhancing Medical Image Classification through Transfer Learning and CLAHE Optimization", Current Medical Imaging, Vol. 21, 2025. doi: 10.2174/0115734056342623241119061744